

Lecture Notes in Electrical Engineering 375

Jason C. Hung
Neil Y. Yen
Kuan-Ching Li
Editors

Frontier Computing

Theory, Technologies and Applications

 Springer

Lecture Notes in Electrical Engineering

Volume 375

Board of Series editors

Leopoldo Angrisani, Napoli, Italy
Marco Arteaga, Coyoacán, México
Samarjit Chakraborty, München, Germany
Jiming Chen, Hangzhou, P.R. China
Tan Kay Chen, Singapore, Singapore
Rüdiger Dillmann, Karlsruhe, Germany
Haibin Duan, Beijing, China
Gianluigi Ferrari, Parma, Italy
Manuel Ferre, Madrid, Spain
Sandra Hirche, München, Germany
Faryar Jabbari, Irvine, USA
Janusz Kacprzyk, Warsaw, Poland
Alaa Khamis, New Cairo City, Egypt
Torsten Kroeger, Stanford, USA
Tan Cher Ming, Singapore, Singapore
Wolfgang Minker, Ulm, Germany
Pradeep Misra, Dayton, USA
Sebastian Möller, Berlin, Germany
Subhas Mukhopadhyay, Palmerston, New Zealand
Cun-Zheng Ning, Tempe, USA
Toyoaki Nishida, Sakyo-ku, Japan
Bijaya Ketan Panigrahi, New Delhi, India
Federica Pascucci, Roma, Italy
Tariq Samad, Minneapolis, USA
Gan Woon Seng, Nanyang Avenue, Singapore
Germano Veiga, Porto, Portugal
Haitao Wu, Beijing, China
Junjie James Zhang, Charlotte, USA

About this Series

“Lecture Notes in Electrical Engineering (LNEE)” is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer’s high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer’s other Lecture Notes series, LNEE will be distributed through Springer’s print and electronic publishing channels.

More information about this series at <http://www.springer.com/series/7818>

Jason C. Hung · Neil Y. Yen
Kuan-Ching Li
Editors

Frontier Computing

Theory, Technologies and Applications

 Springer

Editors

Jason C. Hung
Department of Information Technology
Overseas Chinese University
Taichung
Taiwan

Kuan-Ching Li
Providence University
Taichung
Taiwan

Neil Y. Yen
Aizu-Wakamatsu, Fukushima
Japan

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-0538-1 ISBN 978-981-10-0539-8 (eBook)
DOI 10.1007/978-981-10-0539-8

Library of Congress Control Number: 2016934668

© Springer Science+Business Media Singapore 2016

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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Preface

The International Conference on Frontier Computing—Theory, Technologies, and Applications (FC) was first proposed in early 2010 at an IET executive meeting. This conference series aims at providing an open forum to reach a comprehensive understanding of the recent advances and emergence of information technology, science, and engineering, with themes in the scope of Communication Network Technology and Applications, Communication Network Technology and Applications, Business Intelligence and Knowledge Management, Web Intelligence, and any related field that prompts the development of information technology. This will be the fourth event of the series, in which fruitful results can be found in the digital library or conference proceedings of FC 2010 (Taichung, Taiwan), FC 2012 (Xining, China), FC 2013 (Gwangju, Korea). Each event brings together researchers from worldwide to have excited and fruitful discussions as well as future collaborations.

The papers accepted for inclusion in the conference proceedings primarily cover the topics: database and data mining, networking and communications, web and Internet of things, embedded system, soft computing, social network analysis, security and privacy, optics communication, and ubiquitous/pervasive computing. Many papers have shown their great academic potential and value, and in addition, indicate promising directions of research in the focused realm of this conference series. We believe that the presentations of these accepted papers will be more exciting than the papers themselves, and lead to creative and innovative applications. We hope that the attendees (and readers as well) will find these results useful and inspiring to their field of specialization and future research.

On behalf of the organizing committee, we would like to thank the members of the organizing and the program committees, the authors, and the speakers for their dedication and contributions that made this conference possible. We would like to thank and welcome all participants to the capital city of Thailand—Bangkok. Bangkok is a country with a long and remarkable history. To get a picture of Southeast Asia, this city will certainly be an entry. Though most of the countries may share some similar characteristics, you will find that the culture of Thailand

is very rich from different perspectives, such as art, religion, nomadic lifestyle, food, and music. Bangkok is a world-class and well-known city, with modern facilities and stable weather. We encourage the participants to take this chance to see and experience Thailand, especially the remote counties and the nomadic lifestyle there. We also sincerely hope that all participants from overseas and from Thailand enjoy the technical discussions at the conference, build a strong friendship, and establish ties for future collaborations.

We convey our sincere appreciations to the authors for their valuable contributions and to the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

July 2015

Jason C. Hung
Neil Y. Yen
Kuan-Ching Li

Organization

Steering Chairs

Kuan-Ching Li, Providence University, Taiwan
Jason C. Hung, Overseas Chinese University, Taiwan
Neil Y. Yen, The University of Aizu, Japan

General Chairs

C.S. Raghavendra, University of Southern California, USA
Yi Pan, Georgia State University, USA
Hamid R. Arabnia, The University of Georgia, USA
Hong Shen, University of Adelaide, Australia
Jen-Shiun Chiang, Tamkang University, Taiwan
Qingguo Zhou, Lanzhou University, China

Vice General Chairs

Han-Chieh Chao, National Ilan University, Taiwan
Zheng Xu, Tsinghua University, China
Yasuji Sawada, Tohoku University of Technology, Japan
Eiko Yoneki, University of Cambridge, UK
Kurosh Madani, University of Paris-EST, France

Program Chairs

Keqiu Li, Dalian University of technology, China
Hai Jiang, Arkansas State University, USA

Kehan Zeng, University of Macau, Macau
Meng-Yen Hsieh, Providence University, Taiwan
Zhou Rui, Lanzhou University, China

Vice Program Chairs

Deqiang Han, Beijing University of Technology, China
Hsuan-Fu Wang, Chung Chou University of Science and Technology, Taiwan
Jun-Hong Shen, Asia University, Taiwan
Fang-Biau Ueng, National Chung Hsing University, Taiwan

Workshop Chairs

You-Shyang Chen, Hwa Hsia University of Technology, Taiwan
Wei-Chen Wu, Hsin Sheng College of Medical Care and Management, Taoyuan County, Taiwan
Chengjiu Yin, Kyushu University, Japan
Yan Pei, University of Aizu, Japan

Publicity Chairs

Vladimír Smejkal, Brno University of Technology, Czech Republic
Fei Wu, Zhejiang University, China
Francisco Isidro Massetto, Federal University of ABC, Brazil
Riz Sulaiman, Universiti Kebangsaan Malaysia, Malaysia
Wei Tsang Ooi, National University of Singapore, Singapore
Yusuke Manabe, Chiba Institute of Technology, Japan
Soumya Banerjee, Birla Institute of Technology, India
Tran Thien Phuc, Hochimin City University of Technology, Vietnam
Jindrich Kodl, Authorised expert in security of information systems, cryptology and informatics, Czech Republic
Poonphon Suesawaluk, Assumption University of Thailand, Thailand
Jenn-Wei Lin, Fu Jen University, Taiwan

International Advisory Committees

Jinannong Cao, Hong Kong Polytechnic University, Hong Kong
Su-Ching Chen, University of Florida, USA
Fatos Xhafa, Technical University of Catalonia, Spain

Jianhua Ma, Hosei University, Japan
Runhe Huang, Hosei University, Japan
Qun Jin, Waseda University, Japan
Victor Leung, University of British Columbia, Canada
Qing Li, City University of Hong Kong, Hong Kong
Jean-Luc Gaudiot, University of California, Irvine, USA

Contents

Cloud and Crowd Based Learning	1
Chun-Hsiung Tseng, Ching-Lien Huang, Yung-Hui Chen, Chu-Chun Chuang, Han-Ci Syu, Yan-Ru Jiang, Fang-Chi Tsai, Pin-Yu Su and Jun-Yan Chen	
Artificial Neural Network Based Evaluation Method of Urban Public Security	7
Zheng Xu, Qingyuan Zhou, Haiyan Chen and Fangfang Liu	
Building the Search Pattern of Social Media User Based on Cyber Individual Model	15
Zheng Xu, Xiao Wei, Dongmin Chen, Haiyan Chen and Fangfang Liu	
Design of Health Supervision System Base on WBAN	23
Xinli Zhou	
The Analysis of Hot Topics and Frontiers of Financial Engineering Based on Visualization Analysis	33
Liangbin Yang	
An Efficient ACL Segmentation Method	43
YunBo Rao, XianShu Ding, Jianping Gou and Ying Ma	
Image Haze Removal of Optimized Contrast Enhancement Based on GPU	53
Che-Lun Hung, Zhaohui Ma, Chun-Yuan Lin and Hsiao-Hsi Wang	
Research of Thunderstorm Warning System Based on Credit Scoring Model	65
Xinli Zhou, LiangBin Yang and HaiFeng Hu	

Cloud-Based Marketing: Does Cloud Applications for Marketing Bring Positive Identification and Post-purchase Evaluation? 77
Ching-Wei Ho and Yu-Bing Wang

Decision Analyses of Medical Resources for Disabled Elderly Home Care: The Hyper Aged District in Taiwan 85
Lin Hui and Kuei Min Wang

The Research About Vehicle Recognition of Parallel Computing Based on GPU. 97
Zhiwei Tang, Yong Chen and Zhiqiang Wen

Pseudo Nearest Centroid Neighbor Classification 103
Hongxing Ma, Xili Wang and Jianping Gou

Recommended System for Cognitive Assessment Evaluation Based on Two-Phase Blue-Red Tree of Rule-Space Model: A Case Study of MTA Course 117
Yung-Hui Chen, Chun-Hsiung Tseng, Ching-Lien Huang, Lawrence Y. Deng and Wei-Chun Lee

A Algorithm of Detectors Generating Based on Negative Selection Algorithm. 133
Wu Renjie, Guo Xiaoling and Zhang Xiao

A Comparative Study on Disease Risk Model in Exploratory Spatial Analysis 141
Zhisheng Zhao, Xiao Zhang, Yang Liu, Junhua Liang, Jiawei Wang and Yaxu Liu

An Algorithm for Image Denoising Based on Adaptive Total Variation 155
Guo Xiaoling, Yang Jie and Zhang Xiao

Social Events Detection and Tracking Based on Microblog 161
Guiliang Feng, Yiping Lu, Jing Qin and Xiao Zhang

An Optimization of the Delay Scheduling Algorithm for Real-Time Video Stream Processing 173
Hongbin Yang, Jianhua Guo, Chao Liang, Zhou Lei and Changsheng Wang

Microblogging Recruitment Information Mining. 185
Jing Qin, Yiping Lu, Shuo Feng and Guiliang Feng

Community Trust Recommendation Based on Probability Matrix Factorization 195
Xunfeng Li and Weimin Li

Robust Markov Random Field Model for Image Segmentation 205
 Taisong Xiong and Yuanyuan Huang

Community Clustering Based on Weighted Informative Graph 215
 Yi Xu, Yingning Gao and Weimin Li

A Data Clustering Algorithm Using Cuckoo Search 225
 Mingru Zhao, Hengliang Tang, Jian Guo and Yuan Sun

The Application of Bacteria Swarm Optimization Algorithm in Site Choice of Logistics Center 231
 Mingru Zhao, Hengliang Tang, Jian Guo and Yuan Sun

SIDA: An Information Dispersal Based Encryption Algorithm 239
 Zhi-ting Yu, Quan Qian, Rui Zhang and Che-Lun Hung

Software Behavior Analysis Method Based on Behavior Template. 251
 Lai Yingxu, Zhao Yiwen and Ye Tao

Formalizing Dynamic Service Interaction Based on Pi-Calculus 261
 Yaya Liu, Jiulei Jiang and Wenwen Liu

Applications of Video Structured Description Technology for Traffic Violation Monitoring 271
 Qianjin Tang, Zheng Xu, Zhizong Wu, Yixuan Wu and Lin Mei

Research of Mining Multi-level Association Rule Models. 279
 Wen-Hsing Kao, Chin-Wen Lo, Kuo-Pin Li, Hsien-Wei Yang and Jeng-Chi Yan

Research of the Dimension Combination Strategy Model 289
 Bo-Shen Liou, Ruei-Yang Lin, Kuo-Pin Li, Wen-Hsing Kao and Jeng-Chi Yang

Short Latency Bias in Latency Matrix Completion 301
 Cong Wang, Min LI and Yan Yang

Facial Feature Extraction Based on Weighted ALW and Pulse-Coupled Neural Network 311
 Junhua Liang, Zhisheng Zhao, Xiao Zhang, Yang Liu and Xuan Wang

Event Representation and Reasoning Based on SROIQ and Event Elements Projection 325
 Wei Liu, Ning Ding, Yue Tan, Yujia Zhang and Zongtian Liu

The Research and Application of Data Warehouse’s Model Design 337
Zhangzhi Zhao, Jing Li, Yongfei Ye, Yang Liu and Yaxu Liu

Question Recognition Based on Subject 353
Li-fang Huo, Li-ming Zhang and Xi-qing Zhao

Development of a Mobile Augmented Reality System to Facilitate Real-World Learning 363
Kai-Yi Chin, Ko-Fong Lee and Hsiang-Chin Hsieh

A Simple Randomized Algorithm for Complete Target Coverage Problem in Sensor Wireless Networks. 373
Weizhong Luo, Zhaoquan Cai and Zhi Zeng

A Novel Enveloped-Form Feature Extraction Technique for Heart Murmur Classification 379
HaoDong Yao, BinBin Fu, MingChui Dong and Mang I. Vai

Research on Network Security Strategy Model. 389
Anyi Lan, Bo Li, Rongsheng Huang, Xiao Zhang and Guiliang Feng

Investigating on Radioactivity of LBE and Pb in ADS Spallation Target 395
Yaling Zhang, Xuesong Yan, Xunchao Zhang, Jianqi Chen, Qingguo Zhou and Lei Yang

Design of Farmland Environment Remote Monitoring System Based on ZigBee Wireless Sensor Network 405
Yongfei Ye, Li Hao, Minghe Liu, Hongxi Wu, Xiao Zhang and Zhisheng Zhao

Attractions and Monuments Touring System Based on Cloud Computing and Augmented Reality. 417
Deqiang Han, Zongxia Wang and Qiang Zhang

Constructing Weighted Gene Correlation Network on GPUs. 429
Guanghui Yang, Sheng Zhang, Yuan Tian, Ping Lin, Jiang-Feng Wan, Qingguo Zhou and Lei Yang

Design of Scalable Control Plane via Multiple Controllers 441
Wenbo Chen, Xining Tian and Zhihao Shang

Research on Learning Record Tracking System Based on Experience API 451
Xinghua Sun, Yongfei Ye, Li Hao, Zexin An and Xiaoyu Wang

An EF6 Code-First Approach Using MVC Architecture Pattern for Watershed Data Download, Visualization and Analysis System Development Based on CUAHSI-HIS 459
 Rui Gao, Yanyun Nian, Lu Chen and Qingguo Zhou

Student-t Mixture Modelling for Image Segmentation with Markov Random Field. 471
 Taisong Xiong, Yuanyuan Huang and Xin Luo

Integrated Genetic Algorithm and Fuzzy Logic for Planning Path of Mobile Robots. 481
 Shixuan Yao, Xiangrong Wang and Baoliang Li

Characterization of Noise Contaminations in Realistic Heart Sound Acquisition 491
 Jun Huang, Booma Devi Sekar, Ran Guo, MingChui Dong and XiangYang Hu

Independent Component Analysis of Space-Time Patterns of Groundwater System 503
 Chin Tsai Hsiao, Jui Pin Tsai and Yu Wen Chen

Analysis of the Status Quo of MOOCs in China. 515
 Li Hao, Xinghua Sun, Chunlei Zhang and Xifeng Guo

Detection for Different Type Botnets Using Feature Subset Selection 523
 Kuan-Cheng Lin, Wei-Chiang Li and Jason C. Hung

Rotation Invariant Feature Extracting of Seal Images Based on PCNN 531
 Naidi Liu, Yongfei Ye, Xinghua Sun, Junhua Liang and Peng Sun

The Taguchi System-Two Steps Optimal Algorithm Based Neural Network for Dynamic Sensor Product Design. 541
 Ching-Lien Huang, Yung-Hui Chen, Chun-Hsiung Tseng, Tian-Long John Wan, Lung-Cheng Wang and Chang-Lin Yang

Accurate Analysis of a Movie Recommendation Service with Linked Data on Hadoop and Mahout. 551
 Meng-Yen Hsieh, Gui-Lin Li, Ming-Hong Liao, Wen-Kuang Chou and Kuan-Ching Li

A Method of Event Ontology Mapping 561
 Xu Wang, Wei Liu, Yujia Zhang, Yue Tan and Feijing Liu

A Research on Multi-dimensional Multi-attribute String Matching Mechanism for 3D Motion Databases 575
 Edgar Chia-Han Lin

An Novel Web Service Clustering Approach for Linked Social Service 583
Wuhui Chen, Banage T.G.S. Kumara, Takazumi Tanaka, Incheon Paik and Zhenni Li

Cloud Computing Adoption Decision Modelling for SMEs: From the PAPRIKA Perspective 597
Salim Alismaili, Mengxiang Li and Jun Shen

Cost Analysis Between Statins and Hepatocellular Carcinoma by Using Data Mining Approach 617
Yu-Tse Tsan, Yu-Wei Chan, Wei-Chen Chan and Chin-Hung Lin

Hospital Service Queue Management System with Wireless Approach 627
Manoon Ngorsed and Poonphon Suesaowaluk

A Smartphone Based Hand-Held Indoor Positioning System 639
Lingxiang Zheng, Zongheng Wu, Wencheng Zhou, Shaolin Weng and Huiru Zheng

A Variational Bayesian Approach for Unsupervised Clustering. 651
Mu-Song Chen, Hsuan-Fu Wang, Chi-Pan Hwang, Tze-Yee Ho and Chan-Hsiang Hung

Virtualized Multimedia Environment for Shoulder Pain Rehabilitation 661
Chih-Chen Chen, Hsuan-Fu Wang, Shih-Chuan Wang, Chih-Hong Chou, Heng-Chih Hsiao and Yu-Luen Chen

Multimedia Technology with Tracking Function for Hand Rehabilitation 671
Ying-Ying Shih, Yen-Chen Li, Chih-Chen Chen, Hsuan-Fu Wang, Shih-Wei Chou, Sung-Pin Hsu and Yu-Luen Chen

LBS with University Campus Navigation System 681
Jiun-Ting Chen and Ya-Chen Chang

An Efficient Energy Deployment Scheme of Sensor Node 689
Cheng-Chih Yang, Hsuan-Fu Wang and Yung-Fa Huang

Channel Equalization for MIMO LTE System in Multi-path Fading Channels 697
Hsuan-Fu Wang, Mu-Song Chen, Ching-Huang Lin and Chi-Pan Hwang

All-Digital High-Speed Wide-Range Binary Detecting Pulsewidth Lock Loops 705
 Po-Hui Yang, Jing-Min Chen and Zi-Min Hong

A BUS Topology Temperature Sensor Cell Design with System in Package Application. 713
 Po-Hui Yang, Jing-Min Chen and Ching-Ken Chen

The Off-Axis Parabolic Mirror Optical Axis Adjustment Method in a Wedge Optical Plate Lateral Shearing Interferometer . . . 721
 Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee and Ya-Hui Hsieh

Two-Mirror Telescope Optical Axis Alignment by Additive Color Mixing Method 731
 Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee and Ya-Hui Hsieh

Design of Relay Lens Based on Zero Seidel Aberrations 743
 Kuang-Lung Huang, Yu-Wei Chan, Jin-Jia Chen and Te-Shu Liu

The Optical Spectra Analysis of 4 LED White-Light Sources Passing Through Different Fogs 753
 Chien-Sheng Huang, Ching-Huang Lin, Guan-Syuan Hong and Hsuan-Fu Wang

High Resolution Camera Lens Design for Tablet PC 761
 Ching-Huang Lin, Hsien-Chang Lin, Ta-Hsiung Cho, Hsuan-Fu Wang and Cheng-Chieh Tseng

Two-Wavelength Optical Microscope Optical Axis Adjustment by Five Incident Parallel Laser Beams. 773
 Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee and Wei-Hsin Chen

The Correlation Analysis Between the Non-contact Intraocular Pressure and Diopter. 783
 Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee and Ching-Chung Chen

Automated Tool Trajectory Planning for Spray Painting Robot of Free-Form Surfaces. 791
 Wei Chen and Yang Tang

The Research of Analysis Addiction of Online Game 801
 Jason C. Hung, Min-Hui Ding, Wen-Hsing Kao, Hui-Qian Chen, Guey-Shya Chen and Min-Feng Lee

**Parameter Estimation of Trailing Suction Hopper
Dredger Dredging Model by GA 811**
Zhen Su and Wei Yuan

CPP Control System Design of Ship Based on Siemens PLC 817
Liang Qi and Shengjian Huang

**The Surface Deformation Prediction of Ship-Hull Plate
for Line Heating 827**
Liang Qi, Feng Yu, Junjie Song and Xian Zhao

**The Framework Research of the Internet of Things
in Dispatching Emergency Supplies 841**
Tongjuan Liu, Yanlin Duan and Yingqi Liu

**Simulation and Optimization of the AS/RS Based
on Flexsim 855**
Tongjuan Liu, Yanlin Duan and Yingqi Liu

**Design and Experiment of Control System for Underwater Ocean
Engineering Structure Inspection and Cleaning
Remotely Operated Vehicle 865**
Haijian Liu, Zhenwen Song, Song Liang, Lu Chang, Renyi Lin,
Wei Chen and Qingjun Zeng

**Using Experiment on Social Learning Environment
Based on an Open Source Social Platform 881**
Jing-De Weng, Martin M. Weng, Chun-Hong Huang
and Jason C. Hung

**User Authentication Mechanism on Wireless Medical
Sensor Networks 887**
Wei-Chen Wu and Horng-Twu Liaw

**Application of Cloud Computing for Emergency Medical
Services: A Study of Spatial Analysis and Data Mining
Technology 899**
Jui-Hung Kao, Feipei Lai, Bo-Cheng Lin, Wei-Zen Sun,
Kuan-Wu Chang and Ta-Chien Chan

**Social Event Detection and Analysis Using Social
Event Radar 917**
Jin-Gu Pan and Ping-I Chen

**Social Network and Consumer Behavior Analysis:
A Case Study in the Retail Store 927**
Pin-Liang Chen, Ping-Che Yang and Tsun Ku

Novel Scheme for the Distribution of Flyers Using a Real Movement Model for DTNs 937
 Tzu-Chieh Tsai and Ho-Hsiang Chan

A Study of Two-Dimensional Normal Class Grouping 949
 Ruey-Gang Lai and Cheng-Hsien Yu

Visualized Comparison as a Correctness Indicator for Music Sight-Singing Learning Interface Evaluation—A Pitch Recognition Technology Study 959
 Yu Ting Huang and Chi Nung Chu

A Fuzzy Genetic Approach for Optimization of Online Auction Fraud Detection 965
 Cheng-Hsine Yu

A Study on the Use Intention of After School Teachers Using Interactive e-Learning Systems in Teaching 975
 Chih-Ching Ho and Horng-Twu Liaw

Bibliometric Analysis of Emerging Trends in High Frequency Trading Research 985
 Jerome Chih-Lung Chou, Mike Y.J. Lee and Chia-Liang Hung

Interactive Performance Using Wearable Devices: Technology and Innovative Applications 993
 Tzu-Chieh Tsai, Gon-Jong Su and Chung-Yu Cheng

Usability Evaluation of Acoustic-Oriented Services on Mouse Manipulation: Can Manipulation with Dual Senses Be Good? 1007
 Chi Nung Chu

Effect of We-Intention on Adoption of Information System Embedding Social Networking Technology: A Case of Cloud Drive 1017
 Jerome Chih-Lung Chou

Improving Project Risk Management of Cloud CRM Using DANP Approach 1023
 You-Shyang Chen, Chien-Ku Lin and Huan-Ming Chuang

Improving Project Risk Management by a Hybrid MCDM Model Combining DEMATEL with DANP and VIKOR Methods—An Example of Cloud CRM 1033
 Chien-Ku Lin, You-Shyang Chen and Huan-Ming Chuang

Using VIKOR to Improve E-Service Quality Performance in E-Store 1041
Chien-Ku Lin, You-Shyang Chen, Huan-Ming Chuang and Chyuan-Yuh Lin

Study on the Intellectual Capital and Firm Performance 1051
Chiung-Lin Chiu, You-Shyang Chen and Mei-Fang Yang

Voluntary Disclosure and Future Earnings 1059
Chiung-Lin Chiu and You-Shyang Chen

A Smart Design of Pre-processing Classifier for Impulse Noises on Digital Images 1065
Jieh-Ren Chang, Hong-Wun Lin and Huan-Chung Chen

An Effective Machine Learning Approach for Refining the Labels of Web Facial Images 1073
Jieh-Ren Chang and Hung-chi Juang

Using the Data-Service Framework to Design a Distributed Multi-Levels Computer Game for Insect Education 1085
Chih-Min Lo and Hsiu-Yen Hung

Financial Diagnosis System (FDS) for Food Industry Listed in the Taiwan Stock Exchange (TWSE) 1091
Cheng-Ming Chang

Classification Rule Discovery for Housing Purchase Life Cycle 1097
Bo-Han Wu and Sun-Jen Huang

Algorithms of AP+ Tree Operations for IoT System 1107
Qianjin Tang, Zhizong Wu, Yixuan Wu and Jinfeng Ma

Dynamic Storage Method of Big Data Based on Layered and Configurable Technology 1115
Wenjuan Liu, Shunxiang Zhang and Zheng Xu

MIC-Based Preconditioned Conjugate Gradient Method for Solving Large Sparse Linear Equations 1123
Zhiwei Tang, Hailang Huang, Hong Jiang and Bin Li

Modeling and Assessing the Helpfulness of Chinese Online Reviews Based on Writing Behavior 1131
Chenglei Qin, Xiao Wei, Li Xue and Hongbing Cao

The Average Path Length of Association Link Network 1139
Shunxiang Zhang, Xiaosheng Wang and Zheng Xu

The Intelligent Big Data Analytics Framework for Surveillance Video System 1147
 Zheng Xu, Yang Liu, Zhenyu Li and Lin Mei

The Intelligent Video Processing Platform Using Video Structural Description Technology for the Highway Traffic. 1153
 Zheng Xu, Zhiguo Yan, Zhenyu Li and Lin Mei

The Scheme of the Cooperative Gun-Dome Face Image Acquisition in Surveillance Sensors 1159
 Zhiguo Yan, Zheng Xu, Huan Du and Lin Mei

Vehicle Color Recognition Based on CUDA Acceleration 1167
 Zhiwei Tang, Yong Chen, Bin Li and Liangyi Li

Video Retargeting for Intelligent Sensing of Surveillance Devices 1173
 Huan Du, Zheng Xu and Zhiguo Yan

Web Knowledge Acquisition Model Based on Human Cognitive Process 1179
 Xiaobo Yin and Xiangfeng Luo

An Investigation on the Relationship Among Employees' Job Stress, Satisfaction and Performance 1185
 Che-Chang Chang and Fang-Tzu Chen

Research on Influence Factors of the Formation of Virtual Innovation Clusters. 1193
 Dong Qiu and Qiu-Ming Wu

Research on the Development Path of New-Type R&D Organization in Guangdong Province, China 1201
 Li Huang

Analysis of Technology Diffusion Among Agricultural Industry Clusters by Game Theory 1209
 Chun-Hua Zheng and He-Liang Huang

Weakness of Zhang-Wang Scheme Without Using One-Way Hash Function 1217
 Zhi-Pan Wu

Weakness of an ElGamal-Like Cryptosystem for Enciphering Large Messages 1225
 Jie Fang, Chenglian Liu and Jieling Wu

Study of Kindergartner Work Pressure Based on Fuzzy Inference System. 1233
 Jie Fang

Comment on ‘The Hermite-Hadamard Inequality for R-Convex Functions’	1245
Zhi-Pan Wu	
Author Index	1249

Cloud and Crowd Based Learning

**Chun-Hsiung Tseng, Ching-Lien Huang, Yung-Hui Chen,
Chu-Chun Chuang, Han-Ci Syu, Yan-Ru Jiang, Fang-Chi Tsai,
Pin-Yu Su and Jun-Yan Chen**

Abstract Speaking of new teaching methodology, “Flipped Classroom” is undoubtedly a very popular one. The basic concept of flipped classroom is to have students learn by themselves before attending a “real” class at school. Once the background learning stage is performed outside of the class time, tutors have free time to lead students to participate in higher-order thinking. However, as shown in the report of Katie Ash, the performance of the flipped classroom method is in fact still arguable. Our survey shows that the contents offered by most modern e-learning systems are relatively static. Consider how fast new information appeared on the Web! Of course, teachers, or material providers, can upload new contents to e-learning systems. However, creating contents requires efforts. Today, work load of our teachers is already heavy, so expecting teachers to update contents very frequently is not practical. The researchers believe that one of the major challenges faced by e-learning systems today is the richness of contents.

Keywords e-learning · Crowd-sourcing

C.-H. Tseng · C.-C. Chuang · H.-C. Syu · Y.-R. Jiang · F.-C. Tsai · P.-Y. Su · J.-Y. Chen
Department of Information Management,
Nanhua University, Dalin Township, Taiwan, ROC
e-mail: lendle_tseng@seed.net.tw

C.-L. Huang
Department of Industrial Management,
Lunghwa University of Science and Technology, Taoyuan 33306, Taiwan, ROC
e-mail: lynne@mail.lhu.edu.tw

Y.-H. Chen (✉)
Department of Computer Information and Network Engineering,
Lunghwa University of Science and Technology, Taoyuan 33306, Taiwan, ROC
e-mail: cyh@mail.lhu.edu.tw

1 Introduction

Speaking of new teaching methodology, “Flipped Classroom” is undoubtedly a very popular one. The basic concept of flipped classroom is to have students learn by themselves before attending a “real” class at school. Once the background learning stage is performed outside of the class time, tutors have free time to lead students to participate in higher-order thinking. Typically, online materials such as videos and slides will be used in the background learning stage. To make the system effective, in most cases, the materials should be accessible from the Web. However, as shown in the report of Ash [1], the performance of the flipped classroom method is in fact still arguable. How is the quality of the materials? How active students are in the stage? How good is the performance evaluation method? Is there a reasonably-designed feedback system? Do practical tools provided for coaches? These are all important factors affecting whether a flipped classroom system is successful or not. Our survey shows that the contents offered by most modern e-learning systems are relatively static. Consider how fast new information appeared on the Web! Of course, teachers, or material providers, can upload new contents to e-learning systems. However, creating contents requires efforts. Today, work load of our teachers is already heavy, so expecting teachers to update contents very frequently is not practical.

The challenges pointed out above are not freshly new in the Web 2.0 age. We have an explosive amount of information. It is beyond the capabilities of traditional material providers to always keep their material up-to-date. At the very beginning of the Web 2.0 age, the issue is solved with crowd sourcing. That is, instead of relying on few material providers to update their Web sites or their blogs, we simply allow everyone to become material providers. However, crowd sourcing is only a partial cure. There are already a few on-line knowledge-sharing Web sites utilizing crowd sourcing technologies. Among them, “Yahoo Answers” is a well-known example. To the best of the researchers’ memory, there is no strict study for this, however, generally speaking, “Yahoo Answers” is not treated as an e-learning Web site. A possible reason is, materials on “Yahoo Answers” are too diverse and not strictly-organized.

The situation motivates this research. Looking at the enormous amount of information on the Web, the researchers wonder how to leverage the information in e-learning systems. Simply automatically collecting “similar” contents together will make little contribution to learning due to the diversity of the Web. In this research, the goal is to propose a module that can facilitate the following functionalities:

1. be automatic
2. generate structured information
3. take advantage of the crowd sourcing technology
4. adopt the cloud technologies

2 Related Works

Although Web search engines today are usually considered efficient, in some circumstances, they are not, especially when semantics and human intelligence are of concern. Some queries simply cannot be answered by machines alone. In such cases, human input is required [2]. The research field is typically named as crowd search or crowd searching. It is not an easy task to mediate between responses from human beings and search engines, and thus the research field is very challenging.

Crowd search is highly related with social networking [3]. The opinions collected within friends and expert/local communities can be ultimately helpful for the search task. For example, the question “find all images that satisfy a given set of properties” can be difficult for machines to proceed, but with the help of human intelligence, answering the question becomes simpler [4]. A special query interface that let users pose questions and explore results spanning over multiple sources was proposed in [5]. Another type of crowd search and crowd sourcing is social bookmarking. As shown in Heymann’s research work [6], social bookmarking is a recent phenomenon which has the potential to give us a great deal of data about pages on the web.

Various crowd search and crowd sourcing systems have been proposed. For example, the research of Parameswaran proposed a human intelligence-based methodology for solving the human-assisted graph search problem [7]. Amazon’s Amazon Mechanical Turk is a commercial product that enables computer programmers (known as Requesters) to co-ordinate the use of human intelligence to perform tasks that computers are currently unable to do [8].

3 Cloud and Crowd Based Learning Module

The proposed system is separated into several sub modules: the data processing sub module, the data provider sub module, and the feedback processing sub module. Figure 1 depicts the overview of the system.

The data processing sub module is responsible for collecting data from the Web and extracting information from them. Data collected from the Web is full of noises and is unstructured. Interpreting and reusing unstructured data is difficult. The sub module will extract structured data according to some pre-defined rules from it. The sub module is based on the researchers’ previous work: the Object-Oriented Schema Model (OOSM). The main issue to be addressed by this sub module is the unstructured nature of the Web. OOSM is in fact a grammar model. Here, we emphasize a database-centric design. Why is the concept of database important? The content of the Web is simply too diverse and ambiguous. To make information extracted from the Web usable for learning, we have to at first make it structured and thus it can be read and processed by applications easily. Databases are certainly

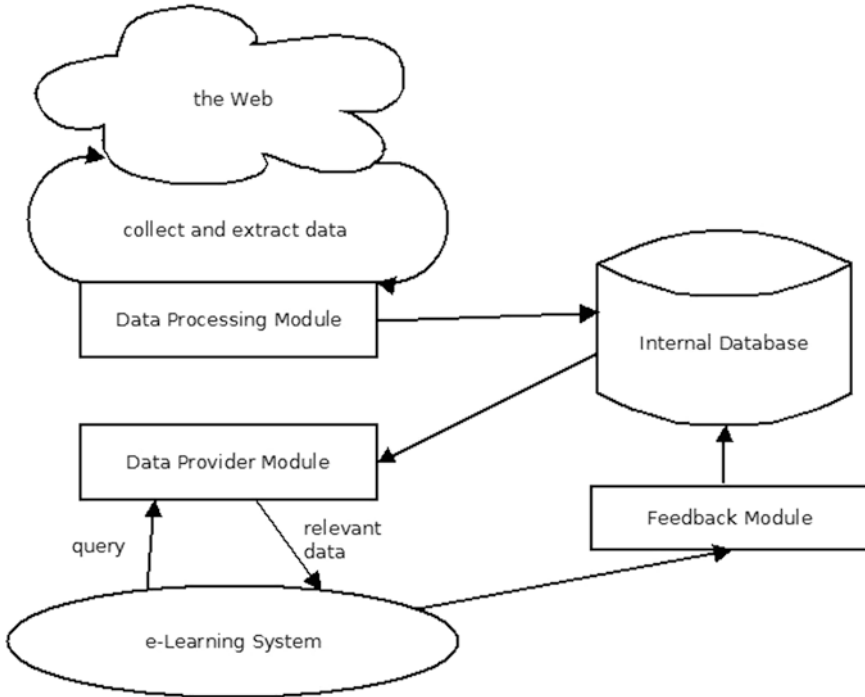


Fig. 1 System overview

structured information sources. OOSM contains three components: the schema model, the mapper tool, and the database.

The data provider sub module provides two sets of utilities: query utilities and transformation utilities. In most cases, e-learning systems interact with the data provider sub module to acquire materials related to their contents. For the purpose, the data provider sub module defines the following function:

$$query(db_namespace, db_localname, criteria, sortby)$$

Here, `db_namespace` and `db_localname` are used for pointing to a specific database defined in the data processing sub module. The data provider sub module allows the following types of criterias:

1. static id: each entry extracted by the data processing sub module will be automatically assigned an unique id; e-learning systems can utilize this id to make a static link
2. null: by specifying null in this field, the data provider sub module will not perform any filtering
3. attribute filter: a JSON-style query language will be developed for detailed querying

Furthermore, without a solid feedback system, it will be impossible to evaluate the effectiveness of the proposed system. The feedback sub module offers two types of feedbacks:

1. direct feedbacks: collected from users by requiring users to submit rating information directly
2. indirect feedbacks: collected by analyzing user behaviors such as the click sequences

For direct feedbacks, simple functions are defined:

$$\begin{aligned} &rate(domain, db_namespace, db_localname, contentId, score) \\ &tag(domain, db_namespace, db_localname, contentId, tags) \end{aligned}$$

As the names suggest, the functions are used for providing ratings and tags (labels) to content provided by a database. Scores give an overview of qualities of contents. By calculating average scores of contained contents, we can also have scores for databases. Tags reveal the characteristics of contents.

Indirect feedbacks are acquired through the analysis of click logs and tags. Click logs record user behavior when they are reading contents. How long did they stay on a specific content? How many contents are read before they leave? Which content is most frequently read after a user reads a specific content? Analyzing these tells us the popularity and potential (implicit) relationships of contents. On the other hand, tags represent users' subjective feeling about content. Roughly speaking, contents with the tags have something in common. Certainly, ambiguities can cause problems. An example is, without contextual information, we can never make sure of the correct meaning of the word "apple". Though completely getting rid of ambiguities is almost impossible, technologies such as word stemming can be used to alleviate the issue. Additionally, after collecting a proper amount of tags, grouping tags into clusters can help determine the similarities between tags and thus in turn reduce the level of ambiguities. Feedbacks are used for calculating scores of Web resources according to the following functions:

$$\begin{aligned} &score\ of\ content(C) : \alpha \times direct\ scores + \beta \times indirect\ scores, \alpha + \beta = 1 \\ &score\ of\ database(D) : \frac{\sum_{i=1}^n \rho_i C_i}{number_of_contents_in_the_database}, \rho = relative\ popularity \end{aligned}$$

4 Conclusions and Future Work

In this manuscript, the concept of a cloud and crowd based learning module is proposed. The proposed mechanism is aimed at solving the lack of content of existing e-learning systems. The proposed method utilizes crowd intelligence to collect related materials from the Web. Besides, to lower the complexities of

adopting the proposed method, we separate the concept into several sub modules and propose a cloud-based deployment structure. For now, the implementation is still in its prototyping stage. In the future, the following goals are set:

1. complete the implementation
2. integrate the implementation with an e-learning system for evaluation
3. develop a sound evaluation matrix

References

1. Ash K (2012) Educators evaluate ‘flipped classrooms’. <http://www.edweek.org/ew/articles/2012/08/29/02el-flipped.h32.html>
2. Franklin J, Kossmann D, Kraska T, Ramesh S, Xin R (2011) CrowdDB: answering queries with crowdsourcing. In: Proceedings of the 2011 ACM SIGMOD international conference on management of data, pp 61–72
3. Bozzon A, Brambilla M, Ceri S (2012) Answering search queries with CrowdSearcher. In: Proceedings of the 21st international conference on world wide web, pp 1009–1018
4. Parameswaran A, Garcia-Molina H, Park H, Polyzotis N, Ramesh A, Widom J (2012) CrowdScreen: algorithms for filtering data with humans. In: Proceedings of the 2012 ACM SIGMOD international conference on management of data, pp 361–372
5. Bozzon A, Brambilla M, Ceri S, Fraternali P (2010) Liquid query: multi-domain exploratory search on the web. In: Proceedings of the 19th international conference on world wide web, pp 161–170
6. Heymann P, Koutrika G, Garcia H (2008) Can social bookmarking improve web search? In: Proceedings of the 2008 international conference on web search and data mining, pp 195–206
7. Parameswaran A, Sarma A, Garcia-Molina H, Polyzotis N, Widom J (2011) Human-assisted graph search: it’s okay to ask questions. In: The proceedings of the VLDB endowment vol 4, pp 267–278
8. Amazon, <https://www.mturk.com/mturk/welcome>

Artificial Neural Network Based Evaluation Method of Urban Public Security

Zheng Xu, Qingyuan Zhou, Haiyan Chen and Fangfang Liu

Abstract In a Smarter City, available resources are harnessed safely, sustainably and efficiently to achieve positive, measurable economic and societal outcomes. Enabling City information as a utility, through a robust (expressive, dynamic, scalable) and (critically) a sustainable technology and socially synergistic ecosystem could drive significant benefits and opportunities. In this paper we propose a model based on Grid Management System. This model is based on grid cycle providing grid capturing, grid sharing, grid enhancing and grid preserving. Moreover, our model shares grid that supports the law of knowledge dynamics. Later we illustrate a scenario of Pudong District of Shanghai for independence issues. An Artificial Neural network (ANN) based simulation applying the proposed Grid Management System model is also described at the end of this paper to validate its applicability.

Keywords Artificial neural network · Public security, grid management system

Z. Xu (✉)

Tsinghua University, Beijing, China
e-mail: xuzheng@shu.edu.cn

Z. Xu

The Third Research Institute of Ministry of Public Security, Shanghai, China

Q. Zhou

Changzhou Party School of CPC, Changzhou, China

Q. Zhou

Changzhou Administrative College, Changzhou, China

F. Liu

Shanghai University, Shanghai, China

H. Chen

East China University of Political Science and Law, Shanghai, China

1 Introduction

In a Smarter City, available resources are harnessed safely, sustainably and efficiently to achieve positive, measurable economic and societal outcomes. Data (and then information) from people, systems and things in cities is the single most scalable resource available to City stakeholders but difficult to publish, organize, discover, interpret, combine, analyze, reason and consume, especially in such an heterogeneous environment [1–4]. Indeed data is big and exposed from heterogeneous environments such as water, energy, traffic or building. Most of the challenges of Big Data in Smart Cities are multi-dimensional and can be addressed from different multidisciplinary perspectives e.g., from Artificial Intelligence (Machine Learning, Semantic Web), Database, Data Mining to Distributed Systems communities. Enabling City information as a utility, through a robust (expressive, dynamic, scalable) and (critically) a sustainable technology and socially synergistic ecosystem could drive significant benefits and opportunities. While research efforts in Big Data have mostly focused on the later stages of the process of making sense of the sea of data (e.g. data analytics, query answering, data visualization, etc.), in the context of Smart Cities, where heterogeneous data originates from multiple municipal and state agencies with little to no coordination, major hurdles and issues continue to impede progress toward these later stages. These key unaddressed issues are often related to information exploration, access, and linking.

Recent research and experiments suggest that artificial neural network (ANN) can be a candidate for nonlinear series forecasting [5]. ANN is typical intelligent learning paradigm, widely used in some practical application domains including: pattern classification, function approximation, optimization, forecasting and many others [6]. Opposed to traditional forecasting approaches, ANN has a strong self-learning and self-organizing ability so it can tackle any nonlinear problem. As a classic method of ANN, BP neural network model is widely used in forecasting area. Using neural networks has the limitations of large complexity and also fails because of over-fitting, local optima. On the other hand, RBFNNs, with only one hidden layer, have the ability to find global optima. In addition to less computational complexity, simulations performed in the literature reveal that the RBFNN produces superior performance as compared to other existing ANN-based approaches. Hence the works on task scheduling using RBFNN became an established and an active area of academic research and development [7]. In this paper we propose an E-Governance model based on Grid Management System. This model is based on grid cycle providing grid capturing, grid sharing, grid enhancing and grid preserving. Moreover, our model shares grid that supports the law of knowledge dynamics. Later we illustrate a scenario of Pudong District of Shanghai for independence issues. An Artificial Neural network (ANN) based simulation applying the proposed Grid Management System model is also described at the end of this paper to validate its applicability.

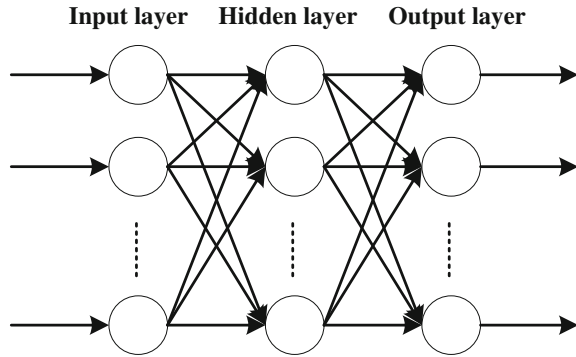
2 Related Work

Data Grids primarily deal with data repositories, sharing, access and management of large amounts of distributed data. Many scientific and engineering applications require access to large amounts of distributed data; however, different data could have their own format. In such Grid systems many types of algorithm, such as replication, are important to increase the performance of Grid-enabled applications that use large amounts of data. Also, data copy and transfer is important here in order to achieve high throughput. To successfully realize the vision of scientific Grid applications, the commission of the Next Generation Grid (NGG) [8] recognized the challenging research topics in future Grid systems including guaranteed QoS, reliability, and service performance, which are of vital importance in the Grid-computing service. Nevertheless, most commercial vendors in reality endeavor to increase commercial interests by efficient Grid-computing service and cost-effective Grid trade, but if the Grid system were utterly driven by commercial interests rather than QoS, a commercial buyer would not embrace the Grid-computing service to deal with critical computing jobs. Therefore, a number of RMS concerning service efficiency, service cost, or the requirements of QoS have been recommended in various complex Grid environments [9]. Grid computing was conceived to connote the idea of a “power grid:” namely, applications can plug into the Grid to draw computing resources in the same way electrical devices plug into a power grid to draw power. Analogous to a power grid, it views geographically distributed computing capabilities, storage, data sets, scientific instruments, knowledge, and so on as utility resources to be delivered over the Internet seamlessly, transparently, and dynamically as and when needed. The Grid is built upon two fundamental concepts: virtualization, i.e., individuals and/or institutions with the required resources or common interests can dynamically form a virtual organization (VO) that enables rapid assembly and disassembly of resources into transient confederations for coordinated problem solving, and dynamic provisioning, i.e., resources provision is transient, dynamic, and volatile without guarantee of availability, central control for accessibility, and prior trust relationships. Grid computing offers a promising distributed computing infrastructure where large-scale cross organizational resource sharing and routine interactions are commonplace.

3 Basic Methodology

The Back Propagation (BP) neural network with self-adaptive and self-organizing characteristics can be very effective in dealing with nonlinear problems. Over the years, BP neural network model is widely applied in forecasting area. A BP neural

Fig. 1 The structure of BP neural network



network model comprises an input layer, one or more hidden layers and an output layer. Each layer comprises a number of nodes connected by weight-value. The BP network structure is shown in Fig. 1. BP network learning process consists of forward propagation and backward propagation. During the process of forward propagation, input samples are sent from the input layer to the hidden layer and finally to the output layer. The output results are produced after this process. Then turn to the back propagation stage if there is a big difference between output results and expected results. In the back propagation, output error is reversed back to input layer, by modifying connection weights between neurons of each layer. These two propagations repeat iteratively to adjust connection weights and node biases in order to eventually minimize the error function. It's known that BP neural network is trained by Back Propagation (BP) algorithm [10].

RBF neural network is a kind of three-layer static feed-forward neural network consists of input layer, hidden layer and output layer. A typical RBF network structure is similar as Fig. 1 shows. The difference between RBF network and BP network is that it uses Gaussian function as the transfer function from the input layer to the hidden layer. Gaussian function is a local activation function and it is activated within a small extent so that the network has the local learning ability [11]. For the same problem, a RBF neural network requires more nodes in hidden layer but it has shorter training time and higher learning speed than a BP neural network.

Elman neural network is a regression neural network consists of four layers: input layer, hidden layer, undertake layer and output layer, as shown in Fig. 2. The input layer, hidden layer and output layer are similar as forward network. Its special feature is that the undertake layer has the ability to remember output value of hidden layer a time before and then use it as input value to hidden layer next time [12]. This type of network has a function of remembering dynamically so it can deal with dynamic information accurately.

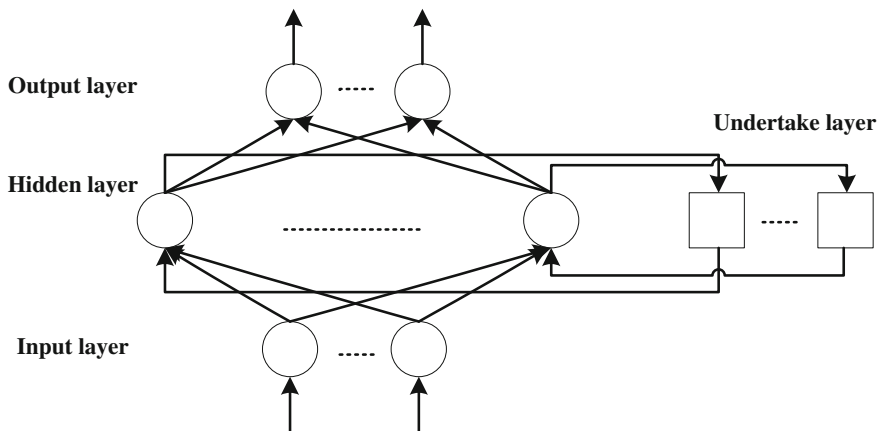


Fig. 2 The structure of Elman neural network

4 Urban Public Security Management Network Platform

Urban public security management network platform can be divided into three levels, the most above level is urban public security emergency response commanding center, the second level is management center, emergency preplan management center, database, information briefing center, and the third level is made up of safety and prevention system, firefighting control center, and a monitoring terminal of different danger sources from business and enterprises. Among them, the third level is the key of urban public security management network, safety and prevention systems and firefighting control centers of business and enterprises should make effective monitoring and management of various firefighting facilities under their protective area, and transmit their effective status to urban public security management center. In addition, monitoring terminals should be built to various dangerous sources, and transmit their status to the center.

In this way, the status of dangerous sources can be enquired from this center, and if the dangerous sources are in a wrong or non-regular state, evaluations shall be made and conclusions and corrective measures against can be made and sent to respective management departments. If corrections have not been made within validity, then the grade of danger shall be raised, and the evaluation report shall be delivered to emergency response commanding center. The safety and prevention system, firefighting control center, and a monitoring terminal of different danger sources shall make regular self-inspection and maintenance, and send reports to the public security database for central processing. By the effective management of third level information terminal, various dangerous sources can be investigated and made effective monitoring and control, thus a solid groundwork of urban public security management system can be laid eventually.

5 Conclusions

In a Smarter City, available resources are harnessed safely, sustainably and efficiently to achieve positive, measurable economic and societal outcomes. Enabling City information as a utility, through a robust (expressive, dynamic, scalable) and (critically) a sustainable technology and socially synergistic ecosystem could drive significant benefits and opportunities. In this paper we propose a model based on Grid Management System. This model is based on grid cycle providing grid capturing, grid sharing, grid enhancing and grid preserving. More-over, our model shares grid that supports the law of knowledge dynamics. Later we illustrate a scenario of Pudong District of Shanghai for in-dependence issues. An Artificial Neural network (ANN) based simulation applying the proposed Grid Management System model is also de-scribed at the end of this paper to validate its applicability.

Acknowledgments This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, 2013AA014603, in part by National Key Technology Support Program under Grant 2012BAH07B01, in part by the National Science Foundation of China under Grant 61300202, 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009, in part by the China Postdoctoral Science Foundation under Grant 2014M560085, the project of Shanghai Municipal Commission of Economy and Information under Grant 12GA-19, and in part by the Science Foundation of Shanghai under Grant 13ZR1452900, 12ZR1411000.

References

1. Wang L, Ke L, Liu P (2015) Compressed sensing of a remote sensing image based on the priors of the reference image. *IEEE Geosci Remote Sensing Lett* 12(4):736–740
2. Liu P, Yuan T, Ma Y, Wang L, Liu D, Yue S, Kolodziej J (2014) Parallel processing of massive remote sensing images in a GPU architecture. *Comput Inf* 33(1):197–217
3. Wang L, Ke L, Liu P, Ranjan R, Chen L (2014) IK-SVD: dictionary learning for spatial big data via incremental atom update. *Comput Sci Eng* 16(4):41–52
4. Wang L, Geng H, Liu P, Ke L, Kolodziej J, Ranjan R, Zomaya AY (2015) Particle swarm optimization based dictionary learning for remote sensing big data. *Knowl-Based Syst* 79:43–50
5. Crone SF, Hibon M (2011) Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *Int J Forecast* 27(3):635–660
6. Zhang G, Patuwo BE (1998) Forecasting with artificial neural networks: the state of the art. *Int J Forecast* 14:35–62
7. Lu XL, Lu XQ, Liang XH (2009) Neural network and its application to prediction of nonlinear time sequence. *Syst Eng Theory Appl* 17(4)
8. Patel P, Marwala T (2006) Forecasting closing price indices using neural networks. In: *Proceedings of IEEE international conference on systems, man and cybernetics, 2006*. SMC 06, pp 2351–2356
9. Blank SC (1991) Chaos in futures markets? A nonlinear dynamical analysis. *J Futures Markets* 11(6):711–728
10. Yu L, Wang S (2008) Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ* 30(5):2623–2635

11. Haiming Z, Xiaoxiao S (2013) Study on prediction of atmospheric PM_{2.5} based on RBF neural network. In: Proceedings of IEEE fourth international conference on digital manufacturing and automation (ICDMA), pp 1287–1289
12. Yongchun L (2010) Application of Elman neural network in short-term load forecasting. In: International conference on artificial intelligence and computational intelligence (AICI), vol 2. IEEE, pp 141–144

Building the Search Pattern of Social Media User Based on Cyber Individual Model

Zheng Xu, Xiao Wei, Dongmin Chen, Haiyan Chen and Fangfang Liu

Abstract As the Web enters Big Data age, users and search engines may find it more and more difficult to effectively use and manage such big data. On one hand, people expect to get more accurate information with less search steps. On the other hand, search engines are expected to incur fewer resources of computing, storage and network, while serving the users more effectively. After more and more personal data becomes available, the basic issue is how to generate Cyber-I's initial models and make the models growable. The ultimate goal is for the growing models to successively approach to or become more similar as individual's actual characteristics along with increasing personal data from various sources covering different aspects. In this paper, we propose the concept of search pattern, summarize search engines into three search patterns and compare them in order to seek the more efficient one. We propose a new search pattern termed as ExNa, which can be incorporated into search engines to support more efficient search with better results.

Keywords Search pattern · Social media · Cyber individual model

Z. Xu (✉)

Tsinghua University, Beijing, China
e-mail: xuzheng@shu.edu.cn

Z. Xu

The Third Research Institute of Ministry of Public Security, Shanghai, China

X. Wei

Shanghai Institute of Technology, Shanghai, China

D. Chen

Software College of Northeastern University, Shenyang, China

F. Liu

Shanghai University, Shanghai, China

H. Chen

East China University of Political Science and Law, Shanghai, China

1 Introduction

As the Web enters Big Data age, users and search engines may find it more and more difficult to effectively use and manage such big data. On one hand, people expect to get more accurate information with less search steps. On the other hand, search engines are expected to incur fewer resources of computing, storage and network, while serving the users more effectively. New types of search engines are emerging to solve the problem. In particular, faceted search [1, 2], ontology-based search [3], concept-based search [4], and rule-based search [5] all aim to improve search engines in some aspects and contribute to the development of so-called “next generation” search engines (NGSEs).

With rapid advances of computing and communication technologies, we are stepping into a completely new cyber-physical integrated hyper world with digital explosions of data, connectivity, services and intelligence. As individuals facing so many services in the digitally explosive world, we may not be aware of what are the most necessary or suitable things [6–9]. Hence, the appearance of Cyber-I, short for Cyber-Individual, is a counterpart of a real individual (Real-I) to digitally clone every person [10, 11]. The study on Cyber-I is an effort to re-examine and analyze human essence in the cyber-physical integrated world in order to assist the individuals in dealing with the service explosions for having an enjoyable life in the emerging hyper world.

After more and more personal data becomes available, the basic issue is how to generate Cyber-I’s initial models and make the models growable. The ultimate goal is for the growing models to successively approach to or become more similar as individual’s actual characteristics along with increasing personal data from various sources covering different aspects. The focus of this research is on the initialization and growth of Cyber-I’s models. The initial models are generated based on the personal data acquired in a Cyber-I’s birth stage, while the growing models are built with the personal data continuously collected after the birth. We proposed three mechanisms for Cyber-I modeling to enable the models growing bigger, higher and closer successively to its Real-I.

A big question here is then that in order to achieve NGSEs, what types of search patterns should NGSEs support. With an aim to help find a possible answer to this big question, in this paper we adopt an inside-out approach by first defining *Search Pattern (SP)* as the combination of index structure, user profiles, and interaction mechanism, which can describe the features related to the search process more comprehensively, including those of NGSEs. Then, we summarize current search engines into three types of search patterns. By comparing and analyzing different patterns, we try to identify what features a “next generation” search engine (NGSE) should have and what search patterns NGSEs should support. Based on this, we propose a new search pattern named ExNa by defining its model and basic operations. To validate the newly proposed ExNa search pattern, we conduct experimental studies upon a semantic search engine named NEWSEARCH, and the results show that KNOWLE equipped with ExNa can improve the holistic

efficiency of the search system. A search pattern may be good at a special aspect of a search engine, such as the precision of searching, the storage of index, the I/O, and so on. ExNa is good at the holistic efficiency when compared with search engines of other search patterns.

In this paper, we propose the concept of search pattern, summarize search engines into three search patterns and compare them in order to seek the more efficient one. We propose a new search pattern termed as ExNa, which can be incorporated into search engines to support more efficient search with better results.

2 Related Work

User models are also known as user profiles, personas or archetypes. They can be used by designers and developers for personalization purposes so as to increase the usability and accessibility of products and services. With the development of personalized systems, like e-learning systems, a lot of personal data can be collected. In order to find some personal features to give appropriate advices or recommendations, the user model should be established in service systems. However, many such kind of user models is application-specific or service-specific which cannot be used by other applications/services. To overcome this barrier of the user models between different applications, a generic user model system (GUMS) was proposed to support interoperability among different user modeling systems [12]. The GUMS is able to exchange contents of user models, and use the exchanged user's information to enrich the user experience. Life logging is utilized to automatically record user's life events in digital format. With continuously capturing contextual information from a user and the user's environment, personal data increases fast and becomes huge. The most of lifelog systems are putting more emphases on personal data collection, storage and management [13]. Lifelong user modeling is trying to provide users such models accompanied with users' whole life [14]. This idea or vision is attractive, but no general mechanism has been made and no practical system has been built yet. Lifelong machine learning (LML), received great attention in recent years, is to enable an algorithm or a system to learn tasks from more domains over its lifetime [15].

3 The Search Pattern

ExNa is not a simple integration of the Narrow SP and the Expand SP. ExNa is expected to have a free styled interaction, a more efficient index structure which should be rich semantics, less storage, and abundant interaction paths, and a flexible user profile to support all kinds of service. Some conflicts should be resolved in ExNa, such as the conflict between the rich semantics and the huge storage.

And some problems should be solved in ExNa too, such as how to realize the free styled interaction, and how to build a flexible user profile to support all kinds of services.

Although ExNa is a little like the integration of Narrow SP and Expand SP, it just includes the interaction paths of Narrow SP and Expand SP. As the definition of Search Pattern shown, SP consists of three parts and the search path is only the representation of the entire SP.

Based on the discussions of Linear SP, Narrow SP and Expand SP, we compares them as per the structure of index, the storage of index, the semantics of index, the interactive mechanism and user profiles. We strive to find a new search pattern by integrating the advantages of the current SPs as many as possible. Clearly, Narrow SP and Expand SP work in vertical and horizontal directions, respectively. Narrow SP may rapidly narrow the search scope with the support of hierarchical index structure. Expand SP may expand the search based on some semantic relations, thereby facilitating user search with fuzzy terms. Take both vertical and horizontal directions into account, the index of the new SP should be a structure of multi-layered, in which different layers denote the indices of different granularities. Besides, the web resources of the same layer should be organized as a semantic link network. We name the index structure as the multi-layered semantic link network index structure. Rich semantics should be included in the new index structure to support efficient search. The semantic relations between layers support the narrow search. The semantic link network may hold several kinds of semantic relations to support the expand search. Storing rich semantic information needs more storage than the inverted index. The multi-layered and community structure in semantic link network may reduce the storage of index to a large extent. We expect the storage space to be at the medium level which is much less than a single layer network structure such as Expand SP. With the support of the multi-layered semantic link network index structure, a user may interact with the index from both vertical and horizontal directions, which form a free-styled interaction mechanism. To support the free-styled interaction, the proper structure of user profiles should be a multi-layered network too, so as to record user.

4 The Basic Data for Social Media Profile

SNS profile. The online social networking service (SNS), like Facebook.com, is a great way to find out more about you, which allows anyone with an email address to create a profile complete with pictures and a variety of specific personal information. Personal information is voluntarily supplied by the user and usually contains information such as Major, Hometown, Relationship, Status, Political Views, Interests, Favorite Music/Movies/Books/Quotes, and an “About Me” section which contains a short description of the user someone you have just met. The SNS profile play an important role during the initialization of Cyber-I modeling since it contains some context information that is able to be utilized. For instance, taking a user’s age, occupation or hometown into consideration will better locate the user or give

the user a better service or more applications will be added in order to generate more personal data.

Preference Choice. It has long period study/research in the area of psychology, and psychological research give us proof that the recognition of user preferences could reflect something deep inside the user, such as the characteristics, the trait. And such preference could also lead to influence the selection and instantiation of the action that achieve the user's target. In this thesis, we start from the simple color preference, which may not be sensitive for someone's privacy concern and generally speaking, everyone has his/her own loved color. Color preference is an important aspect of visual experience that influences a wide spectrum of human behaviors. Secondly, we suggest the user to choose the other optional preference choices, which are available as Foods, Sport, Movie and Music. If users are willing to choose those (we are strongly suggest to do this), the model can get and know your properties of different aspects in order to generate a better initial model for you and provide you more services/apps. The function of preference choices will be talked in detail in the next section.

Browsing History, App Usage and Activity Tag. In order to fetch the information concerning the user's activity on PC, we make use of the software "Manic Time" to implement those functions, which could generate the data into the different CSV files. The files can be uploaded manually into the database and can be processed by our Java program in processor database. We calculate the total time and the times you open one software during your working on PC. Meanwhile, the frequently visited website can also be analyzed through this Java program. After analyzing the CSV files, the consequences can be demonstrated on the form of pie chart or bar graph based on the Google Chart Visualization API. What's more, we can generate further results, such as the top 3 favorite websites, what application or even what kinds of information are preferred. For the professional like employees and students who are in front of computer every day, the activity tags like "go for lunch" "afternoon nap", "time for dinner" can also be demonstrated in the results and able to be stored into database for modeling.

Movement log. In order to collect the movement log of the number of steps of the day, UP of Jawbone Company, which is a wearable activity recording device, can be used. Further, it is possible to synchronize and visualize the data at any time measured by using the UP smartphone application. In addition, since it measurably every day, logging your exercise conditions on an ongoing basis. UP is possible to use about 1 week on a single charge which is designed that user can wear everyday with waterproof function and with just 22 g weight body in wristband. In this research, the number of steps was collected and through synchronizing with smartphone to transfer data to the server of JAWBONE, steps and exercise situation and the consumption of calories each date can be stored. Further, it is possible to access the home page of JAWBONE, obtaining CSV format data in the account page when have been registered. Analysis is performed to get the number of steps for knowing the motion state through the data obtained from the UP in our present study.

5 Conclusions

To overcome the shortcomings of traditional search engines, many new search engines are designed based on some new technologies. Each search engine has its advantages and disadvantages. So we try to summarize the current search engines to find the features of search engines of next generation. In this paper, we propose the concept “Search Pattern” to describe the most important features of search engines. We classify current search engines into three Search Patterns: Linear Search Pattern, Narrow Search Pattern, and Expand Search Pattern. We present a novel search pattern ExNa based on the comparison of Linear Search Pattern, Narrow Search Pattern, and Expand Search Pattern. Then, we model ExNa and definition its basic operations to help developing search engines of next generation.

Acknowledgments This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, 2013AA014603, in part by National Key Technology Support Program under Grant 2012BAH07B01, in part by the National Science Foundation of China under Grant 61300202, 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009, in part by the China Postdoctoral Science Foundation under Grant 2014M560085, the project of Shanghai Municipal Commission of Economy and Information under Grant 12GA-19, and in part by the Science Foundation of Shanghai under Grant 13ZR1452900, 12ZR1411000.

References

1. ICSTI Insight: Next generation search. http://www.icsti.org/IMG/pdf/insight_2010_july.pdf
2. Tunkelang D (2009) Faceted search. In: Synthesis lectures on information concepts, retrieval, and services, vol. 1(1), pp 1–80
3. Alisi T, Bertini M, D’Amico G, Del Bimbo A, Ferracani A, Pernici F, Serra, G (2009) Sirio: an ontology-based web search engine for videos. In: Proceedings of the 17th ACM international conference on multimedia, pp 967–968
4. Shehata S, Karray F, Kamel M (2007) Enhancing search engine quality using concept-based text retrieval. In: Proceedings of IEEE/WIC/ACM international conference on web intelligence, pp 26–32
5. Pilz T, Luther W, Ammon U, Fuhr N (2006) Rule-based search in text databases with nonstandard orthographym. *Literary Linguist Comput* 21(2):179–186
6. Wang L, Ke L, Liu P (2015) Compressed sensing of a remote sensing image based on the priors of the reference image. *IEEE Geosci Remote Sens Lett* 12(4):736–740
7. Liu P, Yuan T, Ma Y, Wang L, Liu D, Yue S, Kolodziej J (2014) Parallel processing of massive remote sensing images in a GPU archi-tecture. *Comput Inform* 33(1):197–217
8. Wang L, Ke L, Liu P, Ranjan R, Chen L (2014) IK-SVD: dictionary learning for spatial big data via incremental atom update. *Comput Sci Eng* 16(4):41–52
9. Wang L, Geng H, Liu P, Ke L, Kolodziej J, Ranjan R, Zomaya AY (2015) Particle swarm optimization based dictionary learning for remote sensing big data. *Knowl-Based Syst* 79:43–50
10. Yen NY, Ma J, Huang R, Jin Q, Shih TK (2010) Shift to Cyber-I: reexamining personalized pervasive learning. In: Proceedings of the 3rd IEEE/ACM International conference on cyber. Physical and social computing, pp 685–690 December 2010

11. Wei J, Huang B, Ma J (2009) Cyber-I: vision of the individual's counterpart on cyberspace. In: Proceedings of the IEEE international conference on dependable, Autonomic and secure computing, pp 295–302
12. Levene M (2011) An introduction to search engines and web navigation. Wiley
13. Hu WC, Chen Y, Schmalz MS, Ritter GX (2001). An overview of the world wide web search technologies. In: Proceedings of the 5th world multi-conference on system, cybernetics and information, pp 22–25
14. Cutting D, Pedersen J (1989) Optimization for dynamic inverted index maintenance. In: Proceedings of the 13th annual international ACM SIGIR conference on research and development in information retrieval, pp 405–411
15. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM Comput Surv* 31 (3):264–323

Design of Health Supervision System Base on WBAN

Xinli Zhou

Abstract Traditional health care system in the family-oriented application of monitoring system has some disadvantage, which is relatively small, and function relatively single operability is more complex, real-time performance is poor, the price is relatively expensive. In recent years, with the progress of integrated circuit technology and wireless communication technology, wireless body area network (WBAN) systems have got fast development. The application system base on BAN technology also has received more and more people's attention. This paper presents a general framework BAN-based health care system, mainly introduces the design of the sensor from the perception layer and network protocol.

Keywords Wireless sensor · Wireless body area network (WBAN) · 802.15.6 · Health care and monitoring system

1 Introduction

Wireless body area network is the product of the rapid development and convergence of microelectronics technology and wireless communication technology. The principle of WBAN is, by various types of sensors perceiving from the body or body surface, collecting physiological signals and transmitting them to the local station, ultimately interact with information center and attain the monitoring and the purpose of medical treatment. Body area network [1] (wireless body area network, BAN or WBAN), also known as BSN (body sensor network), let the network extends to the human body, is an important part of internet of things.

The proposed BAN has been widespread concern on the medical community and the communications sector. IEEE has officially launched the standard of 2012 802.15.6 [2], the standard thought WBAN have a wide range of applications in

X. Zhou (✉)

School of Information Science and Technology, University of International Relation,
Beijing 100091, China
e-mail: Zhouxinli001@126.com

health care, emergency care, specific population monitoring tracking, entertainment and other fields. Chinese scholars have carried out theoretical and applied the study. Berkeley of University of California focuses on studies of wearable BAN, scalability and resource optimization. Based on a variety of communication methods the Chinese University of Hong Kong has built a mixed BAN, and study on BAN and mobile tracking and energy-aware MAC for the relevant research. Korea information and Communications University BSN build new systems from the perspective of energy consumption and communication. In addition, studies in Canada, Germany, Ireland, Brazil, Belgium and Switzerland and other countries in terms of BSN adaptability and adjustability, middleware, signal processing algorithms, health and activity monitoring and network reliability has also made progress [3]. Base on the previous research results, this paper presents a general framework BAN-based health care system, mainly introduces the design of the sensor from the perception layer and network protocol.

2 Systematic Design

Our health care systems base on WBAN combined with sensor technology, wireless communication technology, embedded technology. It can be collected physiological information using wearable wireless sensor nodes without affecting the daily activities to record the trajectory change of the physiological signals in one day, then to analyze such information and data, as well as to determine the occurrence of unforeseen circumstances inform physical condition, and sends the results to a remote medical service center or guardian.

Structural health monitoring system of BAN is generally considered to have three layers [4]: BAN internal, between BAN and aggregation node and the information monitoring center (Fig. 1).

The first layer internal network is composed of a group of sensors monitoring of physiological characteristics, due to limited resources, their functions experienced a simplified design. In the medical field, sensor is capable of measuring and

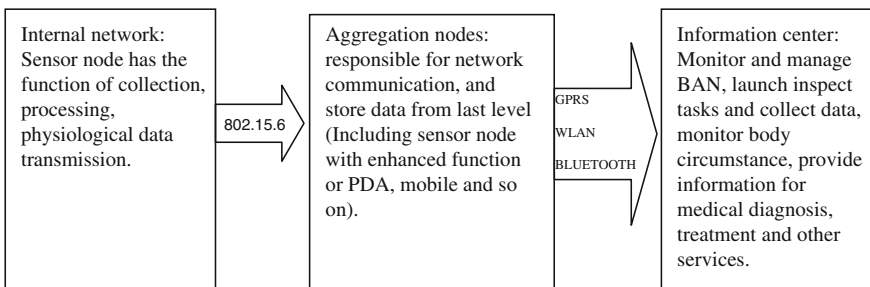


Fig. 1 Architecture of BAN network system

processing body's physiological signals or the environmental information, and transmitting the information to the outer control node. The sensor can also receive an external command to trigger action. In non-medical field, wearable devices (such as Headset, MP3 player and game controller) can include. In this paper, we analyze and design the various types of sensors used in the health monitoring system.

The second layer is a mobile personal server with fully functional design (mobile personal sever) or BSN Head or master nodes, and further includes a base station. It is responsible for the external communication and network, and stores the collected data from first layer. It manages each sensor node or device with low energy consumption, receives and analyzes sensing data and executes user program following provision. Here the base station can be mobile phone with relatively rich resources, PDA with accessable internet or other handheld devices.

The third layer is the data center contains various information such as electronic medical records which provide medical server maintains a registered user, some responding services for the user, medical and nursing staff.

3 Physical Channel Design

In the physical channel, we need to consider the wireless connection, antenna and power three parts.

3.1 Wireless Connection

In the earlier researchers realize the BAN with 802.15.4ZigBee, Wi-Fi or low power Bluetooth or other short-distance wireless communication technology. While considered the existence of large power consumption, easy interference and work in the ISM band of IEEE 802.11 and IEEE 802.15.4, so the NB in PHY layer are defined in the IEEE 802.15.6, ultra wideband (UWB) and human body communication (HBC) three physical layers. When working in different frequency range to meet the needs of different scenarios. Narrow band in the physical layer is mainly responsible for the activation/deactivation of wireless transceiver. In current channel NB can estimate the idle channel (CCA) and data sending and receiving. Narrow band (NB) signal is conformity with the energy level of MICS, and the interference is very low to other equipments. UWB physical layer has two working modes: the default mode and QoS mode. Human body communication (HBC) physical layer work in 4 MHz bandwidth, center frequency in two bands of 16 and 27 MHz, these two bands are available in American, Japan and South Korea, the working frequency in Europe is 27 MHz.

In this paper, the ultra wideband (UWB) the design of health monitoring system was selected. It's default mode is suitable for medical use, QOS model is more suitable for medical applications in high priority.

3.2 Antenna Connection

Considering the practical application of BSN, the antenna design is an important problem. The antenna is a decisive factor in the BSN health care system with reliability and high efficiency of wireless communication link, or mini implant biosensor in the body to ensure long-term monitoring. For the design of the antenna, the main consideration should be given to the following three factors: little reverse radiation, compact structure, little interaction between the bodies. As Verbiest et al. designed a printed monopole antenna of miniature, low cost, has good energy saving effect in the human body surface, but the frequency and the bandwidth of the antenna and the emission efficiency is vulnerable to human disturbance.

Therefore, this design uses the Kang et al. proposed a folding of the ultra wideband (UWB) antenna. It uses the edge structure to achieve from 3.1 to 12 GHz ultra wide band, slightly the adjacency effect only on the human body, and the specific absorption rate (SAR) is much smaller than a single omni-directional antennas, also it can get the better energy-saving effect [5]. Of course, the UWB antenna can use the existing ISM band (2.4 GHz) and frequency of national medical and/or regulatory authorities approved medical implant communication system (middle-income countries) and wireless medical telemetry system (WMTS), and ultra wideband (UWB) frequency.

3.3 Power and Protection

Lithium ion battery can not meet the capacity and volume of the application requirements in BAN. Some scholars suggested by devices of super-capacitors and carbon-nanotube-based, Latre et al. [6] pointed out, can also be converted human body temperature or vibration to electrical energy. For instance, using thermoelectric generator (TEG) to convert the temperature difference between the body and the surrounding environment to electrical energy supply the node power. The latest technology is the study of a new type of biological catalysts of fuel energy [7], by the decomposition of glucose to the human body sensor supply the node power. In this paper, the design of the model is the use of Hewitt through a mesh fabric which is composed of tiny tubes, convert the temperature into energy [8].

4 Design of MAC Layer

Media access control (MAC) is the core content of any communication protocol, which influences the quality of service (QoS). The purpose of MAC is to reduce collisions and achieve the maximum possible throughput of the signal with

minimum delay, thereby increasing the reliability and performance of network communication and maximize energy efficiency.

In the IEEE 802.15.6 standard, the access method in the super-frame period can be divided into three categories: (1) random access mechanism using CSMA/CA or Slotted-Aloha to obtain the channel resources; (2) access to resources by unscheduled Polling/Post improvisation and non scheduled access (access connection free competition); (3) by a predetermined then one or more of the super frame distribution (called the 1-cycle or m-cycle distribution) to get scheduled access for slot (access connection free competition).

Due to the following character of health care monitoring system:

1. Energy is more limited, because the BAN sensor is implanted or placed on the surface, so the energy capacity is limited, replacement cost is large than ordinary sensor network, so the energy efficiency problem is more prominent.
2. The BAN channel environment changes constantly. Many sensors to monitor in the human body will change because of the movement of the body and the surrounding environment. The BAN sensor is on the movable, increases the complexity of the network, can not use the fixed network architecture to consider.
3. The physiological data of BAN sensor for the detection is of regular, stable data stream. For example, in the medical care application, the above MAC protocol were not considered the most physiological information of the human body (such as blood pressure, body temperature) changed little during the day, and most of them are in the normal range, the normal physiological information is not necessary for real-time processing, there should be a selection method.
4. For medical care in the BAN, the transmission delay of abnormal data is fatal consequences, so send emergency data should be a priority. This is a common WSN are considered. For the emergency data of some sudden illness of human body may be the priority must be considered.
5. The BAN node is strategically placed in the body or body surface, no redundant nodes handle communication failure. And the sensor network is to balance the general service quality through the redundant nodes.
6. To sum up, it is very important with the design of MAC layer priority to urgent data real time communication. In this paper, the French scholar LETI proposed hybrid medium access control protocol of priority MAC (PMAC) [9] base on IEEE 802.15.6, here data channel separates from the control channel, give priority to important traffic (traffic emergency).

5 Network Topology

The network architecture of BSN is an important part of the previous system architecture. It is the logic organization of communications equipment in the system (such as sensor nodes). In general, network architecture includes the star topology,

ring topology, mesh topology and bus topology. The choice of network architecture influences by system characteristics, and can be influence on lot of performance of system, such as energy consumption, traffic load capacity, node failure robustness and MAC protocol etc. The destination of the BAN network architecture choice is to better ensure that the wireless communication with low energy consumption and high reliability of data transmission, and the choice of architecture needs to consider the following factors: energy consumption, transmission delay, inter-user interference, node failures and mobility. In normal circumstances, the star topology network structure corresponds to one hop wireless communication mode, while the mesh topology corresponds to the multi hop wireless communication.

The structure of the IEEE 802.15.6 is mainly star network topology, but also there will be a net or mixed topology structure, such as the need for multi-hop communication mode when the nodes far away from the body or body block. So it is in large scale BAN network. In this paper, the choice of network architecture is not a single. From a practical perspective, the scale of general BAN networks and the complex degree is mainly base on the architecture of choice. Application of BAN with less node and simple function will first choice the star topology structure of hop wireless communication. General speaking, for the more nodes or large scale BAN network, to select the mesh topology and hybrid topology is suitable.

The design of this paper is BAN network with mesh or mixed topology structure. In the network there has a lot of work to do. On the one hand, the connection probability model in multi hop BSN network need to propose, without the use of circular coverage model considering the problem of wireless communication link. In addition, in the multi-hop communication architecture there will have multiple communication links between two entities, so the mesh topology structure can improve the reliability of the system. On the other hand, multi-hop communication network or mixed topology corresponds also to wearable sensors and peripheral sensor combination, then by using the method of distributed reasoning or strategy to realize intelligent identification and monitoring, multi-hop wireless communication another role is to construct the control system based on BSN network.

6 Choice of Sensor

The sensor technology is an important foundation to build a health monitoring system. All kinds of sensor miniaturization, intelligent, high precision, low power is necessary to support BAN. Low power to the sensor inside the body is essential for implantation of performance. The sensor must be collecting physiological signals, such as wireless ECG and pulse blood oxygen node and nodes of wireless temperature sensor can measure a variety of important physiological signals such as blood pressure, body temperature, blood oxygen, ECG, EEG, emg, etc.

According to the location of the body, sensor nodes in the monitoring system can divide into 3 categories: 1. sensor nodes implanted in the body, including implanted biological sensors and inhaled sensors (such as a camera pills); 2. sensor nodes

Table 1 Sensor nodes in the health care system

Type	Sensor	Function	Location
Medical area: EEG, brain electrical activity for monitoring; ECG, used to monitor cardiac activity; electromyography, used to monitor muscle activity; respiratory monitoring, monitoring of respiratory system. Also has some simple monitoring such as body temperature, heart rate, blood oxygen, blood pressure, glucose etc.	Saturation of blood oxygen sensor	The concentration of oxygen in the blood, is an important parameter of respiration and circulation	Internal
	Ring type heart rate sensor	Describing the voltage charts during the heartbeat caused by heart or heart	Surface, internal
	Glucose sensor	Subcutaneous implantation of nano measurement of blood glucose	Internal
	Blood sensor	No compression with blood pressure measurement	Surface
	EEG scanner	Monitoring the human brain electrical signal	surface
	Sensor of lung function	Velocity and mass flow rate measurement of human breath, and calculated the forced expiratory vital capacity and pulmonary function of human body	Internal
Geographical location, environment: information: To monitor the patient’s daily activities, to help find the disabled access to environmental information, or the lost people	Temperature and humidity sensor	Real time monitoring object information and advice given by the temperature and humidity information	Surface
	GPS + Pressure sensor	Absolute pressure measurement of gas, used for localization of GPS, that the altitude and weather conditions	Surface
	Vision sensor	Use of laser scanner, or digital camera to obtain the image information of CCD	Surface
	Auditory sensor	Bionics sensor	Surface
	Acceleration sensor	Recognition of human posture and motion	Surface

worn on the body, such as a glucose sensor, pressure sensor, non invasive blood oxygen saturation sensor and temperature sensor; 3. environment nodes around the body and near the body the distance which is (relatively) short used for recognition of human activities or behaviors.

Combined with the existing medical information and sensor development status, this paper designed the sensor nodes in the health care system as follows (Table 1).

According to the technical challenges faced in the design of node in the above, we can consider the aspects of electronic and electrical characteristics of node and the function optimization design, also can consider to increase the battery life to

deal with. Another strategy is to design a transmitter with low voltage low power repeater, high integration and high performance to deal with the above challenges of sensor nodes with low energy consumption. In addition, facing the implantable node energy consumption we hope to reduce the energy consumption to 100 W. The radio interface is still a challenge, the design of a good radio interface and its optimization strategy is also able to promote the performance of the sensor node operation and low energy consumption.

7 Security Design

Health monitoring system is mainly used in human peripheral, the human physiological information and other important data transmit in the network, so it is a very private system, only the authorized user can query and monitoring the network; on the other hand, the WBAN is composed of the main data of the accident notification chain, so it need protected, can not failure otherwise, even failure, once the emergency situation, will cause unimaginable consequences for users. In general, protocol, software should be considered its security at the beginning of design, especially in fields of the protection of confidentiality of data transmission and reliability of the network. There are three security levels defined in IEEE 802.15.6 standard, each level has different security properties, protection level and frame format.

The first level is one-time authentication of both sides of communication. It will inevitably to be a fraud, forgery, interception of information security incident follow-up; third levels encrypted for each data frame to ensure the communication security of sensor networks, but this does not meet the purpose of energy saving. Therefore, this paper will design the safety level of health care system in second levels, namely, a certification need to be done at the beginning of each session, so a choice of tradeoff on safety and energy saving was in consideration (Fig. 2).

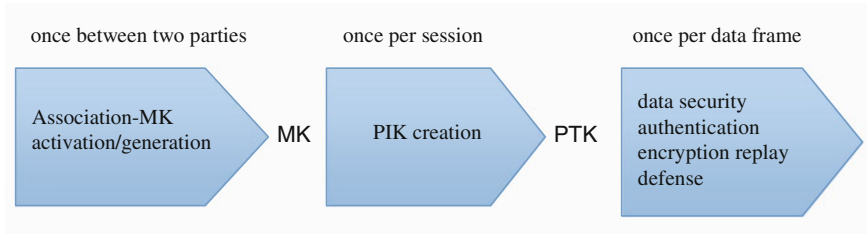


Fig. 2 Three security levels in IEEE 802.15.6 standard

8 Conclusion

In a word, the WBAN will develop towards intelligent, while the sensor nodes develop towards minimized, mobile, implantable and wearable, interactive. We hope to be able to combine various communication technology and network technology, meanwhile, construct the adjustable precision, large-scale, comprehensive health monitoring platform based on BAN. In addition to the measurement of blood pressure, pulse, ECG, EEG, blood, body temperature, blood glucose concentration, action, the surrounding environment information, our system can perform image recognition and intelligent information processing to provide clinical professional degree.

BAN will be more and more involved in data fusion, MAC protocol, energy, parallel and distributed algorithm in the field of technology development. In short, as a new technology, BAN has wide application space in the next year. It is generally considered BAN is worth to research more deeply and widely.

References

1. Benoit L, Bart B, Ingrid M et al (2011) A survey on wireless body area networks. *Wireless Netw* 17(1):1–18
2. 802.15.6-2012—IEEE standard for local and metropolitan area networks—Part 15.6: wireless body area networks. 29 Feb 2012. pp 1–271
3. Gong J, Wan R, Cui L (2012) Research advances and challenges of body sensor network (BSN). *Chinese J comput Res Dev* 47(5):737–753
4. Chen M, Gonzalez S, Vasilakos A, Cao H, Leung VCM (2011) Body area networks: a survey. *ACM/Springer MONET* 16:171–193
5. Hanson MA, Powell HC (2009) Body area sensor networks challenges and opportunities. *IEEE Comput Soc* 58–65
6. Latre B, Braem B, Moerman I et al (2011) A survey on wireless body area networks. *Wireless Netw* 17(1):1–18
7. Sasaki S, Karube I (1999) The development of micro fabricated biocatalytic fuel cells. *Trends Biotechnol* 17(2):50–52
8. Hewitt CA et al. (2012) Multilayered carbon nanotubes/polymer compositebased termoelectric fabrics. *American chemical society* (Jan 2012)
9. Bradai N (2013) New priority MAC protocol for wireless body area networks mobile health'13, July 29. Bangalore, India

The Analysis of Hot Topics and Frontiers of Financial Engineering Based on Visualization Analysis

Liangbin Yang

Abstract The paper did a visualization analysis of co-citation data records regarding to financial engineering which were retrieved from Web of Knowledge by making use of CiteSpaceII software. Through establishing the knowledge map of financial engineering fields, this analysis reflects important figures, articles, knowledge structures, evolution rules of financial engineering industry. Confirms and the research edge and trend of international research on Financial Engineering by detecting subject headings whose word frequency fluctuation are significant.

Keywords Financial engineering · Knowledge map · CiteSpace · Visualization analysis

1 Introduction

Financial engineering has general and special concepts. In this paper we adopt the generalized concept, which refers to the use of engineering means everything to solve the financial problems of technological development. Meanwhile, financial engineering also includes not only the design of financial products, the financial product pricing, trading strategy design and financial risk management, et al. By the late 1980s, with the rapid development of commercial banking, investment banking and securities investment business, John Finnerty proposed the definition of financial engineering, thoughts that financial engineering is a creative solution to the problems of finance including design, development and application of innovative financial tools and financial instruments. The essence of financial engineering is the innovation and practice of financial services, which contains design,

L. Yang (✉)

School of Information Science and Technology, University of International Relations,
Beijing 100091, China
e-mail: ylb@uir.cn

development and implementation of innovative financial tools and financial instruments. It provides a new way of thinking and financial innovations and become the main driving force of modern financial development in last 20 years, and represents the direction of financial development. In recent years, the rapid changes of corporate finance, commercial bank and investment bank led to the birth of a new discipline, people called it as financial engineering. In China, Conducted the research on financial engineering and built the awareness on financial engineering has important practical significance and broad prospects.

Information visualization is an interactive visual representation for abstract data using computer technics to enhance people's cognition on these abstract information. Information Visualization helps people quickly observe, cognitive processing-related information through the visual channel, facilitates to analyze the data, find that the laws and make decisions [1]. Information visualization can also reveal the relationship between information and the information hidden in nature rule. Citation analysis visualization is an important branch of information visualization, it first deal with massive citation data, then use the technology of information visualization to make it easier to observe and understand the information for people, and finally find the hidden rules and patterns in the data. Knowledge Mapping is the theory and methods combined with applied mathematics, graphics, information visualization technology, information science and other disciplines, and metrology citation analysis, co-occurrence analysis and other methods, showing the core structure subjects and develop history, frontier research and the overall framework of multi-disciplinary knowledge integration [2].

This paper is intended to quantitative investigate and visualization analysis for research the field of financial engineering authors, the art and cutting-edge research issue for study, draw a map to show the research frontier in the field of financial engineering, as well as hot field, breaking traditional methods of analysis, makes the majority of scholars to be more intuitive understanding of financial engineering research area.

2 Data and Research Methods

2.1 Data Retrieval

This paper downloads papers in field of financial engineering from the Web of Knowledge database which belongs to SCI by keyword retrieval, our search expression is: Topics = ("financial engineering"). The retrieval time is the May 3, 2012. Select the theme includes all disciplines, library update time for all years, the search results only include journal articles, and the paper language is limited to English. Our search results are as follows: Totally published papers is 440 records

between 1993 and 2012. Each of record includes a data of authors, title, abstract, and literature citations, save as a pure formatted text.

2.2 Research Methods

Herein our visualization tool for citation analysis is CiteSpaceII, the version number is 3.0.R5.

This scientific literature analysis tool developed by Chen Chaomei team of Drexel University in US based on JAVA platform is a pluralistic, sharing time, dynamic network analysis of the new generation of information visualization techniques [3, 4]. CiteSpace is based on the concept of research front in information science and the time-variant duality concept between intellectual bases, and implemented the two complementary views: cluster views and time-zone views [5, 6]. We analyze a particular technology areas and disciplines, by co-occurrence analysis on keyword and co-citation analysis on document, drawing the scientific knowledge map in the field of science, which show the trends of a discipline or field of knowledge in a certain period of development, to form the historic evolution of several research frontiers. Here the CiteSpaceII software can free download at website <http://cluster.ischool.drexel.edu>.

2.3 Data Pre-processing

Before the data processing with formal, we need download the data from the Web of Knowledge database and transform text into format that can be run. That is to separate 440 records as 440 text files that contain only a record, see Fig. 1.

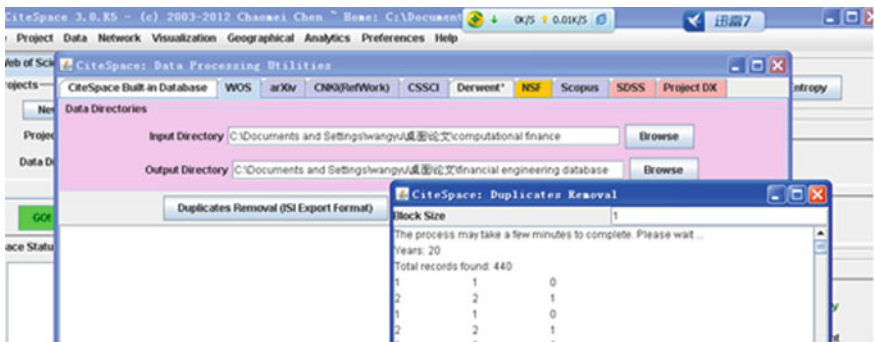


Fig. 1 The pretreatment data format conversion

3 Result Analysis

3.1 Analysis of Intellectual Base of Research Frontier

Intellectual Base of research frontier specified the cited files with term vocabulary. It reflects the situation on the absorption and utilization of advanced concepts in the scientific literature. In CiteSpaceII, it can cluster the co citation network by spectral clustering, the formation of knowledge base of the co citation map of network knowledge [7]. Now we precede visual analysis to the previous downloaded 440 text data using CiteSpaceII. Setting the time scaling value as 2, saying that divided 20 years into 10 periods for segmenting process. Here segmenting process to the data mainly consider the following two aspects: one factor is the CiteSpace software using the principle of “divide and conquer strategy” in the design and operation process; the second factor is easy to analyze the prominent point on evolution of subject and frontier of temporal pattern using of this form. Secondly, setting c , cc , ccv (where c is the literature citations; cc for a total of two literature citations; ccv for literature co-citation coefficient) thresholds are (3, 2, 20), (2, 3, 20), and (4, 3, 20), wherein the specific annual threshold partition is determined by linear interpolation. Finally, according to the different content of analysis, select the corresponding network nodes, such as the choice Cited Reference, Cited Author, Cited Journal, Keyword, Institute as an analysis object, set the threshold value of time slices as 30. Thus, CiteSpaceII can begin analyzing objects with co-citation analysis of the literature, the author co-citation analysis, journal co-citation analysis, keyword co-occurrence analysis, co-institutional analysis, and co-authors analysis and draw the appropriate scientific knowledge maps.

After running the CiteSpace, the result of co-citation network knowledge map on financial engineering literature is Fig. 2, in which includes 241 nodes, 1002 connection lines. In Fig. 2, each node represents a cited literature, extending outward without color circle described in the literature citation time series in different years, the number of citations and the thickness of the circle is proportional to the corresponding year. The node with purple circle represents the key nodes from a cluster transit to another cluster. Assume that a node with a circle of gray as the key node. Here the key nodes has relative the center of high degree and cited frequency, from the view point of knowledge, the key node literatures are classical literature which is generally proposed a new theory or of great theoretical innovation, these nodes may become the key point from one time period to another time transition network. Therefore, to determine the key nodes of research area is the phase of econometric analysis. Supposed that $\Psi\alpha$ and $\Psi\beta$ is main research frontier with paper laded α and β at t moment produced by the knowledge base of $\Omega\alpha = \Phi(\Psi\alpha)$ and $\Omega\beta = \Phi(\Psi\beta)$, then form 2 co-citation clustering with paper α and β , respectively. Connecting the paper $[p(i)]$ on the two clustering path, which describes the character transition from $\Psi\alpha$ to $\Psi\beta$, here $[p(i)]$ called key node. In Fig. 2, includes 7 key nodes of literature, the details is as Table 1.

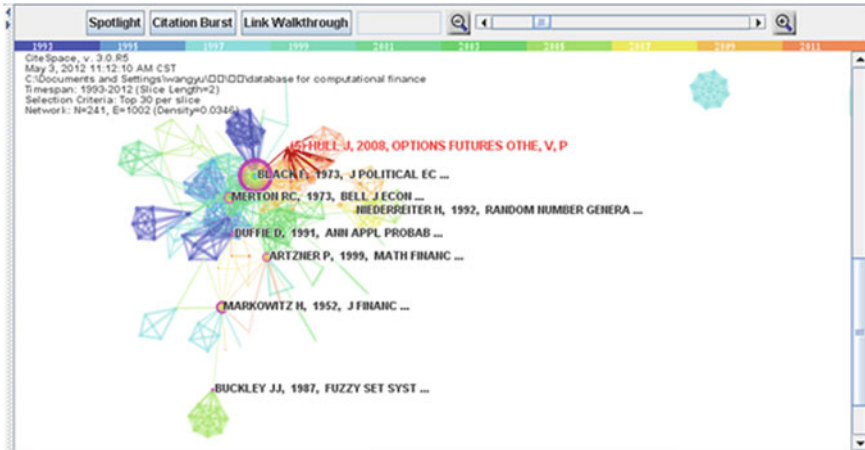


Fig. 2 Financial engineering literature co-citation network knowledge map

Table 1 Information of key nodes of co-citation network knowledge map

Node author	Public year	Paper publication	Co-citation frequency	Central degree
Black F.	1973	J Political Con	55	0.64
Merton R.C.	1973	Bell J Econ	24	0.11
Markonwitz H.	1952	J Financ	22	0.21
Artzner P.	1999	Math Financ	17	0.16
Duffie D.	1991	Ann Appl Robab	8	0.12
Niederreiter H.	1992	Random UMBER Genra	15	0.10
Buckley J.J.	1987	Fuzzy Set Syst	6	0.10

3.2 Analysis of Research Hotspot

From Figs. 2 and 3 can be seen, research and financial mathematics, financial engineering (Financial Mathematics) are inseparable. Financial mathematics is to use mathematical tools to research financial analysis, mathematical modeling, theoretical analysis, numerical calculation and quantitative finance, in order to find the inherent law and used to guide practice. Can be understood as the application of financial mathematics, modern mathematics and computer technology in the financial sector and therefore, mathematical finance is a new interdisciplinary subject, development is very rapid, is one of the frontiers of the discipline of the very active. Through analysis of two important research directions in the field of financial engineering is in Table 2.

From Figs. 2 and 3 we can be seen that research on financial engineering and financial mathematics are inseparable. Financial mathematics is to use mathematical

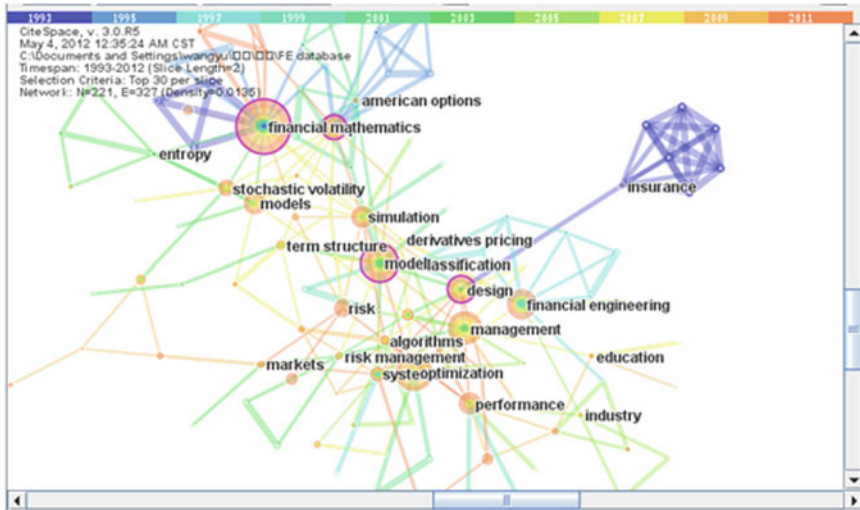


Fig. 3 High frequency keywords knowledge map in financial engineering

Table 2 Vocabulary of research hotspot in financial engineering

No.	Frequency	Keywords	No.	Frequency	Keywords
1	41	Financial mathematics	11	12	Risk management
2	32	Model	12	11	Innovation
3	30	Optimization	13	10	Uncertainty
5	25	Design	14	10	Algorithms
6	20	Simulation	16	9	Interest-rates
7	16	Stochastic volatility	17	6	Insurance
8	13	Valuation	18	6	Credit risk
9	13	Option pricing			

tools to research finance, then proceeds financial analysis, mathematical modeling, theoretical analysis, numerical calculation and quantitative analysis so as to find the inherent regular and used to guide practice. Financial mathematics can be understood as the application of modern mathematics and computer technology in the financial area. Therefore, financial mathematic is a new interdisciplinary subject, development with rapidly, is one of the very active discipline of the frontiers. Table 2 draws the two important research directions in the field of financial engineering.

1. Concept of financial engineering and research scope: Innovation, Design, Optimization, Risk management, et al. can expand the concept and research in the field of financial engineering of the three main aspects. American finance professor John Finnerty divide the scope of the study of financial engineering

into three aspects: The first one is the design and development of new financial instruments, which is a major area of research in the current financial engineering, from the exchange, Options, note issuance facilities, interest rates protocol to index futures, covered warrants, securities depository receipts, zero coupon bonds, convertible bonds, synthetic stock all belongs to this column. The second one is design to reduce the transaction costs of new financial instruments, this part contents includes the optimization of inner operations in financial institutions, the exploration and exploit of arbitrage opportunities in financial markets, and innovation of transaction settlement system. The purpose is to fully mining the profit potential to reduce regulatory costs. The third one is to provide creative solutions for a complete system to solve some financial problems, including various types of risk management, the development and application of technology, innovative cash management strategy, the creation of company's financing structure, the design of corporate mergers and acquisitions program, the implementation of asset securitization and other program contents.

2. Key research object in financial engineering: (a) Option pricing, it is important basis in whole discipline of financial engineering. Option is the acquired option which allowed buyer to buy or sell a certain number of commodities in the future after paid a certain fee options. (b) Stochastic Volatility (SV) model, it is the most active model of income volatility, was first proposed by Clark in the description of joint distribution of stock returns and trading volume, then introduced in econometrics by Harvey et al. (c) Uncertainty trends, the price of investment vehicle is uncertainty in financial markets brings the volatility of returns. The phenomena always are the core issue in the financial field. While research on the volatility of return rate is the basis of analysis such as formation mechanism in capital asset price, financial risk management, financial derivatives pricing, portfolio and so on.

3.3 Research Frontier and Trends Analysis

Research in any field always forward evolution constantly, with the research frontier constant substitute and time lapse, the original research frontier is gradually maturity. It forms the knowledge base of discipline development, in the same time a new research frontier is birth out. The progressive relationship in literature citation expresses the changes in number of words or phrases in research frontier. So, we can detection frequency change to determine the frontier research fields and development trend. Provided by CiteSpaceII of frequency detection technology, we first analysis the previous retrieval data, then set the value of burst term [8], inspect time distribution on term frequency, detect the high rate of frequency change from large subject term. According to the change trend of frequency which is not only the high frequency, we determine the frontier and the development trend in field of financial engineering, as shown in Fig. 4.

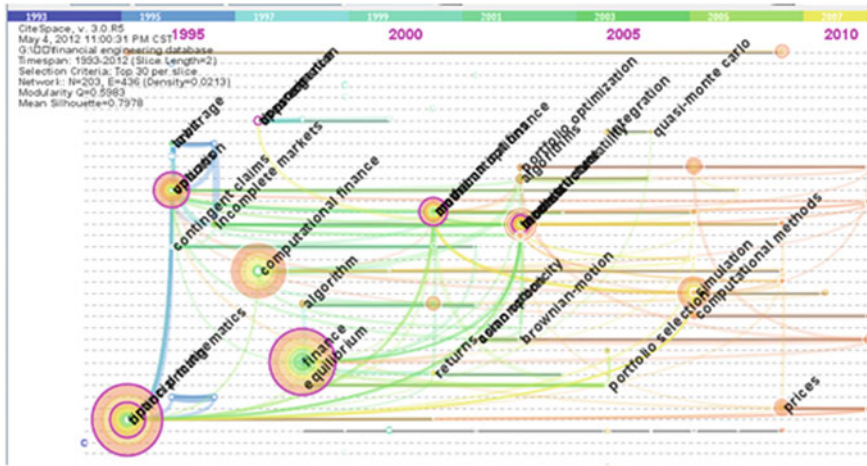


Fig. 4 Knowledge map of research frontier in financial engineering

By setting the appropriate threshold to detect including portfolio selection, multivariate integration, quasi-monte carlo methods, stochastic volatility, portfolio optimization et al. 5 emergence word, we can see the portfolio selection is the biggest change frequency, while quasi-monte carlo methods and portfolio optimization less than the forth. In the following we simply analyze the three important emergence words, then can be detected that the emergence words have good effect on the forward direction.

Portfolio selection is to study how to put the wealth distribution to different assets at a given level of risk to achieve the maximum benefits, or in the case of income to minimize the risks. The balance of risks and benefits always be in the process of investment activity, it is one of the basic problems of investment decision and management. In recent years, with the development of computing technology and information technology, the method of stochastic programming has become a hotspot and frontier of financial engineering in the field of research and practice of dynamic portfolio choice, and achieved certain results. Monte Carlo's method is based on the theory of probability and mathematical statistics. The method and the two fork tree, finite difference methods are all belong to numerical pricing methods. Its essence is the average return of asset price path by simulating the target prediction option and option price estimate. Pricing analysis of Monte Carlo method and Monte Carlo method has been widely used in financial securities, and obtain the good estimate effect. In recent years, the Monte Carlo method and imitate Monte Carlo method are applied more and more widely in financial derivative securities pricing. Base on this theory, the enterprise investment decision analysis method of real option has increasingly become the focus of attention to various people. Portfolio optimization original from the investors always want to get the biggest income at the level of the minimum risk. According to the operation

research of multi-objective programming theory, we can constraint an object on a certain level to transform the multi-objective optimization problem into a single objective optimization problem. After that, the optimal solution is the efficient solution in multi-objective programming. In recent years, more and more investors pay attention to measurement of investment risk, the portfolio optimization problem is sustained attention by them.

4 Conclusions

As we known the knowledge map can help analyze focus on the evolution of the research hotspot in the research area. This paper uses the CiteSpaceII software to draw the knowledge map of the research area of financial engineering, reveals that our research focus mainly concentrated in the innovation and development, option pricing, risk management, etc. By detecting the change trend of term frequency, it can deduce that in next few years the research frontier mainly focused on portfolio selection, stochastic volatility model, portfolio selection, Montecarlo methods etc. The results of this study and research methods and data sources have certain practical significance. But in overall, this study belongs to exploratory research. There are still many shortcomings. For instance, the data retrieval process is too rough to fine, which may affect the amount of literature in scale, causes the results of analysis is not comprehensive. Meanwhile, for threshold to determine, it is largely based on the experience to carry out the selection of threshold with a certain degree of subjectivity to some extent. In the future research, we will further study the other aspects in this field.

Acknowledgements Supported by “the Fundamental Research Funds for the Central Universities”, Project No. 3262015T20, Topic: A new exploration of information research technology of Big Data environment. Project Leader: Liangbin Yang.

References

1. Xiao M, Chen J, Li G (2011) Visualization analysis on the research of mapping knowledge domains based on CiteSpace. *Chin J Libr Inf Work* 66(6):91–97
2. Yang L (2012) Some issues of scientometrics and visualization. *Chin J Intell* 31(4):1–4
3. Zhou J-X (2011) Documents visibilization analysis of information visibilization based on the CitespaceII. *Chin J Inf Sci* 29(1):98–103
4. Li Y, Hou H (2007) Study on visualization of citation analysis. *J Chin Soc Sci Tech Inf* 26(2):301–308
5. Zhao R, Wang J (2011) Research of international social network analysis in frontier domains in visualized information. *Intell Inf Shar* 139(1):88–94
6. Chen C (2004) The searching for intellectual turning points: progressive knowledge domain visualization. In: *Proceedings of the national academy of sciences of the United States of America (PNAS)*, pp 5303–5310

7. Chen C (2006) CiteSpaceII: detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inform Sci Technol* 57(3):359–377
8. Liu Z, Chen Y, Hou H et al (2008) Mapping knowledge domains methods and application. People's Publishing House, Beijing, pp 54–56

An Efficient ACL Segmentation Method

YunBo Rao, XianShu Ding, Jianping Gou and Ying Ma

Abstract Soft-tissue segmentation has always been difficult point in the medical research of diagnosis of soft-tissue defects. Especially for Anterior Cruciate Ligament (ACL) rebuilding surgery, ACL segmentation from all soft-tissue inside knee joint, including Posterior Cruciate Ligament (PCL) and meniscus, is a very important task. In this paper, we propose a novel ACL segmentation method: Space Model Contrast Clustering-based (SMC-based) ACL Segmentation. Unlike the widely used processing method, such as segmentation by MRI gray values and Mimics segmentation drawing, the proposed method relies 3D model of knee joint to segment soft tissue by self-adaptive K-means clustering. Extensional experiments demonstrate that the proposed method can be capable of solving the problem of soft-tissue segmentation well and has achieved higher ACL segmentation efficiency.

Keywords Soft-tissue segmentation · Anterior cruciate ligament · SMC-based ACL segmentation · Self-adaptive K-means clustering

Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

Y. Rao (✉) · X. Ding

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China
e-mail: uestc2008@126.com

J. Gou

College of Science and Engineering, Jiansu University,
Jiangsu 450001, People's Republic of China

Y. Ma

School of Computer and Information Engineering, Xiamen University
of Technology, 361024 Xiamen, People's Republic of China

1 Introduction

With some advanced medical devices, like CT, MRI, being widely used, medical 3D reconstruction has been increasingly important research topic in recent years [1, 2]. 3D reconstruction can largely help doctors detect and diagnose some sickness. Nowadays, more and more surgeries are operated under the guide of scheme designed by computer. Besides, 3D reconstruction has many application in kinds of medical sub-topics, such as tissue segmentation [2], tumor detection and location.

Mostly, the first processing steps is to collect data from patients by CT or MRI. Then reconstruct the 3D model of region of interest (ROI), like the knee joint or brain. Among all these tasks, tissue segmentation is the most basic and simplest one. Because independent ROI can provide the most direct object without any other noise and interference. However, tissue data from CT/MRI can hardly be segmented well due to rough source data and the existed techniques. Especially for soft tissue, segmentation would become much harder [3, 4]. Most of soft tissues has the similar scanning values, and CT scan seems to react same to all soft tissue. For example, inside knee joint, there are mainly three kinds soft tissues, Anterior Cruciate Ligament (ACL), Posterior Cruciate Ligament (PCL) and meniscus. And with the interference of fleshy tissue, it would be a large challenge to segment ROI from them.

ACL segmentation is very important because it the key step of injury detection. With the heavy traffic and more sports activities, more and more people suffer from the injury of ACL. Now the accurate way to ACL injury detection is to observe the biological characteristic of ACL under the arthroscope [5], and this way is called as arthroscopy. However arthroscopy is minimally invasive operation after all, it can hardly be accepted by each patient. Then the ideal way is to reconstruct 3D model of ACL and to detect its physical parameters to give the correct diagnosis. Segmentation of ACL from scanning data of knee joint then become so important.

The most widely used method is segmentation by threshold of gray values and Mimics editor. As said before, gray values is not sensitive among soft tissues, so ACL segmentation by gray values threshold would be cause large error. Editing by people is not also a good choice, because it would need much time for people and it is not impersonal, and cannot be applied by each one. In this paper, we propose a novel ACL segmentation method: Space Model Contrast Clustering-based (SMC-based) ACL Segmentation which relies 3D model of knee joint to segment soft tissue by self-adaptive K-means clustering. Extensional experiments demonstrate that the proposed method can be capable of solving the problem of soft-tissue segmentation well and has achieved higher ACL segmentation efficiency.

The remainder of the paper is organized as follow: the proposed method: SMC-based ACL segmentation is introduced in Sect. 2. The experimental settings and results are demonstrated in Sect. 3. Finally, we make a conclusion objectively in Sect. 4.

2 The Proposed Method

Knee joint is one of the most important tissue to keep stable for people. Nowadays the most widely used way to segment ACL from knee joint is to edit the scanning data using some software like Mimics. The most important factor which would affect result is the experience of editor. Always, only some clinician who have been familiar with the anatomical structure of knee joint are eligible. From the anatomical data, we find that the main interior structure is always the same, especially for adults, and they have the same structure of knee joint. So we think it is feasible to refer the existed space distribution of interior structure of knee joint to achieve ACL segmentation from scanning data. Then the SMC-based ACL segmentation method is proposed. As shown in Fig. 1, the anatomical structure of knee joint is constant to each man. It has the main tissue, as following: bones, including Thighbone, Patella, Fibula and Tibia; soft tissue, like Anterior cruciate ligament, Posterior cruciate ligament, Lateral collateral ligament, Medial collateral ligament, Lateral meniscus, Medial meniscus. Our study in this paper has not taken collateral ligaments into consideration due to they are not connected with ACL and PCL. So the main task is to segment ACL from bones and soft tissues, including PCL and meniscuses. The 3D model of these tissue can be seen like shown in Fig. 2. The structure model with green color is ACL.

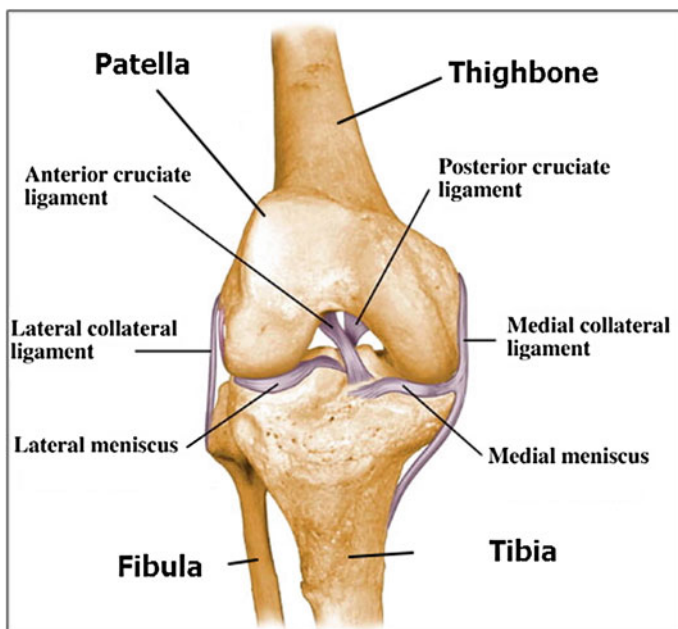


Fig. 1 The anatomical structure of knee joint

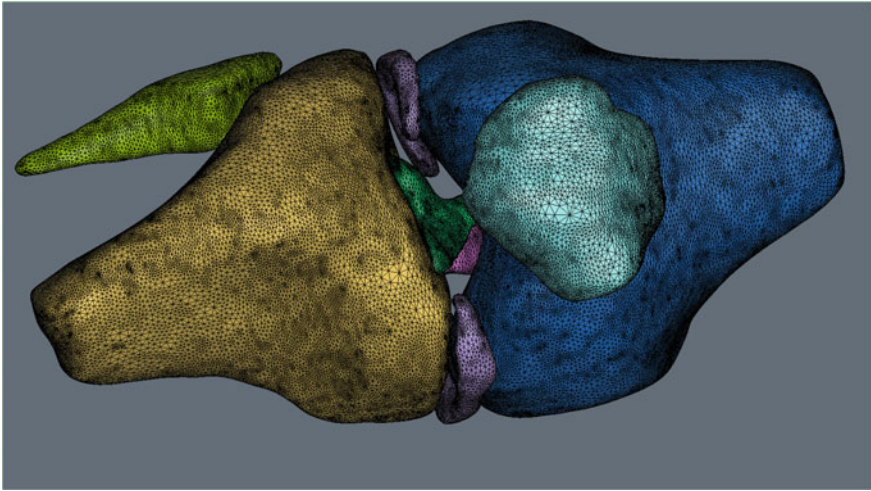


Fig. 2 3D model of the main tissues inside knee joint

For these point clouds, following parameters would have great impact on segmentation result: barycenter (G), angles across with X, Y, Z plane respectively (α, β, δ), and the distance between the point and origin (d). The barycenter G can be computed by these equations:

$$G_x = \frac{\int \int \int_{\Omega} x \rho(x, y, z) dv}{\int \int \int_{\Omega} \rho(x, y, z) dv} \quad (1)$$

$$G_y = \frac{\int \int \int_{\Omega} y \rho(x, y, z) dv}{\int \int \int_{\Omega} \rho(x, y, z) dv} \quad (2)$$

$$G_z = \frac{\int \int \int_{\Omega} z \rho(x, y, z) dv}{\int \int \int_{\Omega} \rho(x, y, z) dv} \quad (3)$$

where G_x, G_y and G_z are three dimensional value respectively, v is the volume of point cloud model, and $\rho(x, yz)$ denotes the density of the model structure. For us human body, each organ is made by some certain materials, so in medical measurement, the density of organ is constant which would greatly help solve the equations above. Then barycenter G of each model can be computed by another simple method, Eq. (4) as follow:

$$G = \frac{1}{N} \sum_{i=1}^N G^i \quad (4)$$

where N is the total number of points in each model, and G^i denotes the i th point in point cloud. Compared to the former equations, Eq. (4) could reduce the time consumption dramatically. The distance d can be computed by Eq. (5):

$$d = \sqrt{x^2 + y^2 + z^2} \quad (5)$$

Then the angles can be solved with distance d using the following equations:

$$\cos \alpha = \frac{\sqrt{x^2 + y^2}}{d} \quad (6)$$

$$\cos \beta = \frac{\sqrt{x^2 + z^2}}{d} \quad (7)$$

$$\cos \delta = \frac{\sqrt{y^2 + z^2}}{d} \quad (8)$$

All of these data from model can be used as arguments into segmentation method. In our study, we choose K-means for the consideration of its simplicity [6–8]. K-means is an unsupervised classification algorithm based on clustering, the basic theory is described in Eqs. (9)–(10).

$$J = \sum_{i=1}^N \sum_{k=1}^K \|X_i - u_k\|^2 \gamma_{ik} \quad (9)$$

$$u_k = \frac{\sum_i \gamma_{ik} X_i}{\sum_i \gamma_{ik}} \quad (10)$$

Assuming that there are N data points in total, they are divided into K cluster. The K-means algorithm is to minimize J , γ_{ik} is 1 or 0. When we get the smallest J , u_k should be met Eq. (10). The iteration cannot be stopped until its convergence. K-means is a high-efficiency method to clustering, however there are two problems hindering on the way. The two existed problems are how to initialize and how to judge convergence well. And the difference in our study is that we use the model data to set initialization and convergence condition.

As the input into K-means, the scanning data of knee joint has the following structure that the algorithm need to segment: Thighbone, Patella, Fibula and Tibia; Anterior cruciate ligament, Posterior cruciate ligament, Lateral collateral ligament, Medial collateral ligament, Lateral meniscus, Medial meniscus. To reduce the time consumption of this task, the algorithm can remove the collateral ligaments directly due to they are outside the knee bones gap. So the number of clusters (K) is 7, 4 kinds of bones and 3 kinds of soft tissues including ACL. Then the barycenter G of these 7 model structures, which are computed by Eqs. (1)–(3), would be used as centriods in initialization processing. And the most important is convergence

condition. Experimental results demonstrate that the best segmentation can hardly be extracted. And the algorithm always miss the and the global optimization between two iterations. In this paper, to reduce the time consumption and the risk of local optimization, we take the variance of angles and distance into consideration of convergence. The convergence condition is shown in Eq. (11).

$$C = \left\| \begin{array}{cc} \Delta d & \Delta \cos \alpha \\ \Delta \beta & \Delta \cos \delta \end{array} \right\| \quad (11)$$

where C denotes the convergence conditional value. The evaluation setting is also important to the segmentation. No matter large or small, it would cause the bad results we cannot stand. We will talk about it in experiments part.

The proposed segmentation method in this paper need the other two traditional segmentation method: threshold segmentation and software editing to remove some other noise data, such as skin and fat. And the proposed processing algorithm is concluded in Algorithm 1 as follow:

Algorithm 1: SMC-based ACL Segmentation

- Step1: Threshold segmentation to extracted bones and soft tissue;
- Step2: erasing editor to remove the noise data;
- Step3: Model data computing using Eqs. (1)–(8);
- Step4: Initialization using Model parameters;
- Step5: K-means iteration until convergence C met condition value.

3 Experiments and Analysis

In this paper, the experimental setting is as follow, hardware: Thunderobot with Inter(R) Core(TM) i7-4710 CPU@2.5 GHz, 16GRAM, NVIDIA GTX 870; software: Windows 8, Mimics 16.0, Geomagic Studio 12.0, MatlabR2012b, SPSS19.0, Geomagic Studio 2012. And the MRI scanning data all are provided by Chongqing Xinqiao Hospital.

Threshold segmentation and erasing editor are two necessary steps in the SMC-based ACL Segmentation method. The basic processing flow of threshold segmentation is shown in Fig. 3. Mimics provides the analysis tool to extract threshold which is different in different patient because it can be affected by some other factors, such as weather, machine. The threshold range of soft tissue of the patient in Fig. 3 is from 1100 to 1800. Then these extracted thresholds can be used to produce segmentation masks. Figure 4a shows us the mask we set in which the green is bone mask while the yellow is soft tissue mask. This figure demonstrates threshold segmentation causes so much noise. Figure 4b shows the results after erasing editor.

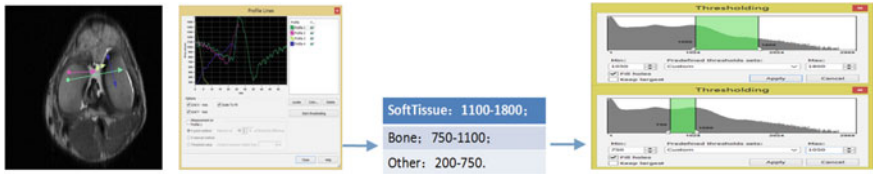


Fig. 3 Threshold segmentation processing flow

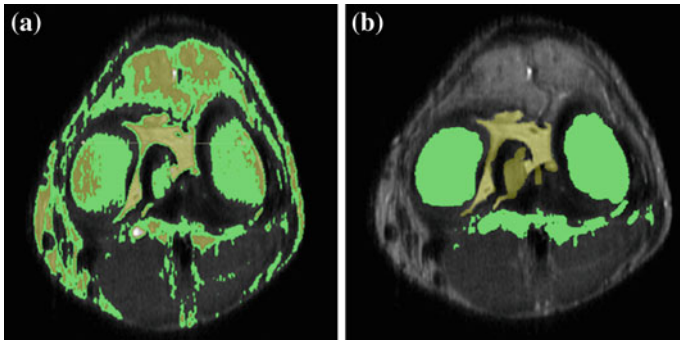


Fig. 4 Threshold and editing segmentation results on tomography

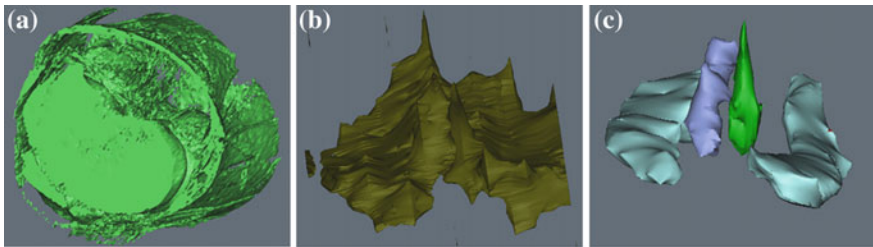


Fig. 5 Results from kinds of method

Threshold and editing segmentation are the most widely used method in medical processing, and it can segment bones from soft tissue well. As shown in Fig. 5a is the 3D objects by threshold masks while (b) is soft tissue 3D object. But the structure of ACL and other soft tissue cannot be observed by any users. The proposed SMC method can segment different soft tissue, like shown in Fig. 5c in which the green is ACL, the red one is PCL while the violet one is meniscus.

In this paper, we also use Geomagic Studio 2012 to optimize the ACL 3D object. Nurbs curve surface is made by some non-uniform ration B-splines [9] which is shown in Eq. (12). And GS2012 can optimize the B splines to remove some corners

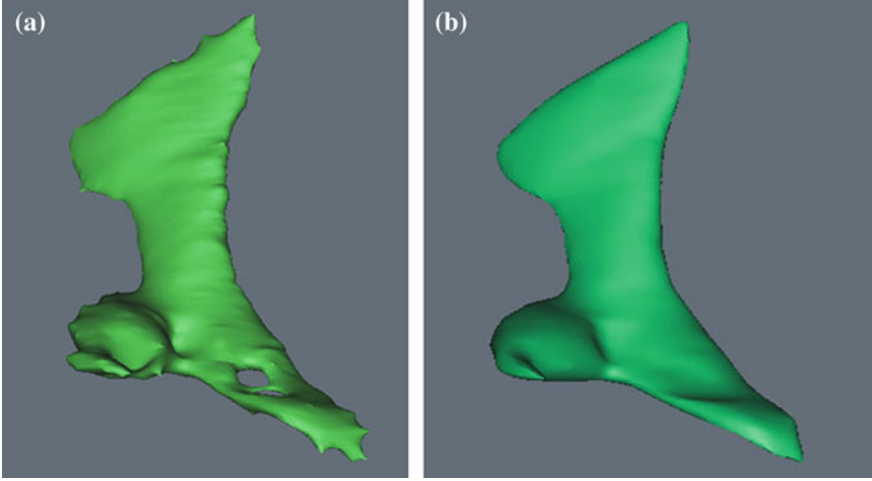


Fig. 6 ACL 3D results and its optimization model

and to make ACL more smooth, like the comparison shown in Fig. 6 in which (a) is the results from SMC segmentation method, while (b) is the optimization result.

$$s(u, v) = \frac{\sum_{i=0}^m \sum_{j=0}^n W_{i,j} P_{i,j} N_{i,k}(u) N_{j,l}(v)}{W_{i,j} N_{i,k}(u) N_{j,l}(v)} \quad (12)$$

where $P_{i,j}$ is corner point, $N_{j,k}(u)$, $N_{j,l}(v)$ are B-splines in u , v directions respectively, $W_{i,j}$ is weight factor.

Figure 7 shows us the segmentation results of 3D model of ACL. In this figure, the segmented ACL has unambiguous edges which would help doctors diagnose. In our study, we use model parameters to initialize K-mean clustering. Experimental results demonstrate that this initialization greatly reduce the iterations and time consumption of algorithm processing. Tables 1 and 2 show the iterations and the time consumption. Of the tables, the data in the parentheses are the results from random initialization. From the comparison, parameters initialization reduces at least half of iterations of random initialization. For ACL segmentation task, iterations are cut down to 1.4 from 7.5, and the time consumption is also reduced into 138 ms from 682 ms.

4 Conclusion

In this paper, we propose a new segmentation method: SMC-based ACL Segmentation which can segment ACL from knee joint well. This new method solves the traditional medical problem, segmentation of soft tissue. The ACL 3D

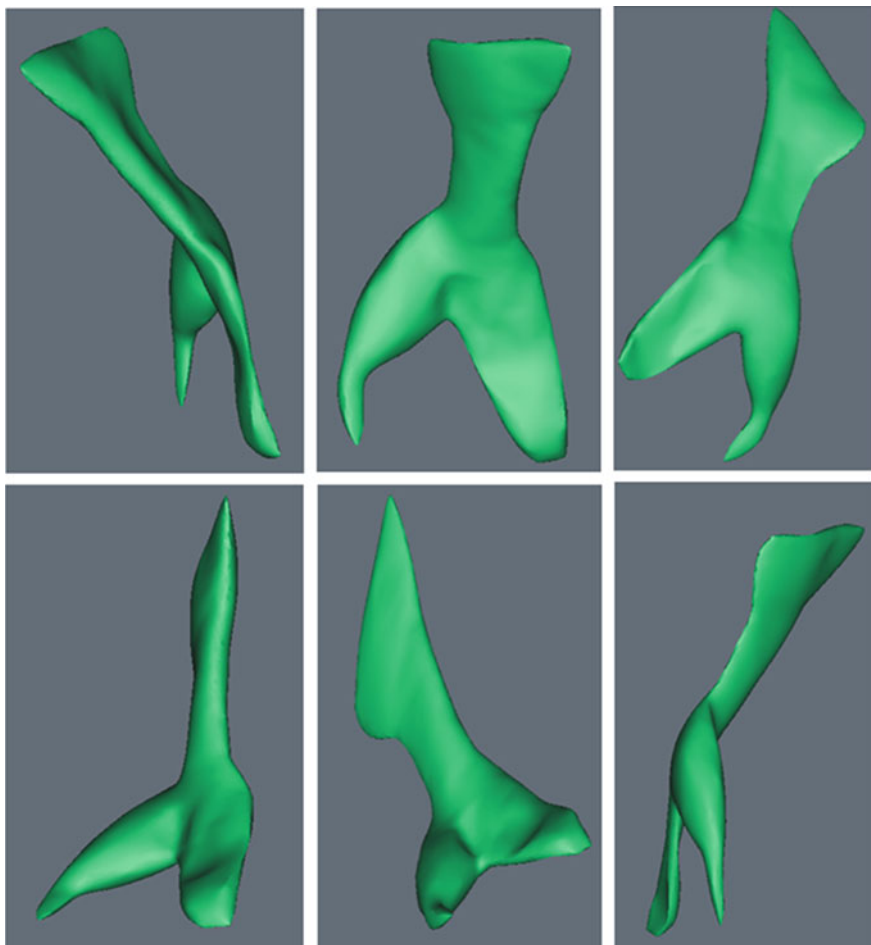


Fig. 7 ACL segmentation results

Table 1 Iterations of K-means in SMC-based ACL segmentation

	Thighbone	Patella	Tibia	Fibula	ACL	PCL	Meniscus
Iterations	6.2 (17.3)	2.1 (4.9)	4.4 (11.9)	1.2 (4.7)	1.4 (7.5)	1.7 (6.9)	3.8 (9.4)

objects from SMC has unambiguous edges which would help doctors diagnose. And model parameters initialization greatly reduce the iterations and time consumption of the proposed algorithm processing.

Table 2 Time consumption of K-means in SMC-based ACL segmentation (*unit* ms)

	Thighbone	Patella	Tibia	Fibula	ACL	PCL	Meniscus
Iterations	485 (1654)	170 (989)	407 (1479)	231 (1074)	138 (682)	206 (941)	467 (1064)

Acknowledgment The authors would like to thank the anonymous reviewers for their helpful comments. This work is partly supported by the National Natural Science Foundation of China (Grant No. 61300092), the Fundamental Research Funds for the Central Universities of China, (Grant No. ZYGX2013J068), Natural Science Foundation of Fujian Province, China (Grant No 2015J05132).

References

1. Koehler C, Wischgoll T, Golshani F (2010) Reconstructing the human ribcage in 3d with x-rays and geometric models. *Multimedia* 17(3):46–53
2. Bouchet A, Pastore JI, Di Meglio L, Robuschi L, Ballarín V (2013) Segmentation and 3D Reconstruction of Microbial Biofilms. *Latin Am Trans* 11(1):324–328
3. Malik OA, Arosha SMN, Zaheer D (2015) An intelligent recovery progress evaluation system for ACL reconstructed subjects using integrated 3-D kinematics and EMG features. *J Biomed Health Inf* 19(2):453–463
4. Saha PK et al (2011) A new Osteophyte segmentation algorithm using the partial shape model and its applications to rabbit femur anterior cruciate ligament transection via Micro-CT imaging. *IEEE Trans Biomed Eng* 58(8):2212–2227
5. Spillmann J, Tuchschnid S, Harders M (2013) Adaptive space warping to enhance passive haptics in an arthroscopy surgical simulator. *IEEE Trans Visual Comput Gr* 19(4):626–633
6. Vora P, Oza B (2013) A survey on K-mean clustering and particle swarm optimization. *Int J Sci Mod Eng* 1(3):24–26
7. Agrawal A, Gupta H (2013) Global K-means clustering algorithm: a survey. *Int J Comput Appl* 79(2):20–24
8. Rajput GG, Patil PN (2014) Detection and classification of exudates using K-means clustering in color retinal images. In: *International conference on signal and image processing*, pp 126–130
9. Rouhani M, Sappa AD, Boyer E (2015) Implicit B-spline surface reconstruction. *IEEE Trans Image Process* 24(1):32–33

Image Haze Removal of Optimized Contrast Enhancement Based on GPU

Che-Lun Hung, Zhaohui Ma, Chun-Yuan Lin and Hsiao-Hsi Wang

Abstract In the domains of computer vision and graphical computation, image haze removal has been a significant issue. By the use of haze removal process, it can significantly improve the visibility of the scene in the image. However, most of the haze removal algorithms bring high computational cost and make algorithms failed in processing huge amount of images. In this paper, we propose a parallel image haze remove algorithm, adopting optimized contrast enhancement approach, to optimize the performance based on GPU platform. The optimization from the proposed algorithm obtains performance acceleration with about 5 times as compared the original version while the haze removal effect is the same. Some haze free images and its original hazy images are shown in the later chapter during this paper. Our work after improvement can process a single picture in a much higher speed after optimization and make it more sufficiently fast for large-scale application which needs image haze removal in computer vision area.

Keywords Image haze removal · Parallel computing · GPU · CUDA

C.-L. Hung (✉)

Department of Computer Science and Communication Engineering,
Providence University, Taichung, Taiwan
e-mail: clhung@pu.edu.tw

Z. Ma

Department of Computer Science and Information Engineering,
Providence University, Taichung, Taiwan
e-mail: g1020435@pu.edu.tw

C.-Y. Lin

Department of Computer Science and Information Engineering,
Chang Gung University, Taoyuan, Taiwan
e-mail: cyulin@mail.cgu.edu.tw

H.-H. Wang

Department of Computer Science and Information Management,
Providence University, Taichung, Taiwan
e-mail: hhwang@pu.edu.tw

1 Introduction

In almost every practical scenario, the images of outdoor scenes often are blurry caused by fog or other types of atmospheric degradation. Due to the atmospheric absorption and the atmospheric scattering, the light of scenery which camera captures is attenuation in different level. Besides, the photometer in air is a part of ingredient of the light of scenery. Image dehazing plays an important role in the domain of image processing. Image dehazing can improve the clarity of the image significantly and revert the color cast of atmospheric scattering of light. Similarity, it can be used to improve the interpretability of images for computer vision and preprocessing tasks.

Generally, the image dehazing algorithms can be classified into two groups: image enhancement and image restoration [1]. The image enhancement algorithms [2–5] are used to improve the contrast in images directly. These algorithms are simple and fast, but they are difficultly to be used to adjust the image characteristics as color changes. The image restoration algorithms [6–8] aim to emphasize features of the image to be suitable for human visual perception. The algorithms have been used to improve the image problem caused by the atmospheric scattering based on strong prior or assumption atmospheric transmission and environmental luminance model. Tan [6] discovers that in general, non-fog image contains higher contrast than the images with fog. The original image can be reverted through maximizing the local contrast of the image. The result is very convincing in theory, but the practical results are not particularly good. Fattal [7] estimates the reflectance and transmittance of the image to infer by assuming that the transmittance and the surface of the projection in the local scene are irrelevant. This approach achieves more accurate and very good dehazing results. However, this method cannot copy with the image with high concentrations of fog.

He et al. [8] proposed an algorithm based on dark channel prior. It presents a simple but effective image prior law—dark channel prior to a single input image dehazing. Dark channel prior is a statistical law for outdoor non-fog image. It is based on a key idea; mostly outdoors image without fog, the each local area exists some pixels in a color channel with low intensity values. Using this effective image prior law and defogging model the algorithm estimates the concentration of fog and revert high-quality images. The experimental results show that this algorithm can achieve the better defogging effect based on a single image, and also obtain more accurate depth image information. However, the processing speed of this algorithm is very slow. The computation time for processing a 1024×768 image is about 60–80 s.

Some of the recent algorithms based on He's algorithm are proposed on improving contrast and luminance of degraded image. Matlin and Milanfar [9] proposed a method based on BM3D [10] and He's algorithm to remove haze and noise from a single image. They also proposed an iterative regression method. Both of these two algorithms can achieve good processed image when the noise level is precisely known. The drawback of these algorithms is that the latent errors by denoising can be amplified when noise level is unknown. Nan et al. [1] proposed a

Bayesian framework to avoid dynamic range compression in He's algorithm. The haze and noise in the input image are removed simultaneously. They adopted an iterative strategy with feedback to achieve the more accurate results than Matlin's approach. For these algorithms above, the computation cost is still high for processing a single image. Kim et al. [11] proposed an algorithm to improve the contrast of the given image and design a cost function in order not to lose too much information while recovering the contrast.

Actually, these image-dehazing algorithms are time-consuming leading to be used as real-time applications difficultly. To enhance the computational performance, Graphic Processing Units (GPUs) has been adopted in many image-dehazing algorithms. Xue et al. [12] proposed haze removal algorithm using dark channel prior implemented on GPU. Valderrama et al. [13] proposed a GPU-based local adaptive algorithm, which uses local statistics of the hazed image, for single image. Ok et al. [14] proposed a GPU-based dehazing algorithm, which executes CPU-based MSR algorithm for each RGB channel and CUDA Gaussian Blur algorithm, for video surveillance.

In this paper, we propose a GPU based image-dehazing algorithm based on by Kim's algorithm [11] through optimized contrast enhancement approach. From the experimental results, the execution time to process 1024×768 image is about more than 2 s by Kim's algorithm on CPU and it can be reduced to 0.4 s by the proposed algorithm. The optimization from the proposed algorithm obtains performance acceleration with about more than 5 times when the image size growing while the dehazing result is as excellent as the original algorithm.

2 Background

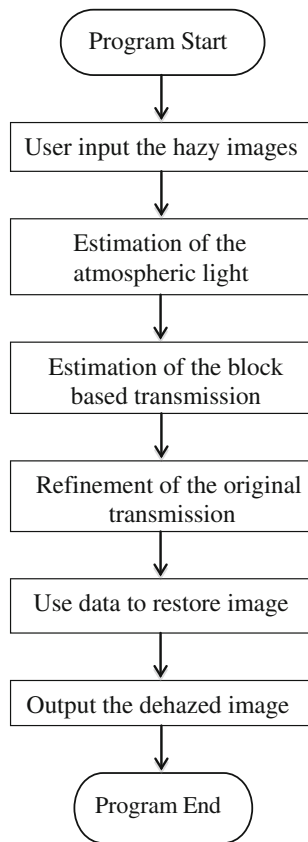
2.1 Haze Formation Model

In computer graphics and computer vision, there is a model describing the observed color of an image in the presence of haze or fog. The model widely used to describe the formation of a haze image is the equation as below [8, 11]:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

Note that in the Eq. 1, $I(x)$ is the observed intensity, $J(x)$ is the scene radiance, A is the atmospheric light, and $t(x)$ is the transmission of the reflected light. The $I(x) = (I_R(x), I_G(x), I_B(x))$ and the $J(x) = (J_R(x), J_G(x), J_B(x))$ represent the original and the observed R, G, B color channels at the specific pixel position x , respectively [15]. Image haze removal is actually to compute J , A , and t from I [16, 17]. Usually, most of the image haze removal algorithms need to calculate these parameters in order to get the haze-free image finally. When the atmosphere is homogenous, the transmission $t(x)$ can be presented as following [18]:

Fig. 1 Modules of image haze removal algorithm



$$t(x) = e^{-p*d(x)} \quad (2)$$

As the Eq. (2), $d(x)$ is the scene depth from the captured camera at pixel position x while p is the attenuation coefficient which is decided by the weather condition.

2.2 Image Haze Removal

The process flow diagram of image dehazing is shown in Fig. 1. Most of haze removal algorithms, either about images or videos, have the similar process that contains these steps to obtain the haze free image. The common steps of dehazing are listed as followings: User input the hazy image to process; Rough estimation of the atmospheric light which is A that represents the ambient light in the atmospheric [6]; Estimation of the block based transmission and the block can be customizing by the user; Refinement of the rough transmission to obtain $t(x)$ using the previous

transmission result. Finally, we can restore the image to become haze-freed according to the hazy image $J(x)$, the computed result A and $t(x)$.

2.3 CUDA Programming Model

Compute Unified Device Architecture (CUDA) is a parallel computing platform and programming model invented by NVIDIA. It enables dramatic increases in computing performance by harnessing the power of GPU. CUDA is an extension of C/C++ which enables users to write scalable multi-threaded programs for CUDA-enabled GPUs [19].

CUDA programs usually contain a special part, called kernel, which will be parallel executed on GPU. The kernel represents the operations or computation to be performed by a single thread and is invoked as a set of concurrently executing threads. These threads are organized in a hierarchy consisting of so-called thread blocks and grids. A grid is a set of independent thread blocks and a thread block is a set of concurrent threads. The total size of a grid (dimGrid) and a thread block (dimBlock) is explicitly specified in the kernel function-call:

```
kernel<<<dimGrid, dimBlock, ... >>> (parameters);
```

Thread communication and synchronization are implemented in the thread blocks so that threads within a thread block can communicate each other through a per-block shared memory and are synchronized using barriers. However, threads located in different blocks cannot communicate or synchronize directly. Besides the shared memory, there are four other types of memory: per-thread private local memory, global memory for data shared by all threads, texture memory and constant memory. Texture memory and constant memory can be regarded as fast read-only caches. One of the optimizing the performance of GPU is to enable threads to access shared memory rather than global memory.

The CUDA architecture consists of a number of streaming multiprocessors (SMs). Each SM contains 8 streaming processors (SPs), which share a per-block shared memory of size 16 KB. All threads of a thread block are executed concurrently on a single SM. The SM executes threads in small groups of 32, called warps. Thus, parallel performance is generally penalized by data-dependent conditional branches and improves if all threads in a warp follow the same execution path.

3 Method

In this section, we will introduce the important works to implement the parallel version of GPU based on the optimized contrast enhancement to improve the computational performance. Each module of the image haze removal shown in

Table 1 Parallel degree of each module

Module	Computing process	Computational complexity by CPU	Parallel degree
Estimation of atmospheric light	Sequencing, comparison	$O(m * n * \log(m * n))$	$O(m * n)$
Estimation of block based transmission	Image partition and matrix operation	$O(m * n)$	$O(m * n)$
Refinement of original transmission	Matrix operation and inversion	$O(m * n)$	$O(m)$
Use data to restore image	Matrix operation	$O(m * n)$	$O(m)$

Fig. 1 has been analyzed to identify the steps that can be parallelized. Then these modules have been implemented on GPU.

3.1 Parallelism Analysis

Based on the computing process of the optimized contrast enhancement algorithm, the computational complexity of the algorithm can be analyzed by the data dependency during the computing process. The parallel degree represents time complexity of the module implemented on GPU. Table 1 shows the result of the parallel degree of each module for $m \times n$ image.

3.2 Dynamic Parallelism

The new feature of dynamic parallelism released in CUDA 5.5 enables the Kepler GK110 GPU to dynamically invoke new threads by adapting to the data by kernel directly. Kernel on GPU has the ability to independently launch additional workloads as needed. The global atmospheric light which is represented the ambient light in the atmospheric is often considered the brightest color in the image. To estimate the atmospheric light more reliably, we should exploit the fact that the variance of pixel values is generally low in hazy regions. As shown in the Fig. 2, a hazy image can be split into four rectangular regions and the score of each region as the average pixel values within the region can be computed. Then, the region with highest score is selected to be split into other four smaller regions until the area of selected region is smaller than the pre-defined threshold. The threshold is set 200 as default.

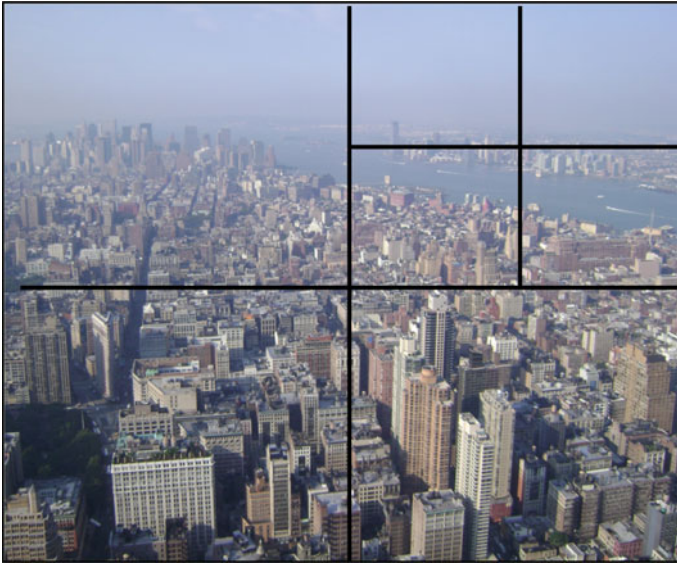


Fig. 2 Atmospheric estimation division

The CUDA feature called dynamic parallelism can be use on the quad-tree subdivision. As shown in the pseudo code, a threshold is defined in the kernel specified to perform the recursive division step.

```

__global__ void Atmospheric_Estimation(...) {
  //define threshold to make the recursion stop;
  if(the size is larger than threshold){
    Atmospheric_Estimation <<<dimGrid,dimBlock, ...>>>( ... );
  }
}

```

3.3 Shared Memory Optimization

The shared memory is on-chip memory of GPU and it has considerable speed as compared to global memory. On CUDA, we launch several hundred of threads in a single block. All the threads in a block which are using the same the shared memory can cooperate with the other in the same block. The build-in function `__syncthreads()` is used to specific the synchronization points of all threads in the kernel; It is the barrier function, and all threads in the block must wait until all the threads in a block arriving the point. From the time of each module of CPU version, the module guided filter cost the biggest part of the total time of program. In the guided filter,

box filtering algorithm is executed to get cumulative function for calculating the integral image, either in one dimension or three dimensions. The box filtering produces the cumulative results of each array which is in X axis or in Y axis. In a factor, the computation between two pixels is not totally independent. If the array data onto the global memory is used for box filter computation, it will decrease the speed apparently because of accessing the global memory too many times leading the performance reduction.

4 Experimental Result

4.1 Haze Removal Effect

We evaluate the haze removal effect of the proposed algorithm on images used in [11]. As shown in Fig. 3, we can obtain very nice images after removing haze by the proposed algorithm. We also apply the proposed method to recover several hazy images and compare the proposed method with Kim's algorithm.

4.2 Performance Optimization

Our platform mainly included CUDA and OpenCV is built on a computing environment of Ubuntu 13.04 with Intel® Core(TM)2 Duo @2.00 GHz 2.00 GHz, Nvidia Tesla K40 with 8G memory. For an image of 1024×768 , the comparison of computational performance is shown in the Table 2. In comparison of the performance of each module, on pure CPU version, atmospheric light estimation costs only 0.04 s, color transmission estimation costs 0.45 s, guided filter costs more than 1.38 s, and restore image cost 0.29 s. 5. Because the refinement of transmission of guided filter is the most time-consuming, we focus our attention on it and make great effort to improve the performance. From the Table 2, the GPU program can earn a speedup which is more than 5 times of the CPU version.

4.3 Performance Comparison

Figure 4 is the comparison of computational performance based on the different hazy image sizes. From the figure, the running time of the CPU dehazing and the proposed algorithms have a linear relationship with the sizes of images. We can get an increasing the trend from the figure, the operating time of the GPU increase in a stable and gentle trend when the image size becomes large, and basically stable at less than 0.4 s. However, the CPU program has bigger and unstable amplitude with the image enlarging.

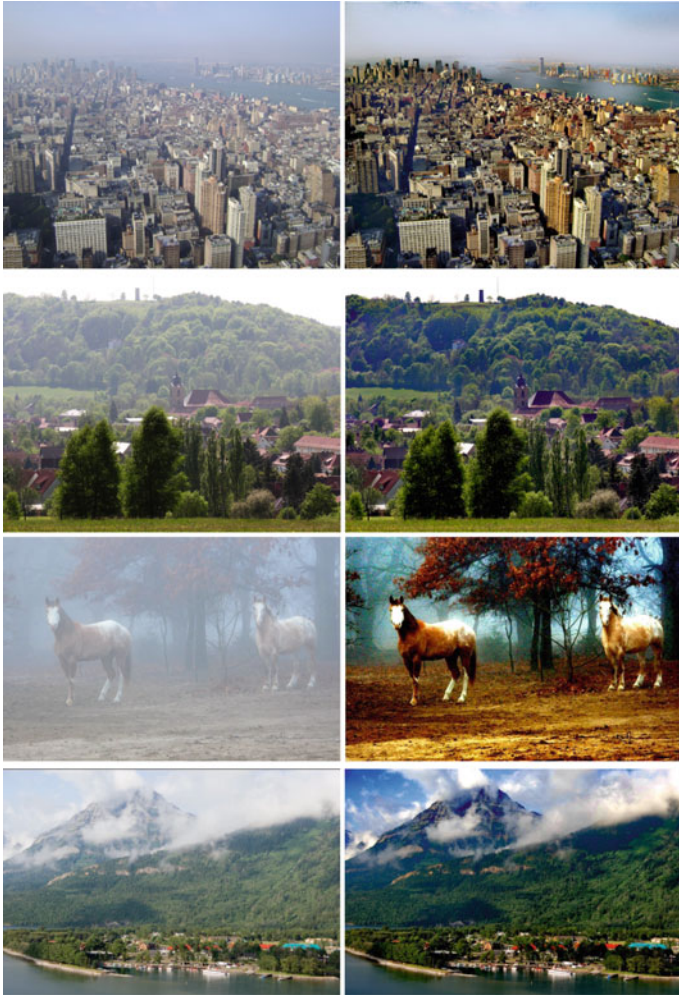
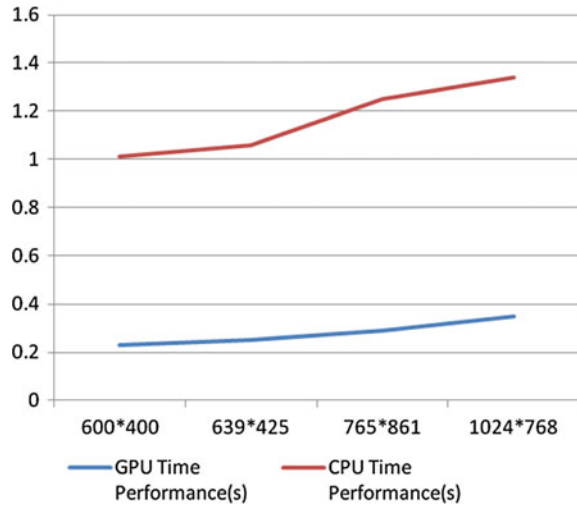


Fig. 3 Images before and after optimization

Table 2 Runtime performance on CPU and GPU

Component name	CPU Runtime(s)	GPU Runtime(s)	Speedup
Estimation of the atmospheric light	0.0483	0.0481	1
Estimation of the transmission	0.4537	0.0727	6.24
Refinement of the guided filter	1.3875	0.2451	5.66
Using data to restore dehazed image	0.2932	0.0573	5.12
Image haze removal	2.1827	0.4232	5.16

Fig. 4 Performance comparison on different image size



5 Conclusion

In this paper, we have proposed a GPU-based parallel image-dehazing algorithm. The proposed algorithm adopts the same strategies as Kim's algorithm, called optimized contrast enhancement, for single image haze removal. The proposed algorithm can achieve very good haze removal effect and reduce the computational cost. Obviously, the proposed algorithm can be used as real time application for removing haze.

Acknowledgment This research was partially supported by the Ministry of Science and Technology under Grants MOST104-2221-E-126 -004 and MOST103-2221-E-126-013.

References

1. Nan D, Bi D, Liu C, Ma S, He L (2014) A Bayesian framework for single image dehazing considering noise. *Sci World J*, Article ID 651986
2. Nan D, Bi D, Xu Y, He Y, Wang Y (2011) Retinex color image enhancement based on adaptive bidimensional empirical mode decomposition. *J Comput Appl* 31:1552–1555
3. Pizer SM, Amburn EP, Austinetal JD (1987) Adaptivehistogram equalization and its variations. *Comput Vis Gr Image Proces* 39:355–368
4. Guo F, Cai Z, Xie B (2011) Videodefoggingalgorithmbasedon fog theory. *Acta Electronica Sinica* 39:2019–2025
5. Li C, Gao S, Bi D (2009) A modified image enhancement algorithm based on color constancy. *Chin Opt Lett* 7:784–787
6. Tan R (2008) Visibility in bad weather from a single image. In: *Proceedings of the 26th IEEE conference on computer vision and pattern recognition*, pp 1–8

7. Fattal R (2008) Single image dehazing. In: Proceedings of the international conference on computer graphics and interactive techniques, pp 1–9
8. He K, Sun J, Tang X (2009) Single image haze removal using dark channel prior. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition workshop, pp 956–1963
9. Matlin E, Milanfar P (2012) Removal of haze and noise from a single image. In: Proceedings of SPIE, pp 82–96
10. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans Image Process* 16:2080–2095
11. Kim JH, Jang WD, Sim JY, Kim CS (2013) Optimized contrast enhancement for real-time image and video dehazing. *J Vis Commun Image R* 24:410–425
12. Xue Y, Ren J, Su H, Wen M, Zhang C (2013) Parallel Implementation and optimization of haze removal using dark channel prior based on CUDA. *Commun Comput Inf Sci* 207:99–109
13. Valderrama JA, Diaz-Ramirez VH, Kober V (2014) Single image dehazing using local adaptive signal processing. In: Proceedings of SPIE 9217, applications of digital image processing, p XXXVII
14. Ok S, Kim M, Cho H (2014) GPU-accelerated dehazing algorithm for video surveillance. *GPC*
15. Herk M (1992) A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels. *Pattern Recogn Lett* 13:517–521
16. Schechner YY, Narasimhan SG, Nayar SK (2001) Instant dehazing of images using polarization. *CVPR* 1:325
17. Nayar SK, Narasimhan SG (1999) Vision in bad weather. *ICCV* 820
18. He K, Sun J, Tang X (2010) Guided image filtering. In: Proceedings of ECCV, pp 1–14
19. NVIDIA: CUDA Programming Guide 2.0. http://www.nvidia.cn/docs/IO/57399/NVIDIA_CUDA_Programming_Guide_2.0Final.pdf

Research of Thunderstorm Warning System Based on Credit Scoring Model

Xinli Zhou, LiangBin Yang and HaiFeng Hu

Abstract Thunderstorms pose great threat to human survival. As traditional statistics and fore-casts have great limitations, this paper applies credit scoring model into thunderstorm warning system. We select related data of the electric field as variables before and during the thunder-storms. And by actual monitoring system of thunderstorm data preprocessing, sampling, binning and so on, by using the method of logistic regression and neural network to deal with, the model based on the combination of location and data of electric field is an effective way of thunderstorm warning. From the angle of the method, the model based on the theory of credit scoring can be faster and more accurate. It can make the probability of forecasting thunderstorm more quantization.

Keywords Thunderstorm warning · Credit scoring model · Logistic regression · Neural networks

1 Introduction

Thunderstorms pose great threat to human survival. Up to now, in China, thunderstorm detection warning is achieved by a two-dimensional thunder-storm location network which is established by provincial meteorological bureaus, and the combination with weather radar and satellite images. It can achieve a wide range of thunderstorm detection and characterization of basic warning. But now, thunderstorm Location Network Technology are facing many problems, such as the scale of warning is too large, the precision is too low, the process of forecast requires a lot

X. Zhou (✉) · L. Yang
School of Information Science and Technology, University of International Relation,
Beijing 100091, China
e-mail: Zhouxinli001@126.com

H. Hu
JOZZON, Yingu Mansion, No. 9 North 4th Ring West Road, Haidian District,
Beijing 100190, China

of human judge, and so on. Therefore, in practice, often applied to the thunderstorm incident confirmation after it happened, but, it is difficult for the early warning and forecast.

In order to make full use of existing observational data, thunderstorm observation data which have already been collected must be fully collated and analyzed. The most commonly used in previous studies is the traditional statistical induction [1] with statistics of distribution of thunderstorm days, thunderstorm hours and thunderstorm density in the various regions over a specific period of time. However, this method has many limitations, the operation is time consuming. And the low statistical speed and not high accuracy are its biggest drawback. When dealing with large amounts of complex data sets, it does not work. Furthermore, thunderstorm days, thunderstorm hours and other parameters through traditional statistical methods have some limitations in practical engineering applications, and cannot reflect the distribution of thunderstorms in statistical area. Based on it, a grid statistical method of thunderstorm parameters appears [2]. It inherits the tradition of thunderstorm parameters of statistical significance. But it requires data consistency both on properties and formats. The accuracy of the statistics division of the grid in different regions also needs to be determined by the thunderstorm features of the region in the past. That also constrains rapid analysis of thunderstorm distribution characteristics.

Thunderstorm data have many features, such as a huge amount of data, property complex, strong timeliness in dynamic, ready to update the database, etc. So for thunderstorm data analysis method should be able to have the characteristics of rapid and efficient data processing, and to meet timeliness and accuracy requirements.

2 Credit Scoring Model

Credit scoring model [3] is a set of decision-making model and supporting technology which has already been used maturely in banks to help lending institutions make loans. By using large amounts of historical data to establish scoring model, banks can predict the probability of default of existing loans or those who apply for loans. The main contents are to collect various influencing factors of repayment of loans from client's application materials, Credit Bureau, bank credit database and other channels, including the applicant's monthly income, outstanding debt and financial assets, how long the applicant has been working in the current position, whether the applicant has record of default or not, whether the applicant owns a house or rents a house to live in, the current income and expenditure of the applicant's account and other information. By using data mining technology based on data statistical analysis, including discriminant analysis, logistic regression, classification trees, neural networks and so on, we establish a credit scoring model. According to scores, we can know the probability of our applicant to be a 'good

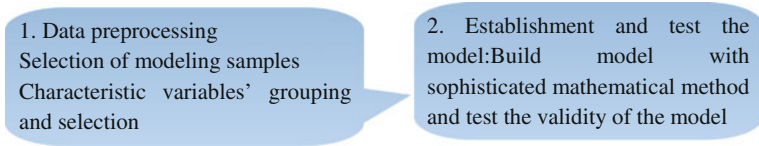


Fig. 1 Credit scoring model development process

customer’ or a ‘bad customer’. So that we can decide to accept or reject the application and provide a reference for credit line and loan pricing.

Credit scoring model is very suitable for thunderstorm prediction judgment. The efficiency of data processing can be greatly improved and the implicit rules can be found when it is used in analytical processing of thunderstorm observation data.

The current single credit scoring models have merits and demerits respectively. For instance, in statistical model, logistic regression does not require a probability distribution of characteristic variables and the same Covariance as the assumptions. The model is of great interpretability and stability. But compared to Artificial Intelligence, the classification accuracy is not so high. The neural network technology is a nonlinear technology without any requirements for data distributions in Artificial Intelligence. It has the advantage of high classification accuracy, but the shortcomings are that it is not interpretable and it is not robust while facing the changes of credit data. A combination of various models will improving the robustness of the combined system [4].

The main technical points of credit scoring model are the source of real samples and variable set-tings. A credit scoring model include two important steps (Fig. 1).

3 Data Sources: Thunderstorm Monitoring System

In order to ensure the stability and the predictive power of the model, data selection criteria are that the data used to model should be adequate and of high quality. In this paper, the data used in the model are from the Jozzon Group’s thunderstorm monitoring system in Beijing Jingshun Road. The system becomes a thunderstorm monitoring network, consisting of six thunderstorm locator and 10 electric field mills. The lightning position indicator detects real-time data and uploads to Amazon cloud computing platform. By the method of magnetic survey and TOA (time of arrival), we can work out the solution of three-dimensional position, then get the real-time location information of thunderstorms. Electric field mills monitor the surrounding atmospheric electrostatic field and upload the field strength values to the Amazon cloud platform database in real time. The system has accumulated a lot of historical data, and the electric field data were about 70 million items just in 2014. So the model provides a true and reliable source of adequate data.

Thunderstorm occurs along with a wealth of electromagnetic pulse radiation. Its frequency range is from low frequency up to high frequencies. By using

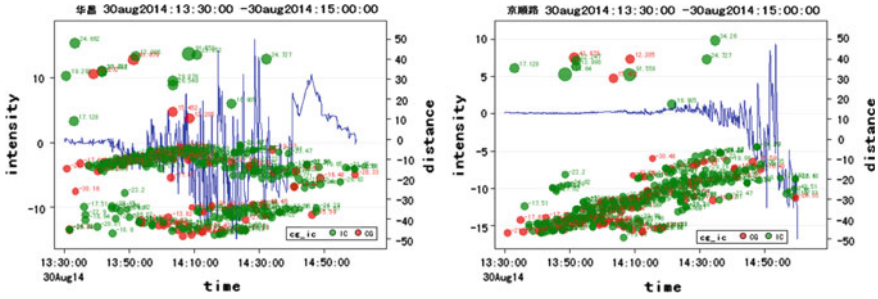


Fig. 2 The changes in the electric field before and during thunderstorms

thunderstorm location system, we can get real-time spatial and temporal distribution, intensity and polarity characteristics of thunderstorm. And these parameters are not only important for thunderstorm detection and defense, but also have an important role in early warning. The model is based on the changes of the target areas' atmospheric electric field (instantaneous change and long-term changes) and the occurrences of thunderstorms have a strong correlation. The left figure is the changes in the electric field when the thunderstorms come from afar. And the right figure is the changes in the electric field during thunderstorms (Fig. 2).

4 Modeling Process

4.1 Data Preprocessing

Data preprocessing is a critical step for providing the basis for the following reliable data mining. Data preprocessing task is to organize raw data through guidance of the defined mining theme and to get clean and accurate data to improve the quality and efficiency of data mining. The Jozzon Group's thunderstorm monitoring system in Beijing has ten electric field instruments. The amount of electric field data and location data in the original database is too large. And each of them has a table. (Electric field data in 2014 were about 70 million) So the efficiency of data processing is low. Before building a model, the table of electric field data and location data should be split into 10 tables. By doing so, the data processing efficiency has been improved. On the other hand, it is consistent with the model logic.

In this paper, the historical data from the database are to be processed in four steps.

- (a) Extracting: Export electric field data and location data from a MySQL database to SAS firstly.
- (b) Grouping: Group the electric field data by number, then calculate the latitude and longitude of thunderstorms and the distances between the locations of the thunderstorms and the electric field mills, preserving the data which are within 50 km.

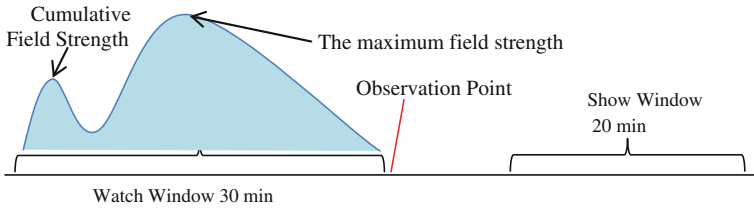


Fig. 3 The observation period and performance period

- (c) Deduplication: Remove the duplicate electric data according to the time, and remove the duplicate location data according to the time and its latitude and longitude.
- (d) Append: Append the extracted data to the original data set and form a data set.

4.2 The Definition of Variables

According to the theory of credit scoring models, generally we set two time points, observation window and show window. So it is divided into two periods—the observation period and the performance period. Observation period is to collect field information, so as to extract the periods of time which can predict the performance of the thunderstorm in the future; performance period is to collect information of thunderstorms, so as to determine whether it is a sample of thunderstorm or not (Fig. 3).

Credit scoring model establishes close contact (see above) between field characteristic variables during the observation period and thunderstorm performance during the performance period, and uses that contact to predict the possibility of thunderstorm in the future. Our concern is what kind of electric field condition will lead to occurrence of thunderstorm, and what kind of electric field indicators will not lead to occurrence of thunderstorm. Therefore, it is the foundation and the goal of building a credit scoring model to define the type of electric field based upon the occurrence of thunderstorm.

1. Independent variables: Electric field indicators

Define six variables: the maximum field strength, cumulative field strength, trends of average method, trends of difference method, number of jitter and standard deviation. Take different time windows to form a series of variables. Trend variables based on mean calculation.

2. Dependent variable

Take a radius of 7 km. The 10-min buffer zone must not have any thunderstorm strikes. Show window will be defined as 1 if thunderstorm strikes occur in 20 min, otherwise 0.

4.3 Sampling

In the credit scoring model, the amount of the sample data must be large enough. There is no single standard for the sufficient number of data. According to Lewise, “good customer” and “bad customer” sample size reached 1500 is sufficient.

Under the circumstances of sufficient amount of database, we choose random sampling method and a stratified sampling method. Random sampling is the general method of randomly taking samples according to the overall data. But, it often lacks of the thunderstorm sample. Stratified sampling method is to randomly take samples from the sample of “thunder” and “not thunder”. And the extraction ratio will be determined as required. So different types of objects can be treated differently to ensure important types of objects is in the sample sufficiently. It is more suitable for determining the sample of credit scoring model.

Overall data sorted by time and take stratified systematic sampling based on whether there is thunder or not. We take 1500 samples in each of the strata, 3000 samples were taken in total. Among them, 2000 as the training set, 1000 as the Test set. Build a model based on the training set and assess the performance of the model on the test set (Table 1).

4.4 Binning

IV (Information Value) is used for variable ordering according to the predictive power of the variables. We use IV to select indicators. WOE means weight of evidence. The risk of a particular option is converted into a linear form which can be easily understood. WOE is used to measure the relative risk. Calculate the value of WOE is to find reasonable interval division of each electric field indicator. And each interval division corresponds to a score card rating.

Binning each variable in sample set, and it is based on:

1. Minimum loss of accumulated IV (information value).

Accumulated IV is to sum the IV of all the binning. It can also be called the variable’s IV. Each IV binning calculated as follows:

Assume that the variation VAR binning m parts: B_1, \dots, B_m . There are n_{1i} samples of thunder and n_{0i} samples of not thunder in B_i . The WOE and IV can be calculated as follows:

$$WOE = \ln(p_{1i}/p_{0i}) \quad IV = (p_{1i} - p_{0i}) * \ln(p_{1i}/p_{0i})$$

Table 1 The explanation of each field in the raw data table

obs_time	Observation point
SelectionProb	The probability of selection. If the total number is 100, take 10 samples, then the probability of selection is: 10/100 = 0.1
SamplingWeight	Sampling weights. It is the reciprocal of the probability of selection
Part	Distinction between training and validation samples 1 represents the training sample, 2 represents the validation sample
Thunder7	The dependent variable of the model. 0–1 variables 0 represents that there is no thunder, 1 represents that there is thunder
The following are independent variables of the model. Variable ended with 10 represent that it is generated within the first 10 min of observation point. Variable ended with 30 represent that it is generated within the first 30 min of observation point	
std_10, std_30	Standard deviation of the electric field intensity
SUM_10, SUM_30	The sum of the absolute value of the electric field strength
MAX_10, MAX_30	The maximum value of the absolute value of the electric field strength
nshake_10, nshake_30	Jitter: if we define the point K as jitter, then, it needs to fit the following conditions d1 = intensity(k-5)—intensity (k) d2 = intensity(k)—intensity (k + 5) If abs (d1) > 0.05, abs (d2) > 0.05 and d1 * d2 < 0, then we define the point K as jitter Among them: intensity (k) means the field strength of the point K, intensity (k-5) means the field strength of the point where 5 s before the point K, intensity (k + 5) means the field strength of the point where 5 s after the point K Number of jitter
trend_10, trend_30	Trend variables based on mean calculation
trend2_10, trend2_30	Trend variables based on the calculation of extreme value

Among them,

$$N1 = \sum_{i=1}^m n1_i, \quad N0 = \sum_{i=1}^m n0_i$$

$$p1_i = n1_i/N_1, \quad p0_i = n0_i/N_0$$

2. Final binning number is about five.

There is no mandatory requirement for the choice of the number of intervals. But in reality, we usually let N range from 5 to 14.

3. WOE (weight of evidence) has a monotonic or U-shaped curve.

4.5 *Logistic Regression or Neural Network*

4.5.1 The Basic Principle of Logistic

There are many ways to build credit scoring model. Since the logistic regression model has many advantages. For instance, it can exclude the impact of individual abnormal data points; it has a strong data processing capability; it can be applied to continuous or categorical arguments; it does not require a multivariate normal distribution and the same Covariance as the assumptions; the results are easy to understand and interpret, and it is widely used in theoretical research and practical applications. By comparison, among the stability and accuracy of statistical methods, Logistic regression is an ideal method. Non-statistical methods, SVM have an advantage which is in the forefront of artificial intelligence methods, but when making scorecard, Logistic model is more effective. This paper intends to use this method.

4.5.2 Neural Networks

Neural networks are currently the most widely used models in the field of personal credit score. Their advantage is the high classification accuracy, and their disadvantage is the lack of interpretability and stability [5]. Multilayer perceptron neural network has a wide range of applications in pattern recognition, function approximation, risk prediction and control. Multilayer perceptron neural network structure consists of an input layer, output layer and the hidden layer. And the hidden layer may be one layer or a multi-layer. Former layer node and back layer node are connected through the neural network weights. There are no coupling nodes in the same layer. The activation function between the input layer and the hidden layer, the hidden layer and the hidden layer is a Sigmoid function generally, but the activation function between the hidden layer and the output layer can be a linear function.

4.6 *Model Checking*

After we have built the model, we need to test its predictive ability and stability by comparison of the difference between the forecast and the actual situation.

4.6.1 The Result of Logistic Regression by Using WOE Instead of Variable Values

Put 2000 samples into a training set and use logistic regression. The variable selection is set to be “stepwise selection method.” The basic idea of stepwise

Analysis of Maximum Likelihood Estimation					
Parameters	Degree of Freedom	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-square
Intercept	1	0.1631	0.0950	2.9466	0.0861
max_10	1	-0.9093	0.1708	28.3345	<.0001
sum_10	1	1.0009	0.1082	85.5099	<.0001
trend_10	1	-0.3879	0.1407	7.6068	0.0058
nshake_10	1	0.2188	0.0978	5.0000	0.0253
max_30	1	-0.2775	0.1392	3.9717	0.0463
sum_30	1	-0.2930	0.1471	3.9687	0.0464
nshake_30	1	0.5468	0.1052	27.0028	<.0001
std_10	1	0.7089	0.1344	27.8067	<.0001
std_30	1	0.5618	0.1144	24.1113	<.0001

Fig. 4 Analysis of maximum likelihood estimation

Analysis Variables: p						
cls	Number of Observations	N	Average	Standard Deviation	Minimum	Maximum
Effec	452	452	0.9133375	0.1355953	0.5267807	0.9995992
False	29	29	0.8350073	0.1652524	0.5036690	0.9984968
Misse	48	48	0.3101100	0.0620589	0.1908976	0.4793185
Norma	471	471	0.1145955	0.0656723	0.0023183	0.4951060

Fig. 5 Analysis variables: p

regression is to introduce variables one by one, and each time we introduce an independent variable, all variables which have been elected need to be tested one by one, and when some variables are no longer significant because of the introduction of new variables, then remove these variables. To ensure that every time variables in the model are significant before introducing new variables, every step of stepwise regression needs to do F-test. (Introduction of a new variable or reject a variable are all called one step of stepwise regression) So repeat this process until there are no significant variables are introduced into the model and no insignificant variables are rejected from the model. So the cycle repeated and the resulting subset is the optimal subset (Figs. 4 and 5; Table 2).

- Effective Alert thunderstorm strike at the show window, the probability of the observation point $p > 0.5$
- False Alert no thunderstorm strike at the show window, the probability of the observation point $p > (1-0.5)$
- Missed Alert thunderstorm strike at the show window, the probability of the observation point $p < 0.5$
- Normal no thunderstorm strike at the show window, the probability of the observation point $p < (1-0.5)$.

Calculate predicted probability of each validation sample by the model results. The second column is the quantile of predicted probability when there is no thunder in actual situation and the third column is quantile of predicted probability. When there is thunder in actual situation. Describe the distribution of the predicted probability according to the quantile. Take the second column for an example, 90 % quantile is 0.27033456, which means predicted probability of 90 % of the sample doesn't thunder is less than 0.27033456. And take the third column for an example,

Table 2 Verify the predicted probability of the sample

Quantile (Definition 5)	No thunder in reality	Thunder in reality
Quantile	Estimate	Estimate
100 % maximum	0.99849676	0.999599
99 %	0.98963365	0.999073
95 %	0.61042832	0.997533
90 %	0.27033456	0.996331
75 % Q3	0.10895120	0.991363
50 % median	0.10895120	0.981675
25 % Q1	0.07827327	0.785954
10 %	0.07827327	0.526781
5 %	0.05455687	0.294652
1 %	0.01160982	0.250032
0 % minimum	0.00231833	0.190898

10 % quantile is 0.526781, which means predicted probability of 10 % of the sample thunders is less than 0.526781. In other world, the predicted probability of 90 % of the sample thunders is more than 0.526781.

When $p = 0.5$, the accuracy rate of thundering is about 90 % and the accuracy rate of not thundering is about 94 %.

When p raised to 0.78, the accuracy rate of thundering is about 75 %, the accuracy rate of not thundering is about 75–90 %.

When p is reduced to 0.29, the accuracy rate of thundering is about 95 %, the accuracy rate of not thundering is about 95–99 %.

The following graph is drawn according to the predicted probability, and y-axis represents the predicted probability, x-axis represents the 1000 training samples (According to whether it thunders and the chronological order).

The distribution of the predicted probability p : blue represents the distribution of the predicted probability p of not thundering in reality; red represents the distribution of the predicted probability p of thundering in reality (Fig. 6).

Fig. 6 The distribution of the predicted probability p

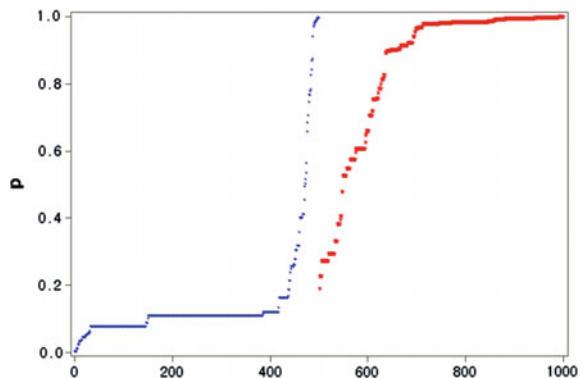


Table 3 Data role = VALIDATE, Target variable = thunder7

Target	Result	Percentage of goal	Percentage of result	Frequency	Total percentage
0	0	82.1632	96.2222	433	48.1111
1	0	17.8368	20.8889	94	10.4444
0	1	4.5576	3.7778	17	1.8889
1	1	95.4424	79.1111	356	39.5556
Omission		Specificity	False alarm	Hit the target	
94		433	17	356	

4.6.2 The Results of Neural Networks

For multilayer perceptron network, the neuron number of hidden layer is 10.

In the hidden layer, combining function is a linear function and activation function is a sigmoid function.

In the output layer, combining function is a linear function and activation function is a logistic function.

In the target layer, error function is Cauchy function and iterative method is quasi-Newton method (Table 3).

4.6.3 Comparison Results

No matter binning or not binning, logistic regression and the neural network methods, the effectiveness of testing within three methods are in an acceptable range (greater than 35 %), but according to the Table 4, it is clear that combining the binning method and Logistic regression method is the best.

Table 4 Comparison results

Calculation accuracy (p = 0.5)			
	Method	Accuracy of thundering (%)	Accuracy of not thundering (%)
After binning	Logistic regression	90	94
	MLP neural network	79	96
Not binning	Logistic regression	80	98
	MLP neural network	71	97

5 Conclusion

This thesis applies the credit scoring model to the statistical processing of thunderstorm data firstly, and achieves certain results. First of all, the results of the credit scoring model can describe the thunderstorm performance trends in different field situations more precisely and it will promote progress in the research and practice of thunderstorm prediction. In addition, data mining methods have the quick calculation and data processing capabilities. Compared to traditional statistical methods, data mining can be more quickly and more accurately in the analysis of the data, which is the ability to meet the thunderstorm Location Monitoring System's requirements of processing a lot of real-time thunderstorm data.

In summary, combining electric field data and location data to build model is an effective means of thunderstorm warning. From the angle of the method, using credit scoring method to build model can get the better results. Compared with the traditional weather forecasting model, the predictions are more specific and the probability of the occurrence of thunderstorm is more quantified.

References

1. Kang P, He J, Zeng R (2006) Statistic analysis on thunderstorm activity regularity of power supply system along railroad section from Golmud to Lhasa in railway from Qinghai to Tibet. *Power Syst Technol* 30(1):55–59
2. Chen J, Zheng J et al (2006) Statistical method of thunderstorm day. *High Volt Eng* 32 (11):115–118
3. Deng C, Wei H, Tang Y (2010) Research on dynamic small business credit scoring at home and abroad. *Stud Int Financ* 10:84–91
4. Zhou ZH, Wu J, Tang W (2002) Ensembling: neural networks: many could be better than all. *Artif Intell* 137:139–263
5. Stepanova M (2001) Using survival analysis methods to build credit scoring models. Dissertation, University of Southampton, Southampton, pp 11–18

Cloud-Based Marketing: Does Cloud Applications for Marketing Bring Positive Identification and Post-purchase Evaluation?

Ching-Wei Ho and Yu-Bing Wang

Abstract With the development of cloud applications for marketing, cloud-based marketing applications are becoming more and more important. The purpose of this study was to demonstrate whether cloud applications for marketing bring positive identification and evaluation through simultaneously considering identification with the community and the company and investigate the behavioral implications from the Facebook community members' perspective. A questionnaire investigation with consumers was conducted in this research for examining five hypotheses. The findings of this study indicated that the interaction on the Facebook brand community can enhance both C-C identifications and consumer post-purchase behaviors. Also this study focused on the 2×2 relationship with C-C identifications and consumer post-purchase behaviors which all had significant and positive effects.

Keywords Cloud-based marketing · C-C identification · Facebook community · Repurchase intention · WOM

1 Introduction

It is impossible to underestimate how much the Internet has changed the world. Online environments have radically changed all the ways we learn, shop, interact and play. According to [16], the term of “cloud marketing” encompasses all of a company’s online marketing efforts. Customers receive notices about sales through their email instead of their post office box. Shoppers can get more information about a product than a whole army of sales staff could offer. The entire business model is based on cloud marketing; using the tools of the Internet to present products,

C.-W. Ho · Y.-B. Wang (✉)
Department of Marketing, Feng Chia University,
100 WenHwa Road, Taichung, Taiwan
e-mail: icebbb@gmail.com

C.-W. Ho
e-mail: chingwei1121@yahoo.com.tw

engage with customers, and push brand messages. Another component of cloud marketing is online applications which help marketing departments operate more effectively. For example, gathering market research used to be a tedious and labor-intensive process. Now it can be done almost instantaneously and at a fraction of the cost. There are two different kinds of cloud applications that are used for marketing purposes: media and tools. Media, e.g. Facebook, Twitter, and YouTube, allows companies or brand players to present a message to the public. Tools, e.g. Google Analytics and Google Social Reports, take common marketing tasks and make them faster and easier. Of all the social networks, Facebook is the most popular and claims to have attracted over 1310 million monthly active users (as of January 2015) since starting in February 2004 [22]. Particularly in Asia, almost 90 % of Asian brands use it as a marketing platform, and 75 % of these brands have developed social media strategies that have been in use for over one year [18]. Facebook has become the top social media by number of users and volume of access or use [8]. Therefore, this current research is going to take Facebook as the study case for the applications of cloud marketing.

2 Literatures and Hypotheses Development

2.1 Cloud Applications for Marketing—Facebook

Of all the cloud applications for marketing, Facebook is the most popular and claims to have attracted over 829 million daily active users (as of June 2014) since starting in February 2004 (www.facebook.com). As one of the important cloud applications for marketing [16], more and more companies find that it is necessary to have a brand presence on Facebook where companies can create brand posts containing anecdotes, photographs, videos, or other materials, and then brand fans can interact with these brand posts by liking or commenting on them [6]. Therefore, the brand community operating on Facebook has become extremely fashionable [23]. This research intends to examine how a company's Facebook community affects consumer identification with community and company.

2.2 Consumer-Community Identification

Recent studies begin by considering the strength of the consumer's relationship with the brand community, which can be illustrated as consumer-community identification [9]. As said by [1], consumer-community identification relates to whether an individual considers himself or herself a member, namely, belonging to the brand community. When a consumer logs on Facebook platform and explores a company's brand page, comments, shares an experience, interacts with other fans,

participates activities or events, or answers comments, this consumer is participating in the community activities. In these interactions resources are being mediated and exchanged with other members in the Facebook community. Therefore, to the degree in which they support information sharing and strengthen bonds among them, Facebook community makes members feel more like insiders. Based on the above discussion, the following hypothesis is put forward:

H1: Higher levels of Facebook community interaction would lead to higher levels of consumer-community identification.

2.3 Consumer-Company Identification

Reference [5] had extended the concept of identification and developed a conceptual framework for consumer-company identification. The essential argument of consumer-company identification is that consumers identify with a company because the company they patronize at least partly satisfies their self-definitional needs, even when they are not formal members of the company [20]. That is to say, customers' needs for self-definition or a sense of belonging can be articulated through developing socially identifying relationships with a company [14]. Furthermore, consumer-company congruence has a positive effect on consumers' evaluation of a company because of their greater commitment toward the firm [15]. As such, when a consumer logs on Facebook and explores a company's brand page, shares an experience, interacts with marketers, asks questions about this company's products or services, or answers comments, this consumer is involving in the consumer-company relationship. In this relationship, meaningful experience and valuable resources are being interacted between members and the company, so the consumer-company identification could be strengthened in such community. Consolidating the theoretical arguments reviewed so far, we hypothesize that:

H2: Higher levels of Facebook community interaction would lead to higher levels of consumer-company identification.

Most previous studies have focused on identification with the brand community [2], but not many have investigated the role played by the company in establishing such relationships and, more specifically, the influence exerted by customer-company identification [17]. Moreover, almost none simultaneously considered identification with the community and the company, investigating the behavioral implications from the Facebook community members' perspective. Therefore, the current study is going to close up this gap. The hypothesis is that:

H3: Higher levels of consumer-community identification would positively influence higher levels of consumer-company identification.

2.4 *Consumer-Company Identification*

Post-purchase intentions have been normally employed as a foundation for predicting consumers' future behaviors [12]. According to [21], post-purchase intentions can be classified into social behavioral intentions (e.g. Word-of-mouth) and economic behavioral intentions (e.g. repurchase).

A research about the brand fan-page from [11] revealed that high usage intensity, get in regular contact with the brand, which in turn should have an effect on their brand relationship and should increase their likelihood for word-of-mouth or repurchase. Besides, Ref. [1] pointed out that community identification can cause both WOM (extra-role) behavior and re-purchasing (in-role) behavior. Therefore, in line with the previous research, we propose the following hypotheses:

H4: Consumer-Community identification positively affects (a) WOM and (b) re-purchase.

Additionally, members with high identification are more prone to contribute the organization with several desirable cooperative behaviors of helping other members and spreading good references [19]. Based on the concept of organizational citizenship behavior [4], consumers are more likely to express their support for an organization, such as a company, by engaging in in-role behaviors, e.g. purchasing products [1] and extra-role behavior, e.g. making recommendations in positive WOM [3]. As such, when customers identify themselves with the organization's vision and value, they are interested in the growth of the organization. Consequently, they express positive behaviors such as encouraging word-of-mouth [10] and re-purchasing. Therefore, in line with the previous research, we propose the following hypotheses:

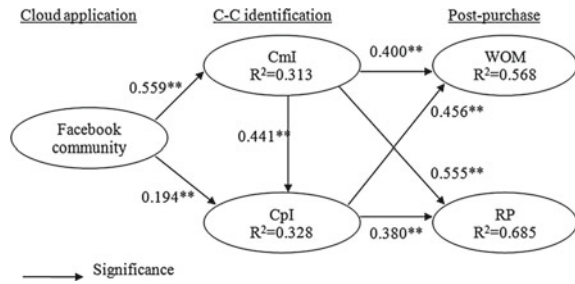
H5: Consumer-Company identification positively affects (a) WOM and (b) re-purchase.

3 **Research Method and Results**

Data were collected by a structured questionnaire developed for the research, adapting those used in previous studies. As the target population in this study consists of people who are members of Facebook communities, the questionnaire was distributed through several posts in social networking sites, such as Facebook. A total of 270 questionnaires were collected with 19 missing values. That is to say, 251 fully completed questionnaires were used for the data analysis. This study mainly used partial least squares (PLS) to test the hypotheses and analyze the data.

Facebook community exerted a significant and positive influence on both consumer-community identification ($H1, \beta = 0.559, p < 0.01$) and consumer-company identification ($H2, \beta = 0.194, p < 0.01$). Therefore, H1 and H2

Fig. 1 Results of the structural model analysis
(Note $**p < 0.01$)



are both supported. The model predicted the path from consumer-community identification to consumer-company identification (H3) and showed that there was a significant and positive relationship between them ($\beta = 0.441, p < 0.01$). So H3 gains supported. Besides, in this test, consumer-community identification indicated a partial mediating effect on company’s Facebook community and consumer-company identification ($\beta = 0.247 > \beta = 0.194$). Furthermore, the paths from consumer-community identification had a significant and positive influence on both word-of-mouth (H4a, $\beta = 0.400, p < 0.01$) and re-purchase (H4b, $\beta = 0.555, p < 0.01$). Meanwhile, the paths from consumer-company identification had a significant and positive influence on both word-of-mouth (H5a, $\beta = 0.456, p < 0.01$) and re-purchase (H5b, $\beta = 0.380, p < 0.01$). Thus, both H4 and H5 are fully supported.

The findings which show in Fig. 1 offer important contributions and implications for both marketing academics and practitioners.

4 Discussions and Conclusions

These results gained from this study offer important contributions and implications for both marketing academia and practitioners. First, this study reveals that consumer interactions with company’s Facebook community have directly positive and significant effects on both C-C identifications (i.e. community and company). Meanwhile, consumer-community identification played a role as a mediator between company’s Facebook community and consumer-company identification. These findings show that the cloud application for marketing, i.e., Facebook, is to bring people with certain similar characteristics together and to facilitate interactions among them. It is consistent with [7, 13] research on brand communities in social media. Besides, this study proves that Facebook community can make fans consider more like insiders (i.e. consumer-community identification) and strengthen connections with consumer-company identification.

Second, the 2×2 relationship with C-C identifications and consumer post-purchase behaviors are all significant and positive but with slightly different effects. The effect on the path between c-community identification and re-purchase

is the strongest ($\beta = 0.555$), while the path between c-company identification and re-purchase is the least effect ($\beta = 0.380$). These relationships have not been found from before literatures and this finding could be viewed as pioneering, setting a benchmark for further research.

Third, in previous studies, most of them focused on identification with the brand community [2] or consumer-company identification [5] respectively. This study proposed an exclusive model of the process by which we can simultaneously consider both C-C identifications (i.e. identification with the community and the company), and investigate the post-purchase intentional behaviors from the Facebook community fans' perspective. We have tested and validated this model and found support for the hypotheses in the context of online community. This finding also could be viewed as pioneering, setting a benchmark for further research.

Finally, this study reveals that cloud applications for marketing (e.g. Facebook in this case) really bring positive identification (for both community and company) and evaluations of the brand.

The following limitations of this study should be addressed. When addressing these limitations, we also suggest directions for future research. First, this study adopted both C-C variables, namely consumer-company and consumer-community identification, as antecedents of post-purchase behaviors. Next time, another C-C variable, namely consumer-cloud identification, could also be considered and examined as antecedents. Second, our sample comprised primarily young adults (under 30 years old); hence, their responses may not be completely generalizable to the population at large. Finally, this study examined a specific cloud application of cloud marketing, namely Facebook, so the results could not be ascribed to other applications of cloud-based marketing. Future researchers could explore both C-C identifications with regard to different types of cloud applications with a specific company or brand settings.

References

1. Ahearne M, Bhattacharya CB, Gruen T (2005) Antecedents and consequences of customer-company identification: expanding the role of relationship marketing. *J Appl Psychol* 90:574–585
2. Algesheimer R, Dholakia UM, Herrmann A (2005) The social influence of brand community: evidence from European car clubs. *J Market* 69:19–34
3. Anderson EW, Fornell CF, Mazvancheryl SK (2004) Customer satisfaction and shareholder value. *J Market* 68:172–185
4. Bateman TS, Organ DW (1983) Job satisfaction and the good soldier: the relationship between affect and employee “citizenship”. *Acad Manag J* 26:587–595
5. Bhattacharya CB, Sen S (2003) Consumer-company identification: a framework for understanding consumers' relationships with companies. *J Market* 67:76–88
6. de Vries L, Gensler S, Leeflang PSH (2012) Popularity of brand posts on brand fan pages: an investigating of the effects of social media marketing. *J Interact Market* 26:83–91
7. Fourmier S, Avery J (2011) The uninvited brand. *Bus Horiz* 54:193–207
8. Hsu YL (2012) Facebook as international eMarketing strategy of Taiwan hotels. *Int J Hosp Manag* 31(2012):972–980

9. Huang HC, Chang CW (2007) Building brand community: a study of VW's club. *Taiwan Bus Perform J* 1(1):1–26
10. Hur WM, Ahn KH, Kim M (2011) Building brand loyalty through managing brand community commitment. *Manag Decis* 49(7):1194–1213
11. Jahn B, Kunz W (2012) How to transform consumers into fans of your brand. *J Serv Manag* 23(3):344–361
12. Kuo YF, Wu CM, Deng WJ (2009) The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Comput Hum Behav* 25(4):887–896
13. Laroche M, Habibi MR, Richard MO (2013) To be or not to be in social media: how brand loyalty is affected by social media? *Int J Inf Manag* 33(1):76–82
14. Mael F, Ashforth BE (1992) Alumni and their Alma mater: a partial test of the reformulated model of organizational identification. *J Organ Behav* 13:103–123
15. Martín L, Ruiz S (2007) “I need you too!” Corporate identity attractiveness for consumers and the role of social responsibility. *J Bus Ethics* 71:245–260
16. Marketing-Schools.org (2012). Cloud marketing. <http://www.marketing-schools.org/types-of-marketing/cloud-marketing.html>. 25 Mar 2015
17. Marzocchi G, Morandin G, Bergami M (2013) Brand communities: loyal to the community or to the brand? *Eur J Mark* 47(1/2):93–114
18. Pon YP, Wang CJ (2012) Which keyword let brand control the digital trend? *Brand News*. <http://www.brain.com.tw/News/RealNewsContent.aspx?ID=17819>. 06 Nov 2012
19. Qu H, Lee H (2011) Travelers' social identification and membership behaviors in online travel community. *Tour Manag* 32:1262–1270
20. Scott SG, Lane VR (2000) A stakeholder approach to organizational identity. *Acad Manag Rev* 25:43–62
21. Smith AK, Bolton RN, Wagner J (1999) A model of customer satisfaction with service encounters involving failure and recovery. *J Mark Res* 36(3):356–372
22. Statistic Brain (2015) Facebook statistics. <http://www.statisticbrain.com/facebook-statistics/>. 25 Mar 2015
23. Trusov M, Bucklin RE, Pauwels K (2009) Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *J Market* 73(5):90–102

Decision Analyses of Medical Resources for Disabled Elderly Home Care: The Hyper Aged District in Taiwan

Lin Hui and Kuei Min Wang

Abstract Facing the fastest ageing fact, Taiwan implements long term care (LTC) program to lessen the load of some demanded home-care disabled-elderly-family. However, the provided medical resources, including the physician, nurse and physiotherapist (PT) would never catch up to the demands for home care. Using Linear Programming to model the current and the alternative home care medical demand-supply situations, it provides the decision maker with a deep insight of the resource allocation problem. The scenario includes three hospitals around the Meinong District in Kaohsiung to be the medical givers. Study shows that current medical resources are insufficient. PT and nurses are in serious shortage. The optimal solution of (physician, nurse, PT) is (15, 27, 22). Elderly under the age of 65 and local home care facilities are excluded in this study.

Keywords Disabled elderly · Medical resource · Linear programming · Model · Demand-supply

1 Introduction

According to Florian Coulmas [4], there are three different types of society based on the proportion of elderly over 65 years old. The categories are as follows. Aging society: 7–14 % of the population are 65 years or older; aged society: 14–20 % of the population are 65 years or older; hyper-aged society: 20 % or more of the population are 65 years or older.

L. Hui

Department of Innovative Information and Technology, Tamkang University,
26247 Yilan County, Taiwan
e-mail: amar0627@gmail.com

K.M. Wang (✉)

Department of Information Management, Shih Chien University,
84550 Kaohsiung, Taiwan
e-mail: willymarkov@gmail.com

According to population policy white paper [10], the ageing index of Taiwan will reach 93.5 % in 2015 and 129.2 % in 2020. This shows that the demographic transition to the fast ageing society is becoming a great challenge not only for the work force and the economy but also the elderly medical care system.

A survey by formal Department of Health in 2010 shows the rate of disabled elderly (age over 64) is 15.42 %. Adapting to the local custom that there are over 80 % of elderly wishing to live with their family or nearby their children so that they may receive care from their family, the government's policy for elderly is in-place. Since 1998, the government has promoted the Long-term Care Programs (LCP) to cope with the ageing storm. The ten-year long care program was initiated in 2007, with the general purpose of providing service to groups of mentally and physically disabled elderly who passed the ADL screening at home or in community.

The objective of this study is to focus on the supply-demand issue for the selected hospitals that are providing medical service to the demanders, who are disabled elderly in the targeted area, Meinong District of Kaohsiung city in Taiwan. Linear Programming (LP) is the mathematical method used to develop a model for analyzing this supply and demand issue.

2 Literature Review

The ageing era has been emerging to the surface of Taiwan so fast that studies and solutions for coping with the unique future are a must for this country. The literature selected for this study include the views in ageing trend, ageing health and care, government's action, medical resource demanded, and LP allocation model in providing solutions for medical demand-supply issue, etc.

The ageing issue going on globally is a fact. However, Taiwan has popped up as one of the fastest ageing countries in the world. Hwang [8] indicated that the rate of population aging in Taiwan is estimated to increase 10 % every 15–20 years, i.e., the elderly population in Taiwan is predicted to grow at a rate of 0.6 % per year. The number of the elderly was 1.47 million in 1993 which will climb up to 3.45 million in 2018 and 6.84 million in 2040 [1].

In fact, there are 23 Districts in Taiwan that have already entered the hyper aged society ahead of the other areas. Meinong District at Kaohsiung City, the target area in this study, is one of them [3]. Ageing also brings the health problem that could be fatal to elderly. The latest investigation with statistic by Ministry of Health and Welfare (MOHW) shows the prevalence of chronic disease for the elderly are as Table 1. The top one that causes of death for elderly is chronic disease, according to MOHW. The top five of these chronic diseases are hypertension (46.7 %), cataract (42.53 %), heart disease (23.9 %), gastric ulcer (21.17 %) and arthritis or rheumatism (21.11 %). The chronic diseases are the major cause of elderly disability.

Table 1 The prevalence of chronic disease for elderly in Taiwan

Age	Number of chronic disease		
	1	2	3
≥65	88.7 %	71.7 %	51.3 %
≥75	90.9 %	76.8 %	56.4 %

In order to recognize the differences among the elderly, the population can be further divided into the following four groups: pre-old (aged 40–64 years), young-old (65–74), middle-old (75–84), and oldest-old (>84), according to Chen [2]. Based on statistics of MOHW, the prevalence of elderly disability in Taiwan is the following: young-old is 7.29 %, middle-old is 20.44 % and oldest-old is 48.59 %. The traditional way of living of the elderly is to live with or live nearby the children so that their daily life can be well taken care of by their family. A survey [14] shows that living with family in Taiwan has increased from 59.95 % in 2005 to 68.46 % in 2009. The willingness of living alone has decreased from 11.32 % in 2005 to 6.85 % in 2009. For home care of the elderly, there are 16.8 % of elderly with problems of autonomy. They mostly depend on their children (48.5 %), spouse (20.2 %) and nursing workers (16.6 %) to take care of them.

The ten year long care program (LTCP) was triggered in 2007 by the Central Government of Taiwan for initiating a policy in coping with the arrival of the ageing era with the responsive alternative in integrated perspective. The aim of this program is to provide assistance for the elderly' daily living activities such as home care, day care and adult placement. It also introduced new international models of elderly care such as group home and unit care to Taiwan. Same as the Western world and Japan, Taiwan also adopts the ageing in place policy. The objective of ten-year LTCP is to establish comprehensive long term care system, where it is able to insure the disabled residence the capability to acquire the appropriate services, to enhance the ability in self-care independently, to improve the living quality, and to maintain their dignity and autonomy. The qualified residents who may receive the services include: the elderly whose age is over 64, 55–64 year old aboriginal, 50–64 year old disabled and the elderly who lives alone with IADL compatible condition. The service includes care, home care, community and home rehabilitation, etc. [11].

Home care is the issue in this study including at least three medical categories. Jia pointed out that in this ten-year LTCP, home care is the service provided by the government, where it is the model of delivering the medical care to the selected elderly by a medical team, which includes a physician, nurse, physiotherapist, medical social worker, occupational therapist, language therapist and dieticians [9].

The medical resource allocation is one of the optimization issues; hence using a right method is essential for seeking out the right solution. There are many available quantitative methods for medical resources analyses. However, According to statistics of Fleissner and Klementiev [6], the most frequently used quantitative models in healthcare fields are 52 in total and grouped into five categories including 11 for each of simulation models and quantitative economic models, 24 of optimization models (LP and Non Linear programming), 6 of Markov models and 10 for others.

Earnshaw and Dennett [5] supposed that the health care system is a gigantic and complex dynamic system where the scientific method is used. Studies [12, 13] showed that LP is one of primary methods for solving medical resource allocation. It can contain the extremely large number of variables and constraints. In fact, LP is a method for finding the best solution for a given mathematical model satisfying certain constraints by maximizing or minimizing an objective function, subject to linear equality and inequality constraints.

3 Scenario

The ageing problem has been a problem for Taiwan that may be a one-way trip. MOHW shows that the latest Taiwan ageing ratio has achieved 11.5 % and predicts that it will be over 14 % in 2017. The targeted District, Meinong, according to latest information [7], has crossed over the 20 % ageing threshold and is approaching 21.7 %, which is 9077 out of 41,668 total residents. It is becoming one of the hyper aged areas in Taiwan. The disabled elderly is estimated to be 1480, distributed into three groups: young-old (344 residents), middle old (719) and oldest-old (417).

Medical resources that can provide Medical care include Cishan hospital, E-DA hospital and Chang Gung hospital.

3.1 Assumptions

Assuming the hospital of Cishan, E-DA and Chang Gung, represented by A, B, C, are able to provide the medical care to the Meinong District.

This action does not concern the other medical facilities, which are smaller in size. Two groups of disabled elderly at home may be determined to have a severe or less severe condition. The severe group is assumed to be over age 84 and the less severe group is from 65 to 84. The total number of elderly in each group, based on current information, is 1063 and 417.

3.2 Concept of Operations

Based on the policy of MOHW, only the disabled elderly who are approved by the authorized appraisal unit can acquire government funding support to the demanded medical care at home. However, the number of visits by the medical resources to these disabled elderly is confined by the following rules:

The maximum number of cases for a physician to handle per day is eight, but the fee of each physician visit would be deducted when the number of cases is over four. There should be no more than 180 cases per month. Each case can only have

Table 2 Medical home care frequency demanded by homecare elderly

	Serious	Less serious
Doctor	n_{11}	n_{12}
Nurse	n_{21}	n_{22}
PT	n_{31}	n_{32}

one visit by a doctor, but without a time limit if the patient is in a critical condition. The maximum number of cases for a nurse to visit per month is one hundred. It is acceptable for a nurse to handle more than one hundred cases, but the visit fee would be at a 40 % discount once it exceeds 100. The maximum number of a Physiotherapist (PT) visit per case each year is six, but no more than once a week. The visiting fee (NT dollar) per case for each medical category is 1035 for a physician, 1370 for a nurse and 1000 for a PT.

In an operation, the first concern is to set up the maximum medical capability in terms of full pay without discount, i.e. the physician can pay at most four visits per day, nurse can have no more than one hundred visits per month and the PT can have sixty visits per month. Time unit in analysis is on a two-month basis. Therefore the number of visits by a doctor/nurse/PT every two months is 176/200/120. It implies the cost of each medical category, C1, C2 and C3 to be 182,160, 260,000 and 120,000.

In addition, the demand of the disabled elderly staying at home is the key factor in this supply-demand study. The disabled elderly are divided into two groups: serious or less serious condition. The serious group may receive more care from the hospital than the group of elderly with less serious conditions. Table 2 shows the medical demand from the disabled elderly, where n_{ij} stands for the number of specific medical resource demanded by these two groups of disabled elderly. The medical resource required by the serious group is two-times more than the less serious group.

4 Modeling

Conceptually, this is a supply-demand analysis by the Linear Programming model, where the objective is to maximize the distribution of medical resources to as many disabled elderly at home as possible. The model includes set, parameter, decision variables, objective function and restraints.

4.1 Definition

The set, parameter and decision variable are defined as the following:

SET

E Elderly group

R Medical resource categories: i.e.) doctor, nurse and PT

H Designated hospitals

Table 3 Definition of the decision variables (Required number of medical category from each hospital)

Hospital category	A	B	C
Physician	X_{11}	X_{12}	X_{13}
Nurse	X_{21}	X_{22}	X_{23}
PT	X_{31}	X_{32}	X_{33}

PARAMETER

- $L_{R,H}$ the maximum number of medical category provided by each hospital
- $N_{R,E}$ the number of medical care (per two months) demanded by the disabled elderly at home
- $M_{R,E}$ the number of visits by each R
- C_R charge of medical care

DECISION VARIABLE

- $X_{i,j}$ Total number of medical resources which can be provided to the disabled elderly.

The decision variable used to represent medical resource is X, which includes doctor, nurse and PT supplying by hospitals that are assigned to supply the targeted area. Definition of the variables is as Table 3.

4.2 Objective Function and Constraints

The model’s objective is to maximize the resource utilization. The purpose is to benefit as many disabled elderly at home as possible to see how much it will cost in the worst case scenario. Therefore, the objective function is expressed to sum up the cost of each medical category from each hospital already multiplied by the number of medical visits to the disabled elderly. The mathematical expression of objective is as follows,

$$Max \sum_{i=1}^R \sum_{j=1}^H C_i X_{ij} \tag{1}$$

The constraints of the model are the relationship of demand and supply as follows:

Constraint 1 is to examine the resource that the medical care can supply to the elderly demand: each medical category in the hospital should be no more than the maximum number that the hospital can provide. i.e.

$$X_{r,h} \leq L_{r,h} \quad \forall e \tag{2}$$

Constraint 2: the total number of medical category for the disabled elderly should be no greater than their demand. The demand of each medical category is the number $N_{R,E}$, which is the sum product of the two groups of disabled elderly groups and their demand, divided by the total number of visits by the medical category. This can be expressed as

$$\sum_h X_{r,h} \leq \frac{N_{r,e}}{M_{r,e}} \quad \forall e \tag{3}$$

$$X_{r,h} \geq 0 \tag{4}$$

5 Analysis

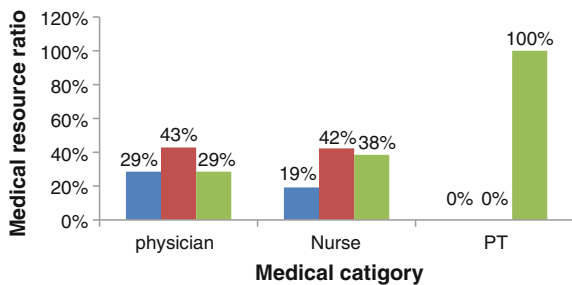
Before reaching the analysis of optimal solution to satisfy the demanders, it is necessary to look into the situation of medical supportability of hospitals.

5.1 Analysis of Medical Distribution Under Concept of Hospital Cluster

Let's denote the medical resource as the set [X11, X12, X13, X21, X22, X23, X31, X32, X33]. The first analysis is to consider the effectiveness of a hospital cluster. Assume the three hospitals are fully capable of delivering the home care medical support. This means that each of hospital's staff is more than enough. The assumed number of physician/nurse/PT is 10/15/10.

The current situation is that this is not a hospital cluster. Among them, there is no coordinated relationship and any of them can deliver the medical service to the demander. This is very similar to the competition of the business market. In Fig. 1,

Fig. 1 The distributed medical resources from each hospital without clustered



the numbers do not belong to a specific hospital. In other words, it can be randomly picked by any of the hospital as long as it is faster than the others. For example, 42 % of nurse can be any hospital’s staff.

5.2 The Current Demand-Supply Analysis

Let the current situation be the base case with the set [1, 2, 4, 7, 8, 14], representing the total medical resource the hospitals currently have available. The major work of medical resources focuses mostly on the hospital’s internal operation and with less attention to home care. As for what ratio of the hospital’s resource should be split out to support home care demand, the answer is still ambiguous (not yet defined) so far. Hence, let’s define the ratio of medical resources that can be assigned to the demand of home care, which is assumed to be 25, 50 % on the base of currently available medical resources. Table 4 shows three cases developed by the current information and case 3 is the base case. To have the medical personnel defined in fraction number is not reasonable at first glance but it becomes recognizable when we transform it into the number of visits to the disabled elderly living at home.

Using the cases in Table 4 with the given medical resources, the result calculated by LP model is shown in Table 5.

In case 1, with 25 % of the hospital’s medical resource for home care, the number of physician/nurse/PT is 6/3.25/1 which creates a shortage gap of $-9/-22.7/-21$ as shown Fig. 2. The extremely scarce medical category is PT. Case 1 is the worst-case scenario for the disabled elderly living in the Meinong District.

The shortage ratio for these three cases is shown in Fig. 3. If hospitals use 25 % of their medical resource for supporting the elderly, the shortage of doctor/nurse/PT will be 60/88/95 %. According to Fig. 2, among medical categories, the shortage of

Table 4 Cases of available medical resources

Hospital	A			B			C		
	Doctor	Nurse	PT	Doctor	Nurse	PT	Doctor	Nurse	PT
Case 1	0.25	0.5	0.25	3.5	1	0.25	2	1.75	0.5
Case 2	0.5	1	0.5	7	2	0.5	4	3.5	1
Case 3	1	2	1	14	4	1	8	7	2

Table 5 The delivered medical resources to home cared elderly and total cost

	X ₁₁	X ₁₂	X ₁₃	X ₂₁	X ₂₂	X ₂₃	X ₃₁	X ₃₂	X ₃₃	Total cost (NTD)
Case 1	0.25	3.36	2	0.5	1	1.75	0.25	0.25	0.5	1,986,918
Case 2	0.5	6.72	4	1	2	3.5	0.5	0.5	1	3,973,835
Case 3	1	6	8	2	4	7	1	1	2	6,592,400

Fig. 2 The current demand-supply situation (Case 1)

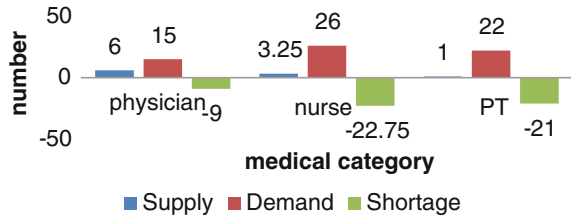
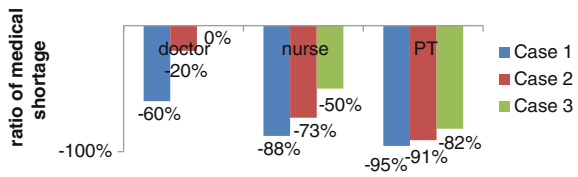


Fig. 3 The medical resource shortage in terms of demand



physicians can be easily resolved. This implies that the denominator is very small. PTs are completely disregarded. Nurses seem to be better, but the shortage situation is still neighboring to PT. In these three cases, nurses and PT have no chance to make up for satisfying the demand.

So far, the medical demand-supply is still unequal in these three assumed cases. The number of medical category is required to increase. But “how much medical resource is enough?” is a question in need of an answer. We need to find out the optimal solution for satisfying the current medical demand.

5.3 Analysis of Finding the Optimized Solution

Let’s assume that case 4 is for increasing the medical support resource from hospitals with the objective of satisfying the demand. When the distributed medical personnel numbers match the demand, the number is the solution to this question. In case 3, the adjusted category is stressed on the nurse and PT. The available support medical personnel from each hospital, denoted as (doctor, nurse, PT)_i where i represents the hospital name in case 4 are (10,15,10)_A, (10,15,10)_B, (10,15,10)_C.

Using the LP model, the optimal solution that satisfies the disabled elderly home care demands: [X₁₁, X₁₂, X₁₃, X₂₁, X₂₂, X₂₃, X₃₁, X₃₂, X₃₃] equals to [10, 5, 0, 15, 11, 0, 10, 10, 2] and costs at most \$12,132,400. The optimal solution for the medical resource devoted to Meinong is (physician, nurse, PT) = (15, 26, 22).

If the budget from the government is enough for the disabled elderly home care case, then the left over will be \$10,405,482, \$8,418,565, \$5,800,000 with respect to the first three cases.

The number of medical resource for each hospital which can satisfy both hospital internal processes and home care needs can be calculated based on the optimal

Table 6 Full hospital medical staff required in order to run both internal operation and the support of disabled elderly home care

Hospital	A			B			C		
Medical category	Doctor	Nurse	PT	Doctor	Nurse	PT	Doctor	Nurse	PT
25 %	8	16	12	6	10	10	4	6	5
50 %	9	17	14	6	11	12	4	6	6

solution, as shown in Table 6. These numbers imply the current situation where the insufficient medical resource may hurdle the implementation of the ten-year LTC program.

6 Conclusions

In addition to the fact of insufficient medical resource devoted to the disabled elderly home care, there are two major findings in this study. First, the disabled elderly home care medical demand-supply issue existed not only in the Meinong District but also the other twenty two areas in Taiwan. Some of them have hospitals or specific medical facilities around but some do not. How this medical support shortage problem can be solved is the key to the success of the ten-year LTC program. Second, the trend of ageing will not stop but speed up; mathematical models may provide the decision maker the insight of future medical demand-supply situations with feasible alternatives This study has excluded disabled patients under the age of 65 and the smaller hospitals that may also available for providing the limited medical care. All of these will be included in our future work with the calibrated model.

Acknowledgments The authors express the appreciation to Cishan Hospital in Kaohsiung City for their support.

References

1. Chen B, Chang CC, Chen C (2014) The estimate of long term care demand on Taiwan—applying GEMTEE model. The public policy conference of Academia Sinica member-Tzong-Shian. Taipei: Institute of Economics, Academia Sinica (in Chinese)
2. Chen CY (2010) Meeting the challenges of eldercare in Taiwan’s aging society. *J Clin Gerontol Geriatr* 2–4 (in Chinese)
3. Chiou ST (2011) Promoting active life of elderly from demographic changes perspective. Health Promotion Administration, Ministry of Health and Welfare, Taipei (in Chinese)
4. Coulmas F (2007) Population decline and ageing in Japan: the social consequences. Routledge, New York

5. Earnshaw SR, Dennett SL (2003) Integer/linear mathematical programming models: a tool for allocating healthcare resources. *Pharmacoeconomics* 21(12):839–851
6. Fleissner P, Klementiev A (1977) Health care system models: a review. Laxenburg, Austria: RM 77-49 ((International Institute for Applied Systems Analysis)
7. Household Registration Office (2014) Population of Meinong District. Kaohisung City government, Kaohisung (in Chinese)
8. Hwang MN (2008) Aging society: emerging issues and perspectives from the Republic of China. *Longevity Prod Experiences Aging Asia* 34–45 (in Chinese)
9. Jia SL (2004) The establishment of home-care facility. *Taiwan elderly medical society News*, pp 18–26 (in Chinese)
10. Ministry of Interior (2013) Population Policy White Paper. Ministry of Interior, Republic of China, Taipei (in Chinese)
11. Ministry of Interior (2007) The ten year long term project. Executive Yuan of Republic of China, Taipei (in Chinese)
12. Propoi A (1978) Optimization models in health care system planning. International Institute for Applied Systems Analysis, Austria
13. Sartipi K, Archer N, Yarmand M (2011) Challenges in developing effective clinical decision support systems. In-Tech Open Access Publishing
14. Statistical division of Ministry of Interior (2009) The situation investigation of elderly in 2009. Ministry of Interior, Taipei (in Chinese)

The Research About Vehicle Recognition of Parallel Computing Based on GPU

Zhiwei Tang, Yong Chen and Zhiqiang Wen

Abstract Vehicle recognition is the important content of intelligent transportation system, there have been many researches on vehicle recognition, and the technology of vehicle recognition based on CPU and DSP cannot meet the needs of the present. This article is about the study of Vehicle recognition and how to realize the GPU algorithm on the CUDA transplantation, make the algorithm parallel, thus speeding up the computation efficiency of vehicle recognition. This thesis is based on the Jeston TK1 development board as the experimental object, achieving high efficiency of GPU image processing.

Keywords GPU · Image processing · Parallel computing · Vehicle

1 Introduction

Intelligent Transportation System (ITS) is an important part of modern transportation; it will be the important direction for future development of science and technology. And the vehicle recognition is the important part of ITS. It is mainly used in highway toll collection, parking management; traffic supervision and so on [1]. There have been many kinds of technology about vehicle recognition. The technology based on image processing is widely used as it is simple and feasible. The Back Propagation (BP) neural network algorithm is a kind of vehicle recognition based on feature extraction. We will remove background noise and extraction the vehicle features. And then use the three layers BP neural network to make the vehicle

Z. Tang · Y. Chen (✉)

The Third Research Institute of Ministry of Public Security, Shanghai, China
e-mail: 99chaoyang@163.com

Z. Wen

Shanghai University, Shanghai, China

identification [2]. However, in the complex environment, it requires accurate and timely, the algorithm is complex and takes much time. If only using the CPU processing has been difficult to meet the current needs. So it is necessary to use the GPU to do the vehicle recognition. The performance of GPU is increasing at a rate 2.8 times per year. It is faster than the CPU performance of 18 months to double the much faster. The application field of GPU is mainly in virtual reality, computer simulation, computer games and so on [3]. GPU plays a very important role in image processing, if the CPU on the image processing algorithm is transplanted to GPU, the speed will be faster. This paper will introduce the research and development of vehicle recognition first, and then will focus on the introduction and implementation of BP neural network algorithm in recognition of vehicle type. There are many kinds of vehicles on the market; there are cars, vans, trucks, tricycle, crane and truck mixer and so on various models. What we have to do is mainly to solve the three kinds of models are common on the market, namely the cars, vans and trucks. Finally, this article will introduce how to realize the vehicle recognition on Linux based on Jeston TK1.

2 The Introduction the Vehicle Identification Technology

Vehicle identification is the important part of intelligent transportation system; the technology has been relatively mature. In general the vehicle type recognition main recognition car, bus and truck three models. It can achieve the vehicle recognition by some methods. First, it can use the radio waves or infrared, in short, it uses the infrared to scan the body shape of the vehicle, and then analysis the body shape to make sure what kind of vehicle it is; Secondly, through the radar to detect vehicles, the main principle of vehicle recognition is the Doppler effect, but the disadvantage of this method is that the cost is relatively high, the technology is relatively more complex; Thirdly, it can measuring the weight of vehicle, the main principle is to detect the gross vehicle weight and axle load, this method often needs to use the other methods together; Fourthly, the widely used method is the induction coil, when the vehicle across the load, the induction coil will have the waveform, and the different vehicles will have different waveforms [4]. The last method is based on image processing method to detect vehicle. With the development of science and technology, image processing technology is developing rapidly [5]. The application of image processing in machine vision has become more and more mature. The main principle of the machine vision recognition models is to use the roadside cameras. The cameras catch the image of vehicle, after image processing to extract characteristic value, at last, through pattern recognition it will know what kind of vehicle it is.

3 The Study of Vehicle Recognition Algorithm

The most important part of vehicle recognition is how to recognition what kind of the vehicle is after get the body information. The machine vision is different from the people’s visual, the machine only know the number “0” and “1”. After the camera gets the picture, it will be transformed to the digital signal by the image processing. The image processing includes the grey image, image segmentation and the image in painting. At last the picture will be something like the Fig. 1 [2].

The key to this article is not the image process ignite point is how to realize what kind of the vehicle is after image processing: big, middle or small. The researcher wills analysis the ratio between height and length. Just like the Fig. 2.

The BP Neural network algorithm is the other method that widely used in this kind of project, the researchers will construct the structure of neural network according to their own needs [6]. In this paper, we used three layers BP network, the input nodes and output are both for 3 [2]. We will see that different kind of vehicle have different ratio, there are three ratios: the top to the length, the top to the height and the front to the back. The machine will analysis these dates to discriminate the vehicle are small, middle or big.

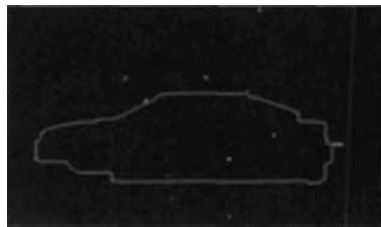


Fig. 1 The body characteristics of vehicle

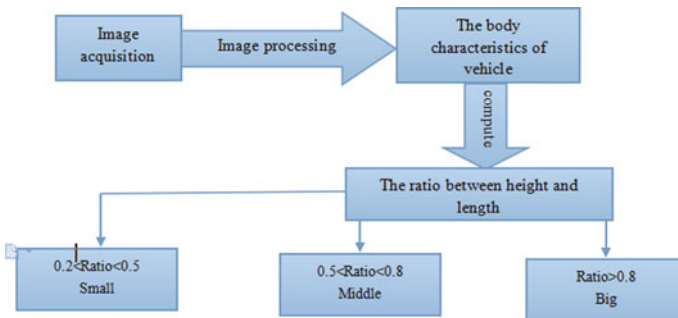


Fig. 2 The block diagram

4 The Realization on Gpu Parallel

Whether it is based on image processing to recognition on matlab or based on BP neural network algorithm, it needs the complex compute on image processing. The image processing includes moving target detection, image gray, image enhancement, image filtering, edge detection, image inpainting and contour extraction [7]. In the past, the researchers will achieve this purpose on pc, like matlab or CPU. These are success in theory, we need to put these theory into practice by put it into the development board. In past, people will use DSP to work for it. In this paper, we will introduce a new development board based on arm which will make the algorithm paralleled like GPU. GPU has powerful multithreaded and parallel processing ability. It will reduce the time on computing.

4.1 The Development Board and Environment Built

In this paper, the development board is the NVIDIA Company released in March 2014 named Jeston Tegra K1. It is claiming to be the first embedded super computer in the world, it is based on Linux, it has 192 CUDA kernel core. And it will be widely used in robot, automatic car and machine visual. The Fig. 3 is the shape of the board.

This board only supports the 64 bit Ubuntu- 12.04 operating system, in this system; we need to download the CUDA6.0 for Ubuntu. We can see the board has almost the entire interface we would use. It has the USB3.0, mini PCIe, HDMI1.4, RS232 serial port and the Gigabit Ethernet. We can develop many things on this board; however, we just use it for the machine vision.

First, it should be connect with pc by the USB and the cable, it should be worthy of attention that the cable is crossover. And then we use the VGA-HDMI line to connect



Fig. 3 Jeston TK1 board

the display, then we can see the Ubuntu on the display. Before we work with the project, we should download the CUDA for the board which is called the CUDA for Tegar could be found on the website of the NVIDIA. It does not need the graphics card of NVIDIA, because the project will be built by cross compiler. Before the compile, the researcher should SCP some library files from the board. The cross compiler means that it uses the pc to build the project but run on the board.

4.2 The Algorithm Parallel

CUDA is the most the most simple and effective to achieve the GPU parallel, it is developed by NVIDIA Company. Its language is just like the C/C++ language. In CUDA language, it would set up a kernel to make the function parallel. This function should only run on GPU without Pumas in this example [8]. Different from the C/C++ language, it is “_global_void”, it is proper belongs to the CUDA language; it means the function should only run on GPU. GPU parallel means it will process all of the pixel point at the same time. Each of the processing is independent, we can see the image as a lot of small image, and each image is one pixel point. Then we can merge the point together as the image after the image processing [9].

5 The Peroration

Vehicle recognition occupies an important position in the intelligent transportation system in the future; the method based on image processing will be a hot research. In order to make the image processing more quickly and efficient, it is necessary to make the algorithm parallel. Whatever using GPU parallel computing ability, it can improve the speed of image processing, to meeting the high efficiency of vehicle recognition efficiency.

Acknowledgment This work was financially supported by National Science and Technology Major Project (No. 2013ZX010033002-003); the national five year science and technology support program funded projects (No. 2012BAH07B01).

References

1. Ying-ying C (2013) The research on vehicle type recognition in intelligent transportation system. University of Electronic Science and Technology of China, Chengdu (in Chinese)
2. Zhi-pan W (2013) The research of vehicle recognition based on BP neural network. Modern Comput (Prof) 04:38–41 (in Chinese)
3. Xia-feng C (2011) GPU-based parallel computing and research in pattern recognition. Comput Digital Eng 08:118–122 + 142 (in Chinese)

4. Yan Z (2007) The method research and completion of identifying the types of the Military Vehicle. DUT (in Chinese)
5. Jolly MPD, Lakshmanan S, Jain AK (1996) Vehicle segmentation and classification using deformable templates. *IEEE Trans Pattern Anal Intell* 18(3):293–308
6. Yau HC, Manry MT (1991) Shape recognition with nearest neighbors isomorphic network. In: *Proceedings of the 1st IEEE-SP workshop on neural networks for signal processing*, vol 9. Princeton, New Jersey, pp 246–255
7. Foley JD, van Dam A, Feiner SK et al (2002) *Computer graphics: principles and practice* second edition in C. China Machine Press and McGraw Hill Education, pp 116–122
8. Cook S (2012) *CUDA programming: a developer's guide to parallel computing with GPUs*. Newness
9. Goldman R (2003) Deriving linear transformation in three dimensions. *IEEE Comput Graph Appl* 23(3):66–71

Pseudo Nearest Centroid Neighbor Classification

Hongxing Ma, Xili Wang and Jianping Gou

Abstract In this paper, we propose a new reliable classification approach, called the pseudo nearest centroid neighbor rule, which is based on the pseudo nearest neighbor rule (PNN) and nearest centroid neighborhood (NCN). In the proposed PNCN, the nearest centroid neighbors rather than nearest neighbors per class are first searched by means of NCN. Then, we calculate k categorical local mean vectors corresponding to k nearest centroid neighbors, and assign the weight to each local mean vector. Using the weighted k local mean vectors for each class, PNCN designs the corresponding pseudo nearest centroid neighbor and decides the class label of the query pattern according to the closest pseudo nearest centroid neighbor among all classes. The classification performance of the proposed PNCN is evaluated on real data sets in terms of the classification accuracy. The experimental results demonstrate the effectiveness of PNCN over the competing methods in many practical classification problems.

Keywords K-nearest neighbor rule · Nearest centroid neighborhood · Local mean vector · Pattern classification

H. Ma (✉) · X. Wang
College of Computer Science, Shaanxi Normal University, Xi'an,
Shaanxi 710062, People's Republic of China

H. Ma
College of Electrical and Information Engineering,
Beifang University of Nationalities, Yinchuan, Ningxia 750021,
People's Republic of China

J. Gou
School of Computer Science and Telecommunication Engineering,
JiangSu University, ZhenJiang, JiangSu 212013, People's Republic of China

1 Introduction

In pattern recognition, k -nearest neighbor rule (KNN) is one of the most widely used nonparametric approaches, and is also deemed to be one of the top 10 algorithms in data mining [1], due to its simplicity and effectiveness for classification. It has been theoretically proven that the KNN classifier has asymptotically optimal performance in the Bayes sense [2, 3]. Moreover, the appeal of KNN is that only a single integer parameter k is required to adjust and any particular statistical distribution of the training data should not be considered [4]. However, the classification performance of KNN and its variants are still degraded by three main issues: the sparse problem, the imbalance problem and the noise problem [5]. The sparse problem is that there are a small number of training samples in many practical classification tasks. In the small training sample size cases, the nonparametric KNN-based classifiers usually suffer from the existing outliers [6]. The imbalance problem is produced when the data in one class heavily outnumbers the data in another class. In this case, the class boundary can be skewed to the class with few samples. The noise problem is the sensitivity to the outliers or noises that exists in KNN and its variants, as they treat both noisy and normal points equally.

In fact, the choice of the neighborhood size k can aggravate the negative influence of the three problems aforementioned on the classification performance of KNN-based methods to some degree [1, 7]. If k is too small, the results can be sensitive to the data sparseness and the noisy points. On the other hand, if k is too large, then the results can be degraded by the introduction of many outliers from other classes in the neighborhood. Thus, the performance of nonparametric KNN-based classifiers can be severely affected by the existing outliers, particularly in the cases of the sparse, imbalance and noise problems.

To overcome the existing outliers, a reliable KNN-based approaches, called the local mean-based KNN rule (LMKNN), is well designed in [8]. It uses the local mean vector of the categorical k nearest neighbors to determine the classes of query patterns. Subsequently, the basic idea of LMKNN has successfully been applied to some approaches, such as the pseudo nearest neighbor rule (PNN) [9], the local mean-based k -nearest centroid neighbor rule (LMKNCN) [10] and other methods [11–13]. As an extension of LMKNN, PNN is also robust to the outliers. It utilizes the distance weighted local learning in each class to design the pseudo nearest neighbor of the query pattern, based on the distance weighted k -nearest neighbor rule [14] and LMKNN. Then, the query pattern is allocated into the class, which the closest pseudo nearest neighbor belongs to. As we know, k -nearest centroid neighbor rule (KNCN), based on NCN [15], is very effective classifier, especially in the small sample size situations [16, 17]. Combined the robustness of LMKNN and effectiveness of KNCN, LMKNCN is introduced in [10]. It employs the local mean vector of k nearest centroid neighbor from each class to decide the class of the query pattern.

To further perform classification on the existing outliers well, especially in the sparse, imbalance and noise situations, we propose the pseudo nearest centroid neighbor rule (PNCN), motivated by PNN and NCN. In this method, we design the pseudo nearest centroid neighbor rather than the pseudo nearest neighbor in each class to classify the query pattern. First, k nearest centroid neighbors of one query pattern are found from each class, and then k local mean vectors corresponding to k neighbors are calculated. Second, the weights for the local mean vectors instead of the neighbors are assigned. Third, the pseudo nearest centroid neighbor in each class are decided by using the weighted sum of distances of k local mean vectors. Finally, the query pattern is classified into the class, which the closest pseudo nearest centroid neighbor belongs to. The classification performance of the proposed PNCN is investigated on real data sets, compared to KNN, LMKNN, KNCN, LMKNCN and PNN. Experimental results suggest that PNCN is effective and robust in such practical situations.

2 Pseudo Nearest Centroid Neighbor Classification

In pattern classification, the recognition rates of the KNN-based nonparametric classifiers are easily affected by the outliers in the issues above. Considering the superiorities of both PNN and NCN, we give a new scheme of designing pseudo nearest neighbor and accordingly propose the pseudo nearest centroid neighbor rule (PNCN), in order to improve the classification accuracy rate. In what follows, for the ease of presentation in the general recognition problem, we first suppose that there is a training set $T = \{x_i \in \mathbb{R}^d\}_{i=1}^N$ with M classes in d -dimensional feature space, and the corresponding class labels are $\{y_1, y_2, \dots, y_N\}$, where $y_i \in \{c_1, c_2, \dots, c_M\}$, a class subset of T from the class c_l is $T_l = \{x_{ij} \in \mathbb{R}^d\}_{j=1}^{N_l}$ with the number of the training samples N_l .

2.1 Nearest Centroid Neighborhood

As we know, the choice of neighborhood plays a critical role in the KNN-based classification [1]. It has been found that Nearest centroid neighborhood (NCN) is a very good alternative to nearest neighborhood [15]. The concept of NCN focuses on the idea that the neighborhood of a query pattern is simultaneously subject to the distance criterion and the symmetry criterion. On the one hand, the neighbors of a query are as close to it as possible by the distance criterion. On the other hand, the neighbors of a query are placed as homogeneously around it as possible with the

symmetry criterion. To seek the neighbors according to NCN, the centroid of a set $Z = \{z_1, z_2, \dots, z_n\}$ should be first defined as

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n z_i. \quad (1)$$

Then, the nearest centroid neighbors of a given query pattern x with both criterions above, are searched through an iterative procedure [15] as follows:

1. Find the first nearest centroid neighbor x_1^{NCN} of x that corresponds to its nearest neighbor.
2. Find the i -th nearest centroid neighbor x_i^{NCN} ($i \geq 2$), which is imposed by the constraint that the centroid of the query x and all previous centroid neighbors, i.e., $x_1^{NCN}, \dots, x_{i-1}^{NCN}$, is the closest to x .

Based on the NCN, KNCN and LMKNCN are introduced in the field of pattern classification [10, 16].

2.2 The Proposed PNCN Classifier

Based on NCN and PNN, we introduce the pseudo nearest centroid neighbor rule (PNCN). It first finds the k nearest centroid neighbors per class in terms of the NCN, and then computes each local mean vector of first j categorical neighbors and allocates the weight for each local mean vector. Finally, the pseudo nearest centroid neighbor per class is designed by using the weighted k local mean vectors corresponding to k nearest centroid neighbors. In the process of making classification decision, PNCN assigns the class label, which the closest pseudo nearest centroid neighbor belongs to among all classes, into the unseen pattern.

Given a query pattern x in the pattern classification problem, the PNCN decides the class label of x as follows:

1. Search k nearest centroid neighbors from T_l of each class c_l for the query pattern x in the training set T , say $T_{lk}^{NCN}(x) = \{x_{lj}^{NCN} \in \mathbb{R}^d\}_{j=1}^k$.
2. Compute the local mean vector $\bar{u}_{lj}^{NCN}(x)$ of the first j nearest centroid neighbors of a query x from class c_l . Let $\bar{U}_{lk}^{NCN}(x) = \{\bar{u}_{lj}^{NCN}(x) \in \mathbb{R}^d\}_{j=1}^k$ denote the set of the k local mean vectors corresponding to k nearest centroid neighbors in the class c_l , and $d(x, \bar{u}_{l1}^{NCN}(x)), d(x, \bar{u}_{l2}^{NCN}(x)), \dots, d(x, \bar{u}_{lk}^{NCN}(x))$ are their corresponding Euclidean distances to x .

$$\bar{u}_{lj}^{NCN}(x) = \frac{1}{j} \sum_{m=1}^j x_{lm}^{NCN}. \quad (2)$$

It should be noted that the local mean vector $\bar{u}_{l_1}^{NCN}(x)$ of the first nearest centroid neighbor $x_{l_1}^{NCN}$ is the same as the first nearest neighbor.

- Assign different weights to k categorical local mean vectors in the same way as the PNN, and the weight \bar{W}_{lj}^{NCN} of the j -th local mean vector $\bar{u}_{lj}^{NCN}(x)$ for the class c_l is determined as:

$$\bar{W}_{lj}^{NCN} = \frac{1}{j} \quad j = 1, \dots, k. \quad (3)$$

- Design the pseudo nearest centroid neighbor $\bar{x}_l^{PNCN}(x)$ of the query point x from class c_l , and c_l can be viewed as the class label of $\bar{x}_l^{PNCN}(x)$. The distance $\bar{d}(x, \bar{x}_l^{PNCN}(x))$ between x and $\bar{x}_l^{PNCN}(x)$ can be defined by the weighted sum of distances of k categorical local mean vectors to x as follows:

$$\begin{aligned} \bar{d}(x, \bar{x}_l^{PNCN}(x)) = & \left(\bar{W}_{l1}^{NCN} \times d(x, \bar{u}_{l1}^{NCN}(x)) + \bar{W}_{l2}^{NCN} \times d(x, \bar{u}_{l2}^{NCN}(x)) \right. \\ & \left. + \dots + \bar{W}_{lk}^{NCN} \times d(x, \bar{u}_{lk}^{NCN}(x)) \right). \end{aligned} \quad (4)$$

- Classify the query point x into the class c , which the closest pseudo nearest centroid neighbor belongs to in the light of Eq. (4) among all classes.

$$c = \arg \min_{c_l} \bar{d}(x, \bar{x}_l^{PNCN}(x)). \quad (5)$$

Note that the proposed PNCN is equivalent to the 1NN, LMKNN, PNN, KNCN and LMKNCN rules only when $k = 1$, and the value of k is no more than N_l .

2.3 The PNCN Algorithm

According to the procedure of the PNCN above, we summarize it in Algorithm 1 by means of the pseudo codes.

Algorithm 1: The pseudo nearest centroid neighbor algorithm

Require:

x : a query pattern, k : number of nearest neighbors, $T = \{x_i \in \mathbb{R}^d\}_{i=1}^N$: a training set.

$T_l = \{x_{lj} \in \mathbb{R}^d\}_{j=1}^{N_l}$: a training subset from class c_l , c_1, \dots, c_M : M class labels.

M : the number of classes in T , N_1, \dots, N_M : number of training samples for M classes.

Ensure:

Predict the class label of the query pattern x by the closet pseudo nearest centroid neighbor among all classes.

Step 1: Calculate the Euclidean distances of training samples in each class c_l to x .

for $j = 1$ to N_l **do**

$$d(x, x_{lj}) = \sqrt{(x - x_{lj})^T (x - x_{lj})}$$

end for

Step 2: Search the k nearest centroid neighbors of x in each class c_l , say $T_{lk}^{NCN}(x) = \{x_{lj}^{NCN} \in \mathbb{R}^d\}_{j=1}^k$.

(i) Find the first nearest centroid neighbor of x in each class c_l , say x_{l1}^{NCN} .

$$[\min_index, \min_dist] = \min(d(x, x_{lj}))$$

$$\text{set } x_{l1}^{NCN} = x_{\min_index}, R_l^{NCN}(x) = \{x_{l1}^{NCN} \in \mathbb{R}^d\}$$

(ii) Find k nearest centroid neighbors of x except x_{l1}^{NCN} in each class c_l .

for $j = 2$ to k **do**

Set $S_l(x) = T_l - R_l^{NCN}(x) = \{x_{ln} \in \mathbb{R}^d\}_{n=1}^{L_l(x)}$, $L_l(x) = \text{length}(S_l(x))$

Calculate the sum of the previous $j - 1$ nearest centroid neighbors.

$$\text{sum}_l^{NCN}(x) = \sum_{r=1}^{j-1} x_{lr}^{NCN}$$

Compute the centroids in the set S_l for x .

for $n = 1$ to $L_l(x)$ **do**

$$\bar{x}_{ln} = \frac{1}{j} (x_{ln} + \text{sum}_l^{NCN}(x)), \bar{d}_{ln}(x, \bar{x}_{ln}) = \sqrt{(x - \bar{x}_{ln})^T (x - \bar{x}_{ln})}$$

end for

Find the j -th nearest centroid neighbor.

$$[\min_index^{NCN}, \min_dist^{NCN}] = \min(\bar{d}_{ln}(x, \bar{x}_{ln}))$$

Set $x_{lj}^{NCN} = x_{\min_index^{NCN}}$, and add x_{lj}^{NCN} to the set $R_l^{NCN}(x)$.

end for

Set $T_{lk}^{NCN}(x) = R_l^{NCN}(x)$.

Step 3: Compute the local mean vector $\bar{u}_{ij}^{NCN}(x)$ of the first j nearest neighbors of x using $T_{lk}^{NCN}(x)$ and the corresponding distance $d(x, \bar{u}_{ij}^{NCN}(x))$ between $\bar{u}_{ij}^{NCN}(x)$ and x .

for $j = 1$ to k **do**

$$\bar{u}_{ij}^{NCN}(x) = \frac{1}{j} \sum_{m=1}^j x_{lm}^{NCN}, d(x, \bar{u}_{ij}^{NCN}(x)) = \sqrt{(x - \bar{u}_{ij}^{NCN}(x))^T (x - \bar{u}_{ij}^{NCN}(x))}$$

end for

Set $\bar{U}_{lk}^{NCN}(x) = \{\bar{u}_{ij}^{NCN}(x) \in \mathbb{R}^d\}_{j=1}^k$, $\bar{D}_{lk}^{NCN}(x) = \{d(x, \bar{u}_{i1}^{NCN}(x)), \dots, d(x, \bar{u}_{ik}^{NCN}(x))\}$.

Step 4: Allocate the weights \bar{W}_{lj}^{NCN} to the j -th the local mean vector $\bar{u}_{ij}^{NCN}(x)$ in the set $\bar{U}_{lk}^{NCN}(x)$.

for $j = 1$ to k **do**

$$\bar{W}_{lj}^{NCN} = \frac{1}{j} \quad j = 1, \dots, k.$$

end for

Set $\bar{W}_{lk} = \{\bar{W}_{l1}^{NCN}, \dots, \bar{W}_{lk}^{NCN}\}$.

Step 5: Design pseudo nearest centroid neighbor $\bar{x}_l^{PNCN}(x)$ using \bar{W}_{lk} and $\bar{D}_{lk}^{NCN}(x)$.

$$\begin{aligned} \bar{d}(x, \bar{x}_l^{PNCN}(x)) &= \bar{W}_{l1}^{NCN} \times d(x, \bar{u}_{l1}^{NCN}(x)) + \bar{W}_{l2}^{NCN} \times d(x, \bar{u}_{l2}^{NCN}(x)) \\ &\quad + \dots + \bar{W}_{lk}^{NCN} \times d(x, \bar{u}_{lk}^{NCN}(x)) \end{aligned}$$

Step 6: Assign the class c of the closest pseudo nearest centroid neighbor to x .

$$c = \arg \min_{c_l} \bar{d}(x, \bar{x}_l^{PNCN}(x))$$

3 Experiments

In this section, we conduct the experiments to validate the classification performance of the proposed PNCN on the benchmark real data sets. The PNCN is compared with KNN, LMKNN, PNN, KNCN and LMKNCN in terms of the

classification accuracy rate, which is taken as one of the effective measures in pattern recognition [8, 10]. In what follows, we should note that the neighborhood size k is for all training samples of a query pattern in KNN and KNCN, while is for training samples of each class in LMKNN, LMKNCN, PNN and PNCN.

3.1 Data Sets

In the experiments, twelve real data sets taken from the UCI Repository [18] are employed. The information of these UCI data sets including the numbers of samples, attributes, classes, training and testing samples is displayed in Table 1. For short, among these data sets, the abbreviated names for ‘Parkinsons’, ‘Transfusion’, ‘Libras Movement’, ‘Cardiotocography’, ‘LandsatSatellite’, ‘Page-blocks’, ‘Image Segmentation’ and ‘Robot Navigation’ are ‘Park’, ‘Trans’, ‘Libras’, ‘Cardio’, ‘Landsat’, ‘Page’, ‘Image’ and ‘Robot’, respectively. Note that the Glass data set originally have seven classes, but in our experiments the five classes with very few samples are deleted.

3.2 Experimental Results

In this subsection, to well demonstrate the classification performance of the proposed PNCN, we do the experiments on the real UCI data sets. One of the advantages of using the real data sets is that they are generated without any knowledge of the classification procedures that it will be used to test. The second

Table 1 The real UCI data sets used in the experiments

Data	Size	Attributes	Classes	Testing samples	Training samples
Sonar	208	60	2	132	76
Park	195	22	2	65	130
Seed	210	7	3	105	105
Wine	178	13	3	59	119
Glass	146	9	2	53	93
Trans	748	4	2	248	500
Libras	360	90	15	90	270
Landsat	6435	36	6	2146	4289
Cardio	2126	21	10	710	1416
Image	2310	19	7	1078	1232
Robot	5456	4	4	1818	3638
Page	5473	10	5	1830	3643

advantage is that the sparse, imbalance, noise problems are usually produced in practical classification, and the outliers for one test sample in these selected real data sets always exist. Since the classification performance of each method is verified by using the validation test, each whole data set is randomly split into a training set and test set. To assess the quality of each method, we perform 10 times on each data set, and the average classification accuracy with 95 % confidence over test sets is viewed as the final performance of each method. For each whole data set, the training and test samples are randomly generated, shown in the Table 1. In the experiments, the parameter of the neighborhood size k takes the value from 1 to 15 with step 1.

To validate the proposed PNCN method on the performance, we first explore the classification accuracy rates of the competing classifiers with varying the neighborhood size k on each real data set. As there is no general way to determine the optimal parameter k in KNN-based methods, it is expected that our PNCN can be more robust to the change of k when the parameter is common for all compared methods. The experimental comparisons of all the classifiers in terms of the classification accuracy is illustrated in Figs. 1 and 2. We can obviously observe that the proposed PNCN almost surpasses the other methods among the preset range of the neighborhood size k on each data set. Compared to KNN, LMKNN, PNN, KNCN and LMKNCN, the classification performance of PNCN at first increases when the values of k is small, and then grows slowly or keeps almost stable as k increases on all the data sets except the Seed data set. Moreover, the best performance of the PNCN is usually yielded at the larger value of the neighborhood size, this fact implies that it can use more nearest neighbors to improve the classification. However, KNN, LMKNN, PNN, KNCN and LMKNCN vary drastically

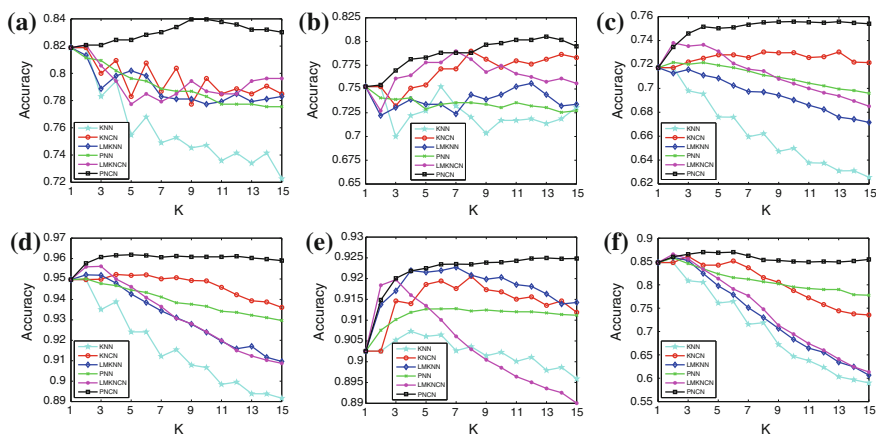


Fig. 1 The accuracy rates of each method via k on each real data set

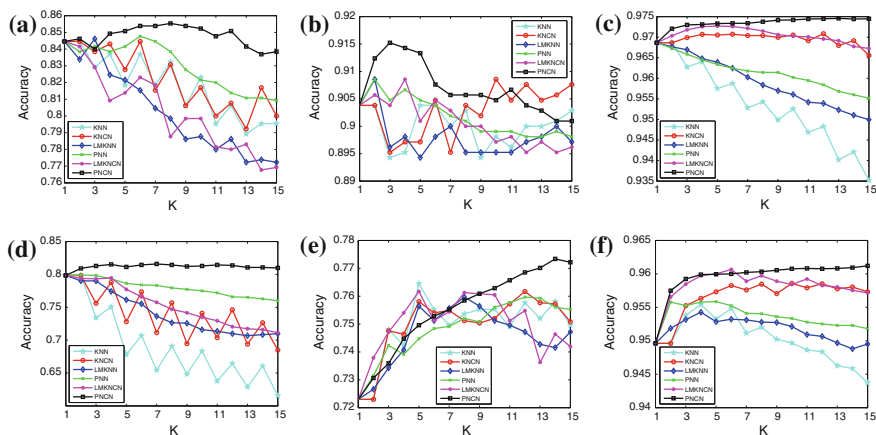


Fig. 2 The accuracy rates of each method via k on each real data set

against the parameter k . It can also be seen that the differences of the classification accuracy rates between PNCN and the other methods are very significant at a larger k . Consequently, we can draw a conclusion that the proposed PNCN is robust to the choice of k with satisfactory performance. This means that the selection of k for PNCN is easier than that for the other KNN-based classifier.

The empirical comparisons of all the competing classifiers are investigated by the maximal accuracy rates (%) of each method with the corresponding standard deviations (stds) and values of parameter k in the parentheses on each real data sets. The classification results of each method on all data sets are given in Table 2. It should be noted that the best classification performance among these methods are indicated in bold-face on each data set. When we look at the best cases in Table 2, the proposed PNCN is found to be very superior to the other methods in most cases. More interestingly, it can be observed that the best classification results of PNCN are nearly yielded at larger values of k on all data sets, compared to the other five methods, shown in Table 2, Figs. 1 and 2. In our experiments, there are a finite number of training samples and the training samples are randomly chosen from each whole real data set, so the value of k can easily affect the classification performance. Nevertheless, the experiments show that the proposed PNCN can use more nearest neighbors to capture enough information, so as to improve the classification performance.

Table 2 The maximal accuracy rates (%) of each method with the corresponding standard deviations (stds) and values of k in the parentheses on each UCI data sets

Data	KNN	KNCN	LMKNN	LMKNCN	PNN	PNCN
Sonar	79.85 ± 3.30	79.85 ± 3.30	79.85 ± 3.30	79.85 ± 3.30	79.92 ± 3.63	81.59 ± 3.69
	(1)	(1)	(1)	(1)	(2)	(7)
Park	84.46 ± 3.28	84.46 ± 3.28	84.62 ± 4.23	84.46 ± 3.28	84.77 ± 4.00	85.54 ± 3.92
	(1)	(1)	(3)	(1)	(6)	(8)
Seed	90.38 ± 2.03	90.86 ± 1.81	90.86 ± 2.92	90.86 ± 2.02	90.86 ± 2.30	91.52 ± 2.44
	(1)	(10)	(2)	(4)	(2)	(3)
Wine	75.25 ± 4.67	78.98 ± 3.50	75.59 ± 5.19	78.98 ± 4.74	75.25 ± 4.67	80.51 ± 3.12
	(1)	(8)	(12)	(7)	(1)	(13)
Glass	81.89 ± 5.85	81.89 ± 5.85	81.89 ± 5.85	82.08 ± 3.90	81.89 ± 5.85	83.96 ± 3.59
	(1)	(1)	(1)	(2)	(1)	(9)
Trans	76.45 ± 1.49	76.17 ± 1.48	75.97 ± 1.71	76.17 ± 1.89	75.97 ± 1.28	77.34 ± 1.33
	(5)	(12)	(8)	(5)	(12)	(14)
Libras	84.78 ± 3.78	86.11 ± 3.02	85.89 ± 3.78	86.56 ± 3.83	85.78 ± 4.12	87.00 ± 3.27
	(1)	(3)	(2)	(2)	(2)	(4)
Landsat	90.73 ± 0.39	92.05 ± 0.59	92.27 ± 0.71	91.98 ± 0.41	91.28 ± 0.41	92.50 ± 0.46
	(4)	(8)	(7)	(3)	(7)	(13)
Cardio	71.73 ± 1.01	73.03 ± 1.15	71.73 ± 1.01	73.80 ± 1.36	72.17 ± 1.38	75.58 ± 1.36
	(1)	(8)	(1)	(2)	(2)	(13)
Image	94.97 ± 0.55	95.22 ± 0.71	95.20 ± 0.54	95.62 ± 0.68	95.01 ± 0.56	96.19 ± 0.57
	(1)	(4)	(2)	(3)	(2)	(5)
Sensor	96.86 ± 0.38	97.08 ± 0.42	96.86 ± 0.38	97.27 ± 0.33	96.86 ± 0.38	97.46 ± 0.38
	(1)	(12)	(1)	(5)	(1)	(13)
Page	95.54 ± 0.50	95.86 ± 0.34	95.43 ± 0.44	96.07 ± 0.52	95.58 ± 0.45	96.12 ± 0.32
	(4)	(10)	(4)	(6)	(5)	(15)

4 Conclusions

In this paper, we propose a new classifier, called the pseudo nearest centroid neighbor rule, with aim of further improving the classification performance. It is motivated by the PNN and NCN. In the new method, we find k nearest centroid neighbors based on the NCN and calculate the k categorical local mean vectors corresponding to the k nearest centroid neighbors. The proposed PNCN designs the pseudo nearest centroid neighbor for each class by using the weighted k local mean vectors, and assigns the class of the closest pseudo nearest centroid neighbor to the query pattern. To investigate the performance of PNCN, we conduct the experiments on real data sets. The experimental results suggest that the proposed PNCN method are promising classifier.

Acknowledgment This work was supported by National Science Foundation of China (Grant Nos. 61162005, 41171338 and 61163002), the Beifang Ethnic University school project (Grant No. 2010Y030), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 14KJB520007), China Postdoctoral Science Foundation (Grant No. 2015M570411) and Research Foundation for Talented Scholars of JiangSu University (Grant No. 14JDG037).

References

1. Wu X, Kumar V, Quinlan JR, Ghosh J (2008) Top 10 algorithms in data mining. *Knowl Inf Syst* 14(1):1–37
2. Cover TM, Hart PE (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
3. Wagner T (1971) Convergence of the nearest neighbor rule. *IEEE Trans Inf Theory* 17:566–571
4. Blanzieri E, Melgan F (2008) Nearest neighbor classification of remote sensing images with the maximal margin principle. *IEEE Trans Geosci Remote Sens* 46(6):1804–1811
5. Guo G, Dyer CR (2005) Learning from examples in the small sample case: face expression recognition. *IEEE Trans Syst Man Cybern* 35:477–488
6. Fukunaga K (1990) Introduction to statistical pattern recognition. Academic Press, San Diego, CA, USA, pp 219–238
7. Gou J, Du L, Zhang Y, Xiong T (2012) A new distance-weighted k-nearest neighbor classifier. *J Inf Comput Sci* 9(6):1429–1436
8. Mitani Y, Hamamoto Y (2006) A local mean-based nonparametric Classifier. *Pattern Recognition Lett* 27(10,15):1151–1159
9. Zeng Y, Yang Y, Zhao L (2009) Pseudo nearest neighbor rule for pattern classification. *Expert Syst Appl* 36(2):3587–3595
10. Gou J, Du Zhang Yi L, Xiong T (2012) A local mean-based K-nearest centroid neighbor classifier. *Comput J* 55(9):1058–1071
11. Yang J, Zhang L, Yang JY, David Zhang (2011) From classifiers to discriminators: a nearest neighbor rule induced discriminant analysis. *Pattern Recogn* 44(7):1387–1402
12. Gou J, Zhan Y, Rao Y, Shen X, Wang X, He W (2014) Improved pseudo nearest neighbor classification. *Knowl-Based Syst* 70:361–375
13. Yang T, Kecman V (2008) Adaptive local hyperplane classification. *Neurocomputing* 71(13–15):3001–3004

14. Dudani SA (1976) The distance-weighted k-nearest neighbor Rule. *IEEE Trans Syst Man Cybern* 6(4):325–327
15. Chaudhuri BB (1996) A new definition of neighbourhood of a point in multi-dimensional space. *Pattern Recogn Lett* 17(1):11–17
16. Sánchez JS, Pla F, Ferri FJ (1997) On the use of neighbourhoodbased non-parametric classifiers. *Pattern Recogn Lett* 18(11–13):1179–1186
17. Sánchez JS, Marqués AI (2006) An LVQ-based adaptive algorithm for learning from very small codebooks. *Neurocomputing* 69(7–9):922–927
18. Bache K, Lichman M (2015) UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml>. University of California, School of Information and Computer Science, Irvine, CA

Recommended System for Cognitive Assessment Evaluation Based on Two-Phase Blue-Red Tree of Rule-Space Model: A Case Study of MTA Course

Yung-Hui Chen, Chun-Hsiung Tseng, Ching-Lien Huang,
Lawrence Y. Deng and Wei-Chun Lee

Abstract Having more than one professional certification is one of the various indicators for the Ministry of Education to promote and evaluate their competencies that the vocational education system students in Taiwan. This is also one way to find the ideal jobs that enhance their competitiveness for vocational students on-the-job. As a result, it is very important for students in vocational education systems to obtain professional certifications. In particular, the more professional licenses they have, the more job opportunities for them. Therefore, we propose a RS (Recommended System) that combines two-phase Blue-Red trees of Rule-Space Model and the best learning path, and it is used to remedy and analyze the learning situation of MTA courses and enhance the pass rate of MTA licenses for students. We classify three SGs (Skill Groups) from the Certiport of Microsoft certification

Y.-H. Chen

Department of Computer Information and Network Engineering,
Lunghwa University of Science and Technology, Taoyuan City, Taiwan
e-mail: cyh@mail.lhu.edu.tw

C.-H. Tseng

Department of Information Management, Nanhua University, Chiayi, Taiwan
e-mail: lendle_tseng@seed.net.tw

C.-L. Huang (✉)

Department of Industrial Management, Lunghwa University of Science and Technology,
Taoyuan City, Taiwan
e-mail: lynne@mail.lhu.edu.tw

L.Y. Deng

Department of Computer Science and Information Engineering, St. John's University,
New Taipei City, Taiwan
e-mail: lawrence@mail.sju.edu.tw

W.-C. Lee

Department of Business Administration, Lunghwa University of Science and Technology,
Taoyuan City, Taiwan
e-mail: liwj@mail.lhu.edu.tw

center in the first phase, and the three SGs (Skill Groups) can be produced as a concept map and Blue-Red trees. In the second phase, The ten chapters of MTA course are classified within the three SGs (Skill Groups) of phase one according to the most similarity in contents between ten chapters and three SGs (Skill Groups). That is, three groups will be created in a MTA course from previous ten chapters. The three groups result in three concept maps and three groups of Blue-Red trees. After that, it is based on the analysis of Rule-Space Mode for all learning objects in each skill group of phase two. We can define the RW (Relation Weight) of every learning object associated with the other learning objects, and separately calculate the Confidence Level values of between two adjacent learning objects from all learning paths. Finally, the optimal learning path can be obtained by the inferred optimal learning path algorithm from the combination of RW (Relation Weight) and CL (Confidence Level). The proposed method can be used to OCLS (Online Course Learning System) that recommended the best learning path of learning objects for learners to online self-learning, or to RS (Recommended System) that provides the basis of self-learning remedies for RFRC (Recommended Form of Remedial Course).

Keywords Rule-Space model · Blue-Red tree · Recommendation system · Relation weight · Confidence level · Learning path

1 Introduction

Today, to provide a remedial teaching method that is intelligent and capable of automatically assess the learning status of learning in distant learning systems is a common trend.

1.1 Knowledge Space Model

Designing a learning path recommendation system that is intelligent and being capable of detecting the learning status of learners for efficient compensation has been a common issue. Especially, developing adaptive learning path for learners with different capabilities and backgrounds has been a trend. Therefore, to increase learners' speed for learning or diversities in learning object selection, we propose the learning concept map, reasonable learning path, and learning path optimization algorithm of adaptive recommendation systems to help learners' learning efficiently and effectively.

We adopt knowledge space cognitive assessment method [1] to infer the learning concept map for learning objects in each learning group. The learning concept map is then used as the basis for deriving the learning tree from all reasonable learning orders. As shown in Fig. 1, assume there are four learning objects in the learning

Fig. 1 Concept map of learning objects

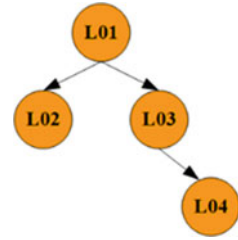
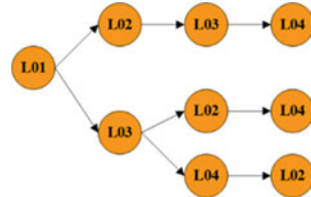


Fig. 2 Learning paths of learning object



tree, parent nodes must be learned before learning child nodes. For example, to learn object L04, one has to learn objects L03 and L01 in advance. Learn object L01 must be learned before learning Learn object L03, for the same reason.

We can gradually deduce the set of all learning objects architecture from an empty with no learning object, and link to a number of different learning path. Therefore, from the above example, it can be inferred many learning paths and associations, as shown in Fig. 2.

1.2 Recommendation System

In recent years, along with the rapid improvements of information industries, the constructions and applications of recommendation systems are wider and wider no matter commercially or academically with amazing achievements. From the point of view of consumers, there are four types of recommendation systems: Personalized recommendation, Social recommendation, Item recommendation and hybrid recommendation (A combination of the three approaches above) [2]. The personalized recommendation recommends things based on the individual’s past behavior, the social recommendation recommends things based on the past behavior of similar users, the Item recommendation then recommends things based on the thing itself.

The major algorithms adopted by recommendation systems include the technologies of data mining, information filtering and retrieval, and collaborative filtering, etc. [3]. Content-based filtering refers to the method of collecting a huge amount of keywords to allow figuring out the relationships among keyword with data mining. It analyzes the types of content, compares the content with user’s

behavior, and records historical content to find out the item the user may need in the future time [4].

Collaborative filtering (CF) refers to the method of summarizing the consuming habits of the crowd to obtain the mathematic model of all members and products as the basis of the recommendation system. There are two different strategies from the point of view of members and products. The first one, the so called user based CF, is to analyze the data of all members to find out similar consumers to predict their tastes against different products. This is also referred to as social recommendation. The other strategy, the so called item-based CF, is to start with the similarity among products. Then analyzing recent best-sellers and giving discounts to needed consumers. This is referred to as item recommendation. A famous example is amazon.com [5].

The rapid improvements of recent network technologies results in more wider and richer adoption of recommendation systems [6, 7]. Park et al. [8] proposed a personalized channel recommendation system including four modules of manager agent, monitoring agent, data preparation agent and recommendation agent, and recommendation algorithm on the wireless Internet platform for getting better performance. Wang et al. [9] then designed and implemented a semantic-based friend recommendation system of Friendbook in social network environment. The Friendbook can recommend potential friends to users by sharing similar life styles from user-centric data collected through sensors on the Android-based smartphone. The recommended results showed the correct predilections in choosing friends from similar life styles.

Therefore, the session 2 introduces the analysis steps of Rule-Space Model. A MTA course schedule based on two-phase blue-read trees is introduced in session 3. The session 4 then infers the best learning path based on Rule-Space Model. A Recommendation System (RS) with learning performance analysis from online learning and testing system is constructed in session 5, and the Sect. 6 then lists some preliminary experiment results. Finally, we will have a brief conclusion and future development in Sect. 7.

2 Rule-Space Model Analysis

To assess the test results of target learners, “Rule-Space Model” is adopted in this research. The method has already been utilized in a huge amount of researches for analyzing the knowledge structure of learners and has gained many positive feedbacks. But how will it performed in the analysis of certification courses is still unknown. The model was proposed by Tatsuoka [10] and was a cognitive diagnosis model developed from psychology measurement theorems and was usually utilized to investigate the relationships between the knowledge structure and problem solving for learners and was adopted in various collaborative learning researches [11, 12]. Here, we utilized Kikumi K. Tatsuoka’s Rule-Space Model to infer the

hierarchical relationship among learning objects and topics and used the results as a basis of learning assessment, analysis, and result improvement.

The analysis steps and flows for Rule-Space Model is listed below:

1. Establish a complete collection of knowledge concepts of a course, and use the concept matrices to illustrate the relationships between knowledge and concepts to obtain the learning concept diagram.
2. Derive the reachability matrix based on the concept matrices. The reachability matrix represents pre-conditions, post-conditions, and the learning sequence of preliminary knowledge concepts.
3. Infer the incidence matrix of all learning paths in the tree structure. This in turn can be used to derive the reduced incidence matrix of all reasonable learning paths, which can be used for Blue-Red tree conversion.
4. Convert the reduced incidence matrix to an ideal attribute matrix and utilized the matrix to obtain the four most important factors:
 1. Examinee: the number of different learning results in the graph.
 2. The ideal item response vectors that can be used to calculate the learning performance of each examinee.
 3. Total scores: calculate the score of learning performance of each examinee from the ideal item response vectors.
 4. Examinee attributes: each examinee represents a kind of learning performance, and an examinee attribute represents a corresponding blue-read tree.

3 MTA Course Planning of Two Phases Blue-Red Trees

The finer the kinds of knowledge are categorized, the more attribute nodes will be generated. The ideal item matrix obtained from Rule-Space Model will also be larger, especially when we have more than 8 attribute nodes, the resulting learning path graph will be too complex for the analysis of learning performance. As a result, two-phase Rule-Space Model analysis is adopted in this research. Coarse-scoped categories are used in the first phase. In the second phase, fine-scoped categories will be utilized. Learners should learn knowledge from fine-scoped categories and assemble coarse-scoped knowledge structures from fine-scoped knowledge. In this research, we categorize the chapters of MTA-Networking-Fundamentals into three skill groups:

- S1: Understanding Network Infrastructures: including Ch02 (Introduction to Computer Networks), Ch06 (Local Area Networking), Ch07 (Wide Area Networks), and Ch08 (Wireless Networks).
- S2: Understanding Network Hardware: including Ch03 (Network Hardware).

- S3: Understanding Protocols and Services: including Ch04 (Protocol), Ch05 (OSI Reference Mode), Ch09 (Networking Services), Ch10 (TCP/IP in the Command-Line), and Ch11 (Network Security).

Each chapter of the MTA course was put into its nearest skill group. Group S1 contains four chapters: Ch02, Ch06, Ch07, and Ch08. Group S2 contains only chapter Ch03 and Group S3 contains Ch04, Ch05, Ch09, Ch10, and Ch11. Then, the three large groups were analyzed according to the order of their skill knowledge and obtained the skill group concept graph as shown in Fig. 3. S1 is required for S2 and S3.

Then, we constructed finer groups within each skill group according to their knowledge concept order. The concept graph of the second phase of the skill group is shown in Fig. 4. Take the course chapter concept graph of S3 as an example, Ch05 is required as a prerequisite for Ch04 and Ch09. And Ch04 is the prerequisite of Ch10. Finally, Ch04 and Ch09 are prerequisites of Ch11.

Skill group concept graph obtained from the first phase is then used to generate the ideal attributes matrices shown in Fig. 5 and the most important attributes: Examinee, Ideal Item Response Vectors, Total Scores, and Examinee attributes. The Examinee attributes of the skill group concept graph of the first phase, which includes E1, E2, E3, and E4 is encoded as 100, 110, 101, and 111. Their ideal item

Fig. 3 Skill group concept of the first phase

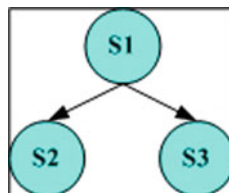
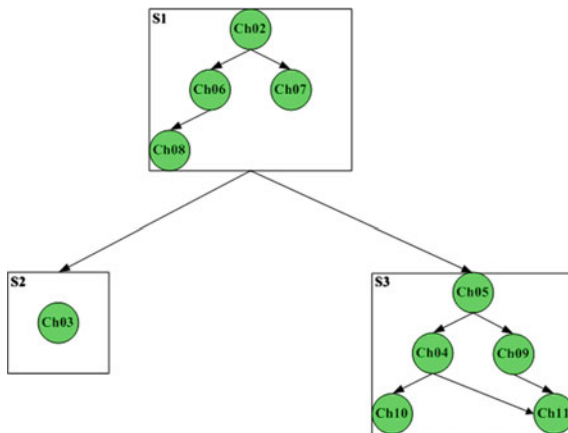


Fig. 4 Course chapter concept of the second phase



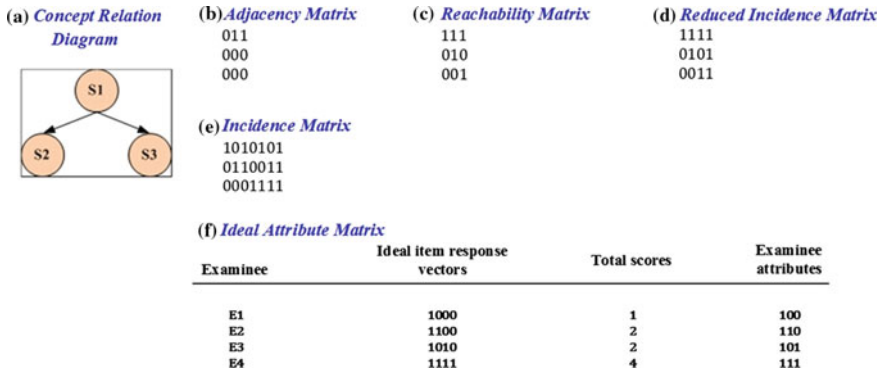


Fig. 5 The ideal attribute matrix based on the inference of rule-space model for the first phase of the skill group

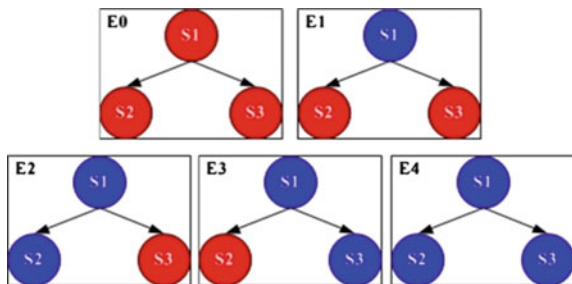
response vectors are 1000, 1100, 1010, and 1111, respectively. The total scores of E4 was then 4, which is a high score.

According to the Rule-Space Model, learners have to pass the learning of the root node by practicing node 1 (the blue node). However, we have to consider the situation in which learners passed none nodes. Therefore, both of the full-red and full-blue trees must be included in Fig. 6. As a result, the total score of E0 is 0000, which represents 0 in total score.

The second phase course chapter concept graphs can generate three sets of ideal attributes matrices through the Rule-Space Model respectively. Take S3 as an example, the matrices shown in Fig. 7 can be derived. The examinee attributes are 10000, 11000, 10100, 11100, 11010, 11110, 11101, and 11111, respectively. The ideal item response vectors are 10000000, 11000000, 10100000, 11110000, 11001000, 11111100, 11110010, and 11111111. The vectors represent total scores: 1, 2, 2, 4, 3, 6, 7, and 8. Hence, S3 acquired the high score 8.

Accordingly, S3, which has a full-red tree, must be included in Fig. 8. The examinee attributes of S3 was thus 00000 which the ideal item response vectors 00000000, which represents 0 in total scores.

Fig. 6 Blue-Red trees from the first phase of the skill group



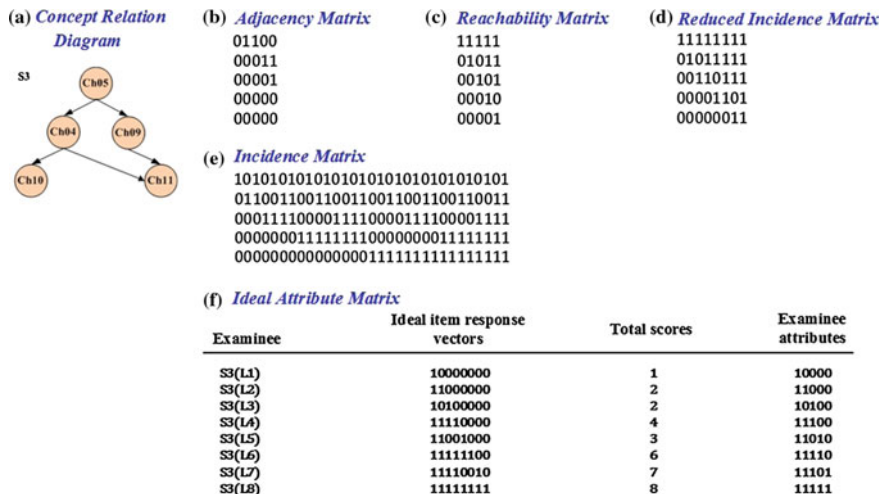


Fig. 7 The ideal attribute matrix based on the inference of rule-space model for the second phase of chapters within S3 skill group

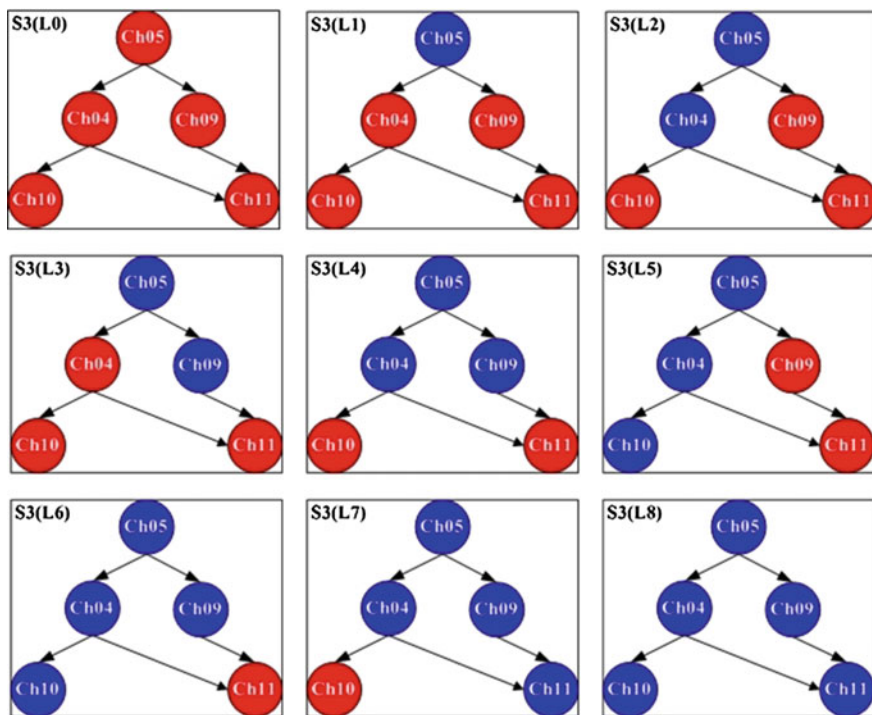


Fig. 8 Blue-Red trees from the chapters within S3 skill group of the second phase

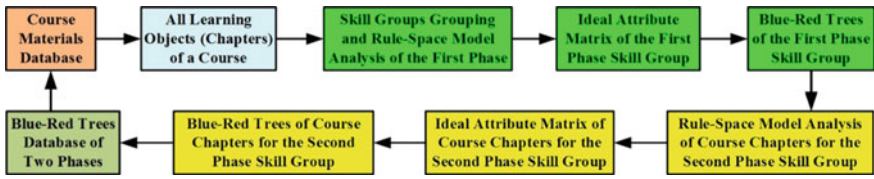


Fig. 9 Flowchart of the production of two phases Blue-Red trees Database

As shown in Fig. 9, we can use Rule-Space Model to generate the “ideal attribute matrices of the first phase skill group” and “the Blue-Red tree of the first phase skill group” for all learning objects of each course. Then, we derive “ideal attribute matrices of the second phase skill group” and “the Blue-Red tree of the second phase skill group” based on the Rule-Space Model analysis for the courses in the second phase skill group. Then, all generated Blue-Red trees will be stored in a Blue-Red tree database to be used as the basis of the developed online course learning system.

4 Inference of the Best Learning Path Based on Rule-Space Model

Definition 1 Assume there are N learning objects numbered from L_i with $i = 1, 2, 3, \dots, N$. Based on the analysis of Rule-Space Model, we have the hierarchical relationship concept diagram of the combination of learning states. We then define the reasonable learning path (LP) consisted of N learning objects as $LP = (S_1 \rightarrow S_2 \rightarrow \dots \rightarrow S_k \rightarrow \dots \rightarrow S_N)$ in which $k = 1$ to N , S_k represents the k -th learning object and S_k expresses a learning object of L_i .

Definition 2 Assume the learning order of N learning objects is encoded as $(S_1, S_2, \dots, S_k, \dots, S_N)$ based on Definition 1, then the Relation Weight (RW) of every learning object S_k and all learning objects is represented as an $1 \times N$ hierarchical matrix $RW_{S_k} = [W_{S_k_S_1}, W_{S_k_S_2}, \dots, W_{S_k_S_k}, \dots, W_{S_k_S_N}]$. The values within matrix RW_{S_k} are expressed the relation rates of the learning object S_k with the relative order of every learning object $(S_1, S_2, \dots, S_k, \dots, S_N)$, and $0 \leq W_{S_k_S_1}, W_{S_k_S_2}, \dots, W_{S_k_S_k}, \dots, W_{S_k_S_N} \leq 1$.

Each relation ratio in Relation Weight (RW) of every learning object S_k represents the values of relation ratio between the learning object S_k and all learning objects. The relation ratio between the learning node and itself is value “1”, which is the highest. The relation ratio between a learning node and its parent learning node is higher and the ratio between a learning node and its child learning node will be lower as value “0”. Accordingly, it can be seen that the Relation Weight (RW) of learning object S_1 is regard as $RW_{S_1} = [W_{S_1_S_1}, W_{S_1_S_2}, \dots, W_{S_1_S_k}, \dots, W_{S_1_S_N}]$, the Relation Weight (RW) of learning object S_2 is regard as $RW_{S_2} = [W_{S_2_S_1}, W_{S_2_S_2}, \dots, W_{S_2_S_k}, \dots,$

$W_{S_2_S_N}]$, so to the Relation Weight (RW) of learning object S_N is regard as $RW_{S_N} = [W_{S_N_S_1}, W_{S_N_S_2}, \dots, W_{S_N_S_k}, \dots, W_{S_N_S_N}]$.

As a result, based on the Rule-Space Model, we can analyze N learning objects and obtain all learning paths with reasonable learning order. The best learning order can be acquired with the Relation Weight (RW) between all pairs of learning objects.

Definition 3 For N learning objects with the learning order $(S_1, S_2, \dots, S_k, \dots, S_N)$, following Definition 1 and Definition 2, if there is a learning path in which the learning order of two adjacent learning objects is defined as $S_t \rightarrow S_{t+1}, t = 1, 2, 3, \dots, N-1$, and the Relation Weight (RW) of S_t, S_{t+1} and all learning objects are defined as $RW_{S_t} = [W_{S_t_S_1}, W_{S_t_S_2}, \dots, W_{S_t_S_t}, W_{S_t_S_{t+1}}, \dots, W_{S_t_S_N}]$ and $RW_{S_{t+1}} = [W_{S_{t+1}_S_1}, W_{S_{t+1}_S_2}, \dots, W_{S_{t+1}_S_t}, W_{S_{t+1}_S_{t+1}}, \dots, W_{S_{t+1}_S_N}]$, then

1. The Confidence Level (CL) of learning from S_t to S_{t+1} is

$$\begin{aligned} CL(S_t \rightarrow S_{t+1}) &= (W_{S_t_S_1} \times W_{S_{t+1}_S_1}) + (W_{S_t_S_2} \times W_{S_{t+1}_S_2}) + \dots \\ &\quad + (W_{S_t_S_t} \times W_{S_{t+1}_S_t}) + (W_{S_t_S_{t+1}} \times W_{S_{t+1}_S_{t+1}}) + \dots \\ &\quad + (W_{S_t_S_N} \times W_{S_{t+1}_S_N}) = \sum_{i=1}^N W_{S_t_S_i} \times W_{S_{t+1}_S_i} \end{aligned}$$

2. For each learning path, the Confidence Level (CL) of two adjacent learning objects are:

$$\sum_{t=1}^{N-1} CL(S_t \rightarrow S_{t+1}) = \sum_{t=1}^{N-1} \sum_{i=1}^N W_{S_t_S_i} \times W_{S_{t+1}_S_i}$$

3. The learning path with the highest value of Confidence Level (CL) is defined as the best learning path, and the max value of Confidence Level (CL) is defined as

$$Max \left(\sum_{t=1}^{N-1} CL(S_t \rightarrow S_{t+1}) \right) = Max \left(\sum_{t=1}^{N-1} \sum_{i=1}^N W_{S_t_S_i} \times W_{S_{t+1}_S_i} \right)$$

As shown in Fig. 10, according to knowledge concept arrangement of all learning objects for a courses in the LMD (Learning Materials Database), we generate the skill group concept graph of the first phase and the course chapter concept map within skill group of the second phase. Then, based on the learning path derivation algorithm of rule space model analysis, we analyze the order of learning objects and induct all learning paths. These paths are then put into the LPD (Learning Path Database). Then, we use the BLPDARSMA (Best Learning Path Derivation Algorithm of Rule Space Model Analysis) to figure out the only one best learning path from all learning path and put into the BLPD (Best Learning Path Database).

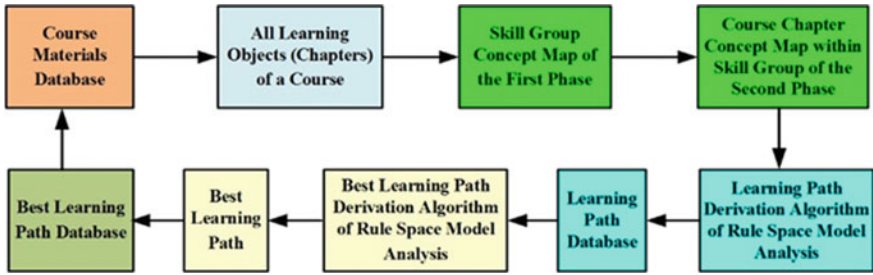


Fig. 10 The best learning path generation flow of the course chapters within skill group of the second phase

Furthermore, we used that for the OCLS (Online Course Learning System) to suggest the best learning paths for learners for their online self-learning.

5 Recommendation System with Learning Performance Analysis from Online Learning and Testing System

We have already proposed a recommendation system with learning performance analysis based on online learning and testing system that combines both the reliability and validity analysis and a feedback improvement mechanism. The build learning and test of recommendation system integrates the blue-red trees database of two phases and the (best) learning path database as shown in Fig. 11. In the figure, the inner circle shows the mechanism of self-improvement learning direction

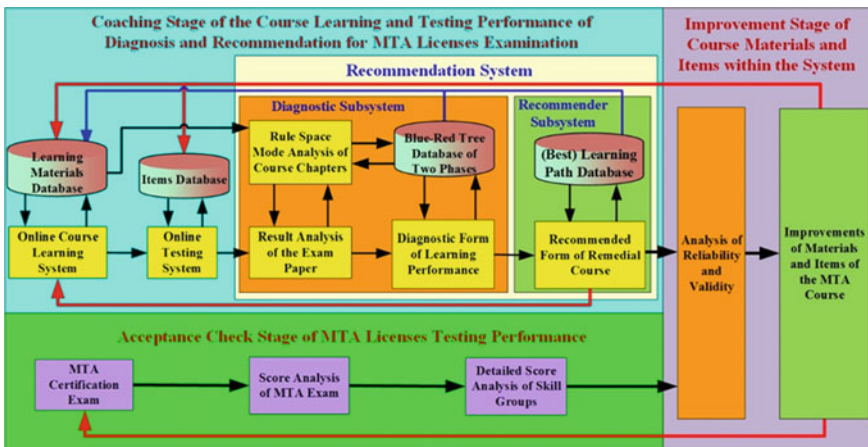


Fig. 11 Flowchart of the recommendation system with learning performance analysis from the online learning and testing system

for students. The outer circle then shows the improvement mechanism of materials and items for MTA Course and increase the performance of MTA license exam while the inner circle has improved positive performance.

In order to auto-analyze the learning performances of learner from the system, we construct a brand-new processing flow of the RS (Recommendation System) from Professor Kikumi K. Tatsuoka’s Rule Space Model [10, 13–15], depicted in Fig. 11. The system can show better or worse learning performance objects when learners have finished these learning and generated Blue-Red trees through strong and weak analysis of learning performances by feedback automatically. They are able to self-learn through the recommendation of best learning path from the recommendation system when learners agree to provide feedback of the RFRC (Recommended Form of Remedial Course). The learner can fit self-regulation processing of the Scaffolding Theory [16, 17] and achieving social learning effectiveness, simultaneously. Therefore, it will be more possessed of functions of self-learning and feedback than the original online test system.

The proposed system consists of three phases: CSCLTPDRMLE (Coaching Stage of the Course Learning and Testing Performance of Diagnosis and Recommendation for MTA Licenses Examination), ACSMLTP (Acceptance Check Stage of MTA Licenses Testing Performance), and ISCMIS (Improvement Stage of Course Materials and Items within the System). Among them, the CSCLTPDRMLE (Coaching Stage of the Course Learning and Testing Performance of Diagnosis and Recommendation for MTA Licenses Examination) in turn contains an OCLS (Online Course Learning System), an OTS (Online Testing System), and a RS (Recommendation System). The RS (Recommendation System) then consists of two sub systems: the Diagnostic Subsystem and the Recommender Subsystem. The execution flow is shown in Fig. 12. Whenever students enter the OCLs (Online Course Learning System) for self-learning, the RS (Recommendation System) will recommend best learning path according the BLPD (Best Learning Path Database) to students to promote their learning performance. After finishing a self-learning and a self-testing through the OCLS (Online Course Learning System) and the OTS (Online Testing System) from a student, the Diagnostic Subsystem can utilize the Rule Space Mode Analysis of Course Chapters and a blue-read tree and corresponding ideal attribute matrix for their learning performance from the DFLP (Diagnostic Form of Learning Performance) will be further generated

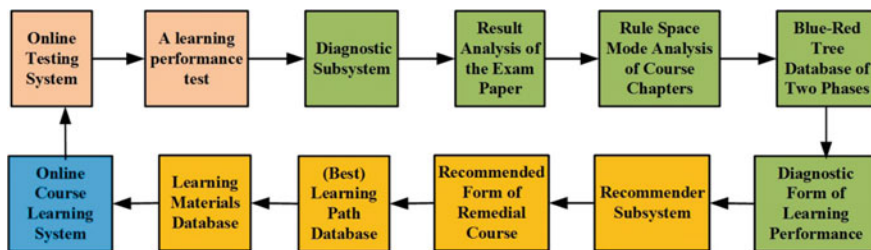


Fig. 12 Flowchart of the diagnostic subsystem and the recommender subsystem

based on the Blue-Red Tree Database of Two Phases as the basis of remedy. Then, the Recommender Subsystem will generate the RFRC (Recommended Form of Remedial Course) based on the suggestion of (Best) Learning Path Database for recommended learning of best learning path from various chapters. This can be used for further self-learning of Computer-Supported Personalized Learning (CSPL).

6 Preliminary Experimental Analysis

To effectively analyze the relation between the performance of remedy for MTA and the pass rate of the certificate, a recommendation system that integrates two-phase blue-red tree rule space model and the recommendation system of the best learning path is proposed in this research. In the first phase, we utilize the skill groups obtained from certiport and identified three skill groups. With the rule space model, a concept diagram and a blue-red tree are generated. Based on the skill groups and the correlations between 10 chapters of MTA, we generate three groups of MTA courses. We analyze each group with the rule space model to obtain learning concept diagram and blue red tree for each to design this course in the second phase. Then, based on the rule space model, we generate learning objects for each skill group and define the weight of them. By calculating the confidence level between each adjacency pairs, web can induct the best learning path. In turn, we use the path in the online course learning system to recommend the best learning order to students.

The proposed recommendation system was adopted to improve the pass rate of MTA. We performed the experiment in “Introduction to Network” course, which is required for 161 students including two daytime classes of freshmen and one evening time class of sophomores. Once students complete online self-learning and testing based on the rule-space model, the system generates a learning performance report and recommend students to improve their weaknesses as shown in Fig. 13.

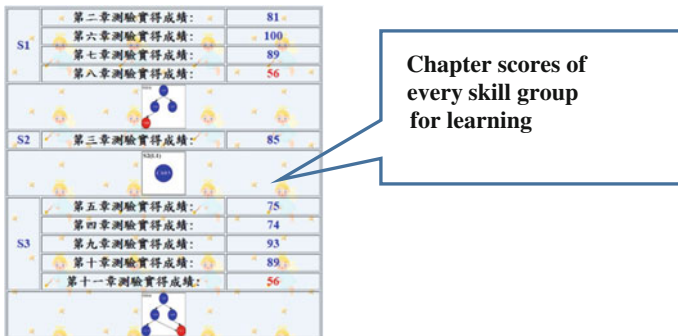


Fig. 13 An example shows the learning performance of Blue-Red trees to correspond to the chapters of skill group from MTA course for a student

The example of the figure appears a diagnostic table including skill groups of a MTA course and learning performance of Blue-Red trees for a learner. A passed chapter is shown as a blue node and the learner must re-learn while a failed chapter is shown as a red node

The differences between pre-tests and post-tests were utilized to evaluate the performance of the proposed method, as shown in Table 1. Before adopting the proposed system, the pre-test mean scores of three classes were 30.65, 35.40, and 30.88 points. With the proposed system, the post-test mean scores were enhanced to 58.54, 59.10 and 54.10 points, respectively. Vast amount of improvements were observed and the improvement rate were 90.10, 66.95 and 75.19 % respectively. Furthermore, the pass rate was increased to 90.76 %, which was a very good result. After three months of counseling and mastery learning, they participate in Networking Fundamentals certification exam of MTA from certification agency of Certiport Portal Site, and there are very good results. The average scores of three classes are about 84.34, 83.29 and 84.53 points separately, and the obtain rates (over 70 points) are 88.52, 92.73 and 91.11 % separately. Therefore, it expresses that 161 students can obtain 146 MTA licenses and the total obtain rate is about 90.68 %.

According to Table 1, we draw a conclusion that the proposed method suits the certification course well. The pass rate of the third grade and fourth grade students was obviously higher than the first and second grade students. The fact shows that Networking Fundamental of MTA covers basic networking concepts, and students of the third and fourth grade should be more familiar with the course. Among them, the correct rate of each skill group can not only help students adjust their learning directions but also help the adjustment of course materials and items of MTA.

Table 1 Learning performance analysis of coaching networking fundamentals course for MTA licenses

Class	Daytime class 1	Daytime class 2	Evening time class
Students	Freshmen	Freshmen	Sophomores
Numbers	61	55	45
Pre-test mean scores	30.65 points	35.40 points	30.88 points
Post-test mean scores	58.54 points	59.10 points	54.10 points
Rates of progress (%)	91.00	66.95	75.19
Average scores of MTA	84.34 points	83.29 points	84.53 points
Average correct rate of S1 from MTA (%)	90.92	88.20	88.82
Average correct rate of S2 from MTA (%)	81.70	82.31	84.24
Average correct rate of S3 from MTA (%)	76.93	76.22	77.60
Numbers of obtaining MTA	54	51	41
Obtaining rates of MTA (%)	88.52	92.73	91.11

As shown in the results, it is for sure that the combination of our recommendation system and the Master theory does help increase the chance for students to pass the MTA exam. The fact shows that the average correct rate of S3 from MTA for three classes are lower compared to S1 and S2, as shown in Table 1. It expresses that students are weaker in the learning of S3 with “Protocols and Services”. Therefore, they must strengthen the learning of S3 chapters. Among them, the correct rate of each skill group can not only help students adjust their learning directions but also help the adjustment of course materials and items of MTA.

7 Conclusion and Future Work

We propose the recommendation system to remedy and analyze students’ learning performance of MTA and improve the pass rate of MTA exam based on the rule space model of two phase blue red tree and the recommendation system of the best learning path. In the first phase, we utilize the skill groups obtained from Certiport and three skill groups are identified. Based on the rule space model, a concept diagram and a blue-red tree are generated. In the second phase, based on the skill groups and the correlations between 10 chapters of MTA, we generate three groups of MTA courses. We analyze each group with the rule space model to obtain learning concept diagram and blue red tree for each to design this course. Then, based on the rule space model, we generate learning objects for each skill group and define the weight of them. By calculating the confidence level between each adjacency pairs, web can induct the best learning path. In turn, we use the path in the online course learning system to recommend the best learning order to students. The correct ratio of each group can not only help students improve their learning performance but also help the improvement of the course materials and tests of MTA. Our experiments show the effectiveness of our system and method.

In the future, we will improve the proposed algorithm and apply it to different domains. One of our goals is to adopt it in the training recommendation system for swimmers or in museum guiding.

Acknowledgments This work was supported by the Ministry of Science and Technology, R.O.C., under Grant MOST 103-2511-S-262-003-.

References

1. Doignon JP, Falzague JCI (1985) Spaces for the assessment of knowledge. *Int J Man-Machine Stud* 23:175–196
2. Zanker M, Jessenitschnig M, Jannach D, Gordea S (2007) Comparing recommendation strategies in a commercial context. *IEEE Intell Syst* 22(03):69–73. doi:[10.1109/MIS.2007.49](https://doi.org/10.1109/MIS.2007.49)

3. Mo Y, Chen J, Xie X, Luo C, Yang LT (2014) Cloud-Based mobile multimedia recommendation system with user behavior information. *IEEE Syst J* 8(1):184–193. doi:[10.1109/JSYST.2013.2279732](https://doi.org/10.1109/JSYST.2013.2279732)
4. Chien SY (2004) A content-based recommendation method for browsing guidance in web-based E-learning system. *Computer Science and Information Engineering*, Mingchuan University, Mingchuan
5. Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Comput* 7(1):76–80. doi:[10.1109/MIC.2003.1167344](https://doi.org/10.1109/MIC.2003.1167344)
6. Khalid O, Khan MUS, Khan SU, Zomaya AY (2014) OmniSuggest: a ubiquitous cloud-based context-aware recommendation system for mobile social networks. *IEEE Trans Serv Comput* 7(3):401–414. doi:[10.1109/TSC.2013.53](https://doi.org/10.1109/TSC.2013.53)
7. Zhang D, He T, Liu Y, Lin S, Stankovic JA (2014) A carpooling recommendation system for taxicab services. *IEEE Trans Emerg Topics Comput* 2(3):254–266. doi:[10.1109/TETC.2014.2356493](https://doi.org/10.1109/TETC.2014.2356493)
8. Park S, Kang S, Kim YK (2006) A channel recommendation system in mobile environment. *IEEE Trans Consum Electron* 52(1):33–39. doi:[10.1109/TCE.2006.1605022](https://doi.org/10.1109/TCE.2006.1605022)
9. Wang Z, Liao J, Cao Q, Qi H, Wang Z (2015) Friendbook: a semantic-based friend recommendation system for social networks. *IEEE Trans Mob Comput* 14(3):538–551. doi:[10.1109/TMC.2014.2322373](https://doi.org/10.1109/TMC.2014.2322373)
10. Tatsuoka KK (1983) Rule space: an approach for dealing with misconceptions based on item response theory. *J Educ Meas* 20(4):345–354
11. Chen YH, Deng LY, Yen NY, Weng MM, Kao BC (2012) One-to-One complementary collaborative learning based on blue-red multi-trees of rule-space model for MTA course in social network environment. In: *The 2012 international conference on human-centric computing (HumanCom 2012)*. Gwangju, Korea, September 6–8 2012, pp 231–239
12. Chen YH, Deng LY, Yen SH, Chen PW, Kao BC, Hsieh YC (2012) Complementary collaborative learning based on blue-red multi-trees inference and performance analysis for social network-case in MTA course. In: *The 5th IET international conference on Ubi-media computing (IET U-Media 2012)*. Xining, China, August 16–18 2012, pp 63–68
13. Tatsuoka KK, Birenbaum M, Tatsuoka MM, Baillie R (1980) A psychometric approach to error analysis on response patterns, (Research Report 80-3-ONR). University of Illinois, Computer-based Education Research Laboratory, Urbana, I11
14. Tatsuoka KK (1981) An approach to assessing the seriousness of error types and predictability of future performance, (Research Report 81-1-ONR). Urbana, 111, University of Illinois, Computer-based Education Research Laboratory (February 1981)
15. Sheehan KM, Tatsuoka KK, Lewis C (1993) A diagnostic classification model for document processing skills. *Educational Testing Service Report* (October 1993)
16. Winnips JC (2001) Scaffolding by design: a model for WWW-based learner support, Dissertation. University of Twente, Enschede
17. Winnips JC, Collis BA, Moonen JJCM (2000) Implementing a scaffolding-by-design model in a WWW-based course considering costs and benefits. In *Bordeau J, Heller R (eds) Processing of ED-MEDIA 2000*. Charlottesville, AACE, VA, pp 1147–1152

A Algorithm of Detectors Generating Based on Negative Selection Algorithm

Wu Renjie, Guo Xiaoling and Zhang Xiao

Abstract There are a lot of redundancy and over all issues in the artificial immune system (AIS) because of using the traditional negative selection algorithm (NSA) to generate detectors. It is the main reason for the high false percentage and high missed percentage in the intrusion detection system (IDS). Therefore, an improved immune detector generation algorithm is put forward. By calculating the optimal size of mature detector set and using the twice match in the improved algorithm. The efficiency of the IDS can be guaranteed. In the last, simulation experiments show that the improved algorithm can cover the more nonself and had a higher detection rate in the IDS.

Keywords Artificial immune system (AIS) · Intrusion detection system (IDS) · Negative select algorithm (NSA) · R-contiguous algorithm

1 Introduction

Biological Immune System (BIS) is a complex organic system composed of immune molecules, immune cells and immune organs. It protects the body from viruses and germs. And it is dynamic, adaptation, distribution and self-organization [1]. The study found that the intrusion detection system (IDS) has the excellent feature similar to the BIS. They all maintain the safety and stability of their own life with Distinguishing self and nonself [2]. In 1994, Forrest introduced artificial immune idea to the intrusion detection and put forward the famous negative

W. Renjie · G. Xiaoling (✉) · Z. Xiao
School of Information Science and Engineering, Hebei North University,
Zhangjiakou 075000, Hebei, China
e-mail: guoxiaoling0494@163.com

selection algorithm. The negative selection algorithm has attracted wide attention since been put forward. But the development in this field in China is still in the initial stage.

2 The Analysis of Negative Selection Algorithm

Negative Select Algorithm (NSA) is the core algorithm in artificial immune system (AIS). It simulae self tolerance of the immune cells. The algorithm first proposed by S. Forrest in 1994. Since proposed, it become the focus and hotspot research in this field. The algorithm includes two stages: the mature detector generation stage and the abnormal detection stage. Figure 1 shows the flow chart of the two stages.

By the analysis it is not difficult to find: (1) In the traditional negative selection algorithm, the string generated by the random way. So it exist duplication and overlap inevitably. This method increase the miss percentage and reduces the detection efficiency of the system. (2) The size of mature detector determines the operating time of the system. It is a worthy depth-discussed study that the relationship between the size of mature detector and detection efficiency. Improvement of negative selection algorithm this paper focuses on the above two problems.

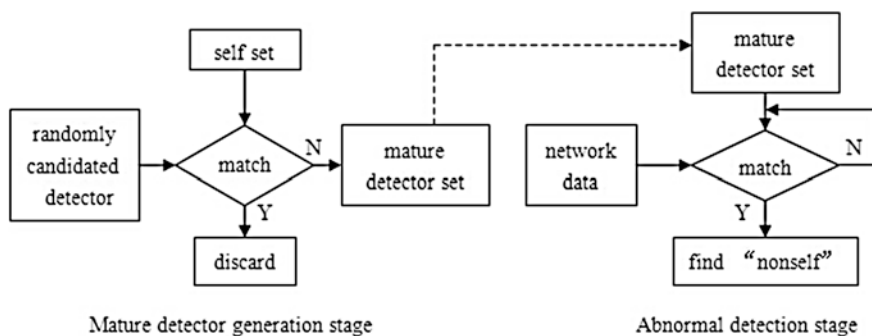


Fig. 1 Negative select algorithm (NSA)

3 An Improved Algorithm for Generating Immune Detectors

3.1 Related Definitions

Definition 1 Self/nonself: the problem domain $U \subseteq \{0, 1\}^l$, Normal network behavior constitute the self set S , Abnormal network behavior or intrusion actions constitute a non self sets N , $S \cup N = U$, $S \cap N = \emptyset$.

Definition 2 r -contiguous algorithm: If it exists some the same continuous bits in the string X and Y , that is, X and Y is r -contiguous. We called matching the success. Or the matching is a failure. r is called the matching threshold [3].

For example : X : 10010111
 Y : 01010100

when $r \leq 4$, X and Y is a successful matching.

Definition 3 String matching Percentage p : In the r -contiguous algorithm, the matching percentage of any two strings X and Y is defined as p .

$$P = \left(1 - \frac{1}{m}\right) * m^{-r} * (l - r) + m^{-r} = m^{-r} \left[\frac{(l - r)(m - 1)}{m} + 1 \right]$$

Among them, $m \in \{0, 1\}^l$, l is the string length, r is the matching threshold. As we discussed the problem is in the binary, so m is 2.

$$P = 2^{-r} \left(\frac{l - r}{2} + 1 \right) \quad (1)$$

Definition 4 The missed Percentage p_m : the Percentage of an nonself string cannot match with mature detector defined as p_m .

$$p_m = (1 - p)^n \quad (2)$$

Definition 5 The optimal detector size n : In order to ensure the system efficiency and time, the minimum detector set must cover the largest nonself space. n represent the optimal detector size. Operation the \ln to the formula (2):

$$\ln p_m = n \ln(1 - p), n = \frac{\ln p_m}{\ln(1 - p)} \quad (3)$$

then bring formula (1) to (3),

$$n = \frac{\ln p_m}{\ln[1 - (l - r)2^{-r-1} - 2^{-r}]} \quad (4)$$

As can be seen from Eq. (4), the mature detector set is only related to the parameters l and r . And it is irrelevant to the self set and the candidate detector size under certain missed percentage. All of that indicates the mature detector size is determined. And n is the optimal number [4].

3.2 Twice Matching of r-contiguous Algorithm

The candidate detectors are generated randomly. So it is Possible that there is a repeat phenomenon. When the candidate detectors become the mature detector by negative selection algorithm, it will also have mutual matching problem. It is the main reason for a large number of redundant mature detector. And it is also the main reason for the high false percentage and missed percentage. So it is necessary to conduct twice *r-contiguous* algorithm. we can match the mature detector with one of the existing detector set. If they does not match, we bring the mature detector into the existing detector set. This method can improve the “nonself” coverage [5], shown in Fig. 2.

1. Do the second *r-contiguous* algorithm for the current D detector and one of the mature detector set. The size of the existing mature detector set is n'
2. Initialize the matching threshold r , the length of a string l , the missed percentage p_m
3. Begin

```

While (  $n' < n$  ) // n is the size of mature detector set
{
  For( $m = 1; m < n'; m ++$ )
    If  $\text{match}(D, D'[m]) \geq r$ 
      (discard,
       Break;)
    Else
      (D was added to the mature detector  $D', n' ++$ )
}
End

```

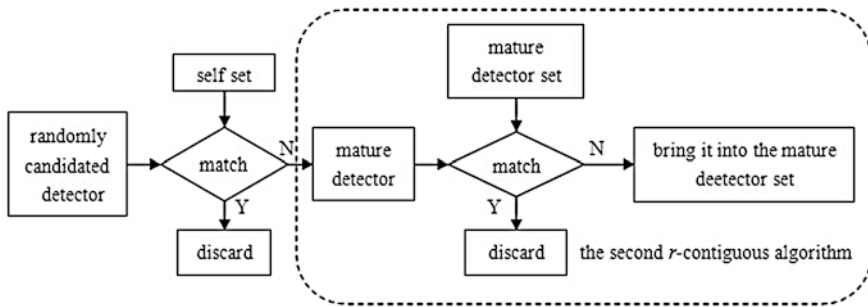


Fig. 2 The second r -contiguous algorithm

4 The Simulation Experiment

4.1 Initialization Parameters

Experiments use the Window XP system and Visual C++6.0 programming tool. The experiment data are chosen from the KDD CUP99 randomly. In order to ensure the accuracy of the experiment, three groups data connections whit diverse scale are chosen. The experiment data are shown in Table 1. The self set is used to generate mature detector set. Normal connections and abnormal connections are used to test for the intrusion detection.

In addition, we must initialize some other parameters in the process of the experiment. Because of each data connection in the KDD CUP99 data set is described with 41 characters and l represents the length of the string, so l is initialized to 4. r is initialized to 9 for the matching threshold. The missed percentage is initialized to 8 %; According to the formula (4): $n = \frac{\ln p_m}{\ln[1-(l-r)2^{-r-1}-2^{-r}]} = 75$.

5 Experimental Results and Analysis

True Percentage (p_t): the system can correctly distinguish self and nonself;

False Percentage (p_f): self is deemed to nonself;

Missed Percentage (p_m): nonself is deemed to self.

Table 1 The experimental data set

Experimental data	Self set	Normal connection	Anomalous connection
1	25,000	10,500	5500
2	35,000	20,500	12,500
3	45,000	30,500	23,500

Above three parameters are used to evaluate the improved algorithm and the traditional algorithm. The calculation of these parameters are as follows:

$$\begin{aligned}
 (p_t) &= \frac{\text{the number of self and nonself correctly detected}}{\text{the total number of test}} \\
 (p_f) &= \frac{\text{the number of self deemed to nonself}}{\text{the number of self}} \\
 (p_m) &= \frac{\text{the number of nonself deemed to self}}{\text{the number of nonself}}
 \end{aligned}$$

The improved algorithm and the traditional algorithm comparison as shown in Table 2. Because of reducing the overlap of the mature detector, we can found that the improved algorithm has greatly improved the detection efficiency.

Table 2 Comparison between NSA and improved algorithm

Data	Method	p_t (%)	p_f (%)	p_m (%)
1	NSA	96.12	9.32	7.12
	Improved NSA	96.76	7.05	5.63
2	NSA	94.01	10.11	8.56
	Improved NSA	96.32	7.45	7.23
3	NSA	93.19	9.88	5.89
	Improved NSA	95.78	9.65	6.02
Average value	NSA	94.44	9.77	7.19
	Improved NSA	96.29	8.05	6.29

Fig. 3 The average value of two algorithms

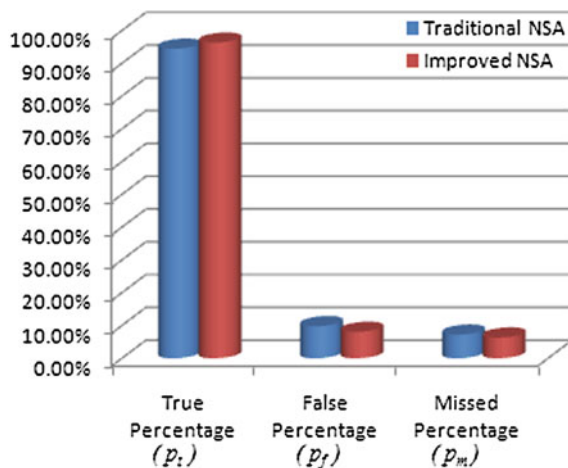


Figure 3 shows the histogram of three average value. It is easy to find that improved algorithm raise the true percentage and reduce the false percentage and missed percentage.

6 Conclusion

Under considering the optimal scale of mature detector set, the improved negative selection algorithm reduces the duplication and redundancy among of the mature detector set by the twice r-contiguous algorithm. The efficiency of intrusion detection is also raised greatly. The problem is that inherent defects of r-contiguous algorithm can not be avoided. And system time will increase due to adding another the matching process these issues will be discussed and improved in the further study.

Acknowledgments Foundation item: The 2013 Natural Science Foundation Project of Hebei North University (Q2013006) and the paper is supported by population health information engineering technology research center in Hebei North University.

References

1. Li T (2004) Computer immunology. Electronics Industry Publishing House, Beijing
2. Forrest S, Parelson A, Allen L et al (1994) Self-nonsel self discrimination in a computer. In: Proceedings of the 1994 IEEE symposium on research in security and privacy, IEEE Compute Society Press, Los Alamos, CA
3. Hofmery S, Forrest S (2000) Architecture for an artificial immune system. *Evol Comput* 7 (1):45–68
4. Jiang Y, Zhao J, Ma Y et al (2014) Generation model of minimum detector based on immune recognition. *Computer engineering and design* 35(5):1598–1601
5. Jin J, Han H, Cui Y (2015) Application of improved negative select algorithm in intrusion detection system. *Electronic Design Engineering* 23(1):7–9

A Comparative Study on Disease Risk Model in Exploratory Spatial Analysis

Zhisheng Zhao, Xiao Zhang, Yang Liu, Junhua Liang, Jiawei Wang and Yaxu Liu

Abstract The present work mainly focuses on the issue of risk model in spacial data analysis. Through the analysis on morbidity data of influenza A (H1N1) across China's administrative regions from 2009 to 2012, a comparative study was carried out among four different estimators SMR, EBPG, EBLN and EBMarsshall as risk model to explore and make improvements for the problems of risk model and pattern of survival distribution in spacial disease analysis. By using R programming language, the feasibility of the above analysis method was verified and the variability of the estimated value generated by each model was calculated. The research on spacial variability of disease morbidity is helpful in detecting epidemic area and forewarning the pathophoresis of prospective epidemic disease.

Keywords Spatial data analysis · Risk model · Disease graphics · R programming language

1 Introduction

Exploratory Spatial Data Analysis (ESDA), using method of health statistics to realize the description and visualization of disease spatial distribution pattern, reveal the spacial distribution of disease risk situation, study data feature and detect the spacial clustering, anomaly and interaction mechanism of data, has become a very important research area in the age of big data. And the choice of model is crucial in the process of practical study. As new research contents for diagnosis and prevention policy in the field of intelligent information processing, spatial disease

Z. Zhao (✉) · X. Zhang · Y. Liu · J. Liang · J. Wang · Y. Liu
School of Information Science and Engineering, Hebei North University,
Zhangjiakou, China
e-mail: zhaozhisheng_cn@sina.com

X. Zhang
e-mail: 780117251@qq.com

data analysis and graphics haven't been studied a lot by Chinese scholars, especially for the issue of spacial model. Some problems related to spacial data analysis cannot be solved directly by observation of data itself, but the process of the observed data generated [1]. Existing mature models and algorithms are mainly applied to the processing of 2-D structured or unstructured data, thus the spacial process of statistical inference is challenging. The choice of model has to be based on scientific investigation and judgement on hypothesis by thorough examination. The choice of model may be different for various scientific areas, and wrong choice of model can hardly lead us to any meaningful conclusion.

In this article, a comparative study on risk model was carried out based on the analysis on morbidity data of influenza A (H1N1) across China's administrative regions from 2009 to 2012 to explore and make improvements for the problems of risk model and pattern of survival distribution in spacial disease analysis. As a newly emerging comprehensive research field, the study on spacial variability of disease morbidity is helpful in detecting epidemic area and forewarning the pathophoresis of prospective epidemic disease, with great theoretical and practical significance.

2 Research Methods and Data Sources

2.1 Research Methods

Exploratory spatial data analysis, the process of which is to analyze and verify the characteristics of spacial information and to form the structure of a deterministic model, is essentially a data-driven analysis method. ESDA methods can be divided into two categories, global statistics and local statistics. Global statistics mainly explores distribution characteristics of a particular attribute in the region, while local statistics probes whether regional information changes smoothly (homogeneous) or mutationally (heterogeneous) [2] by analysis on regional information respectively. And the choice of model will significantly influence the results of analysis.

2.2 Data Sources

National statistical data of influenza A (H1N1) (morbidity data of A (H1N1) on 35 monitoring sites from 2009 to 2012) are shown in Table 1, coming from Datatang website. Data of population across national administrative regions from 2009 to 2012 are shown in Table 2, coming from the official website of National Bureau of Statistics of the People's Republic of China. Results throughout this article were obtained by adopting R programming language.

Table 1 National statistical data of influenza A (H1N1) (morbidity data of A (H1N1) on 35 monitoring sites from 2009 to 2012) (unit: person)

Region	Year			
	2009	2010	2011	2012
Beijing	10,838	288	259	124
Tianjin	959	53	190	12
Hebei	3722	128	262	14
Liaoning	2190	43	33	0
...
Tibet	4966	2	0	0
Ningxia	1426	84	197	78
Xinjiang	5133	121	274	0

Data source datatang.com

Table 2 Data of population across national administrative regions from 2009 to 2012 (unit: ten thousand people)

Region	Year			
	2009	2010	2011	2012
Beijing	1860	1962	2019	2069
Tianjin	1228	1299	1355	1413
Hebei	7034	7194	7241	7288
Liaoning	4341	4375	4383	4389
...
Tibet	296	300	303	308
Ningxia	625	633	639	647
Xinjiang	2159	2185	2209	2233

Data Source www.stats.gov.cn

3 Comparative Research on Different Models

In the A (H1N1) spatial disease analysis research, the first was the division of research area. Different attributes were taken into consideration when data to be categorized into groups or concept hierarchy, with no overlapping. The selection of attributes was measured by the correlation of the problem domain. Important attributes were selected and data unrelated to the problem domain were eliminated, and data were sorted and classified according to the first and the second attributes. Disease risk reflects the morbidity or mortality rate within a period of time. Therefore, the basic data must include the number of people at risk and the number of cases at different regions. When spacial data were sorted out completely, the appropriate size indicator was selected to stratify data according to time and regional property. Morbidity rate was calculated by yearly data and applied to the disease regional study of 31 regions nationwide.

We used P_{ij} and O_{ij} to represent the number of cases in the area i and population at year j , P_i and O_i to represent population and number of cases in each area, with P_+ and O_+ representing total population and the number of cases throughout all regions. The disease risk assessment was generated by studies on the population and number of cases regionally, and we used the formula $E_i = P_i r_+$ to calculate the expected number of cases in region i , in which $r_+ = \frac{O_+}{P_+}$ represented morbidity rate, then compared the number of cases from statistical sources and expectation.

If the group is divided by year, we can use a similar procedure to analyze the population and case distribution for each year, using $r_i = \frac{\sum_j O_{ij}}{\sum_j P_{ij}}$ to calculate incidence rate regionally. Incidence rate can be calculated through dividing the number of cases in layer j by population in layer j , then the expected number of cases in area i can be calculated by $E_i = \sum_j P_{ij} r_j$.

3.1 Statistical Model of Poisson Distribution

We can consider that the number of cases from zone i and time j fitting the Poisson distribution with average $\theta_i E_{ij}$. The relative risk being 1 means the risk of this region is in average. If the relative risk is significantly greater than 1, it means the region has a certain risk. Assume the risk factors have no interaction between regional populations, then the relative risk θ_i only depends on its own region. Basic risk estimation given by the standard mortality of an area (Standard Mortality Ratio) can be calculated by the formula of $SMR_i = \theta_i / E_i$ [3], because the computing of morbidity rate is used to estimate the relative risk, thus data involving cases is usually taken as a numerator, and the population as a denominator. Figure 1 is the SMRs of A (H1N1) in 2009–2012 nationwide.

Fig. 1 China provinces' standard morbidity ratio of influenza A (H1N1) in 2009–2012

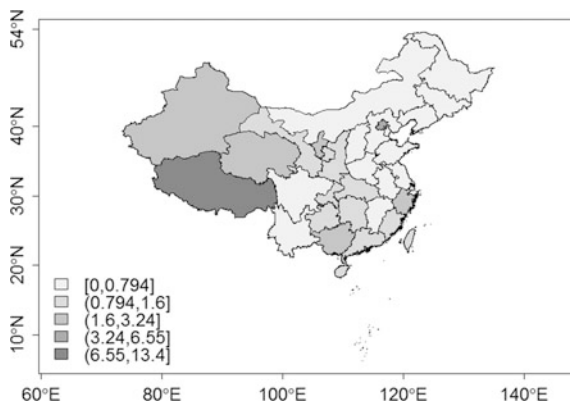


Fig. 2 The confidence interval of SMR

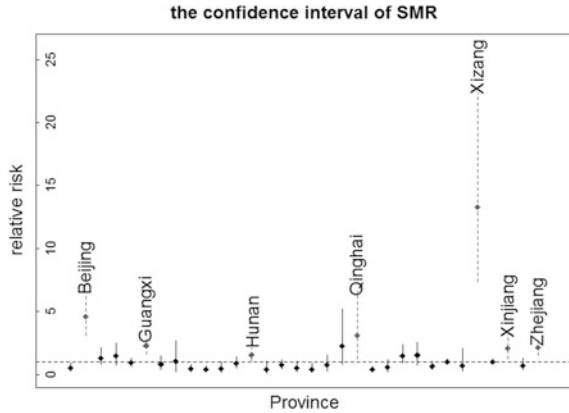


Figure 1 shows spatial distribution situation of influenza A (H1N1) in 2009–2012 nationwide. We can see in China the distribution of H1N1 flu is uneven, with most of the higher risk areas in western and southern China, among which Tibet Autonomous Region, Xinjiang Uygur Autonomous Region, Beijing, Qinghai, Guangxi and Zhejiang take significantly high risk, thus the economic investment and the level of medical assistance may be needed to improve in these areas.

Because θ_i obeys Poisson distribution, the confidence interval of the SMR can be calculated. Figure 2 shows the 95 % confidence interval of SMR, where black dot represents SMR in each region and dotted line indicates the confidence interval being significantly higher than 1. Regions with obviously higher risk (minimum value of ci greater than 1) is marked by a dotted line and corresponding name.

The above graph can also clearly indicate the high-risk areas of influenza A (H1N1) in China, being almost the same as shown in Fig. 1, with Beijing, Guangxi, Hunan, Qinghai, Xinjiang and Zhejiang proved to take high morbidity rate of A (H1N1) once again. Concluded from the above figures, Tibet is the region with the highest risk of influenza A (H1N1) and the SMR confidence interval of this area is much higher than in other parts of China, which suggests that the medical and health condition of the Tibetan region is the lowest in China. The special geographical location and climate characteristics also increase morbidity and mortality rates in Tibet.

3.2 Statistical Model of Poisson-Gamma

Poisson distribution, containing only one parameter of Poisson distribution, is usually the first choice to describe the number of rare events in the unit space, in order to describe a rare event. When the risk of flu obeys the Poisson distribution, the mean is equal to the variance used in the interval estimation of the overall mean. In the Poisson distribution, the mean and variance of O_i are the same. But as a result

of excessive distributed data, the variance may be higher than the average, thus the statistical model needs to be extended. In this case, a simple method is to use the negative binomial distribution to replace the Poisson distribution. When r is an integer, the probability mass function of negative binomial distribution, also known as PASCAL distribution, is:

$$f(k; r, p) = \binom{k+r-1}{r-1} \cdot p^r \cdot (1-p)^k$$

For negative binomial distribution the variance is greater than the average. Because the risk is not homogeneous in different degrees, the greater variance compared with the average in negative binomial distribution, the more serious the homogeneity among regions [4]. Considering the random effect followed Gamma distribution in various areas, we can combine them to use the Poisson-Gamma (PG) model, which can be formed by the following two distributions:

$$O_i | \theta_i, E_i \sim Po(\theta_i E_i)$$

$$\theta_i \sim Ga(v, \alpha)$$

The relative risk θ_i is a random variable, while the distribution of O_i is conditional on the level of θ_i , which is extracted from the Gamma distribution with the average v/α and variance v/α^2 . A shrinkage estimation is calculated by the following formula:

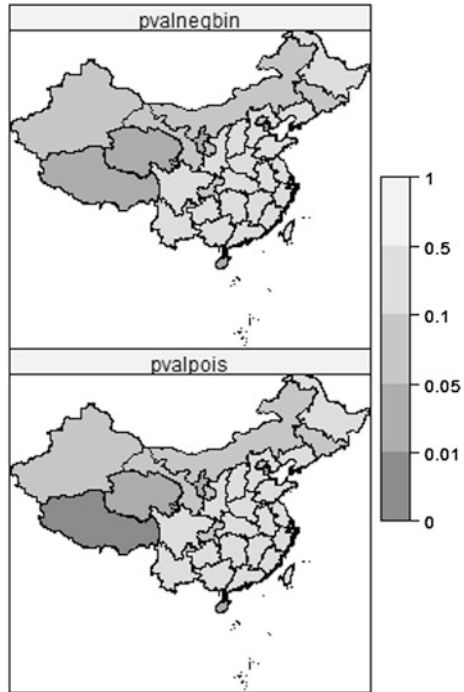
$$E[\theta_i | O_i, E_i] = \frac{E_i}{\alpha + E_i} SMR_i + \left(1 - \frac{E_i}{\alpha + E_i}\right) \frac{v}{\alpha}$$

In this paper, E_i in the areas with low population are usually very small, which can lead to larger changes of SMR_i with very small change of O_i , thus the weight of previous average SMR_i will be small. In addition, if v and α in each region are the same, the data in all regions can be constructed to a posteriori estimator to consider a collection of different areas or neighborhoods [5].

With probability graphics, the p -value of the number of cases for the current used model can be well displayed for observation. Figure 3 is the probability graph of Poisson and Poisson-Gamma models, showing the probability according to the model being higher than from observed data.

The purpose of comparing the above two figures is to show the changes caused by different models. It can be seen that due to the excessive distributed data, the choice of Poisson-Gamma model is better as expected, where p -values are higher for further estimation. However, there are still two areas in the western and northern China with high risk of morbidity.

Fig. 3 Probability graph



3.3 Statistical Model of Log-Normal

Log-normal distribution model, which is frequently used in survival analysis, fits the distribution of survival time of H1N1 cases by completing the maximum likelihood estimate of lognormal model. Given $l = 2d > 0$ as the length confidence interval for estimation parameter, $1 - \alpha (0 < \alpha < 1)$ as the confidence probability, we set $X : \text{LOG}(\phi, \alpha^2)$. Then we study the problem of estimation of μ , based on the relative risk $\beta_i = \log(\theta_i)$ fitting the hypothesis of multivariate normal distribution with mean ϕ and variance α^2 [6]. In this case, the logarithmic relative risk estimator is not denoted by $\log(O_i/E_i)$, but $\log((O_i + 1/2)/E_i)$, because when O_i is 0, the former is meaningless [7].

We use the logarithmic normal distribution to estimate parameter based on EM algorithm, which provides an efficient iterative procedure to calculate the estimation of maximum likelihood of the data. It makes adjustment on missing data by the parameters from previous step when data is incomplete and then updates the estimation value of parameter for the maximum likelihood criterion.

Each iteration is divided into two steps, expectation and maximization, so being called EM algorithm. EM algorithm is used to obtain the estimate values of mean and variance of model, and it repeats the following two steps until convergence [8].

Step 1: estimates the value of E: using the current hypothesis h and observed data X to estimate the probability distribution of Y in order to calculate $Q(h'|h)$.

$$Q(h'|h) \leftarrow E[\ln P(Y|h')|h, X]$$

Step 2: maximize: assuming h is replaced by the hypothesis h' which maximizes the function Q .

$$h \leftarrow \arg \max Q(h'|h)$$

For the Bayes estimator of β_i [6],

$$\hat{\beta}_i = b_i = \frac{\hat{\phi} + (O_i + \frac{1}{2}) \hat{\sigma}^2 \log[(O_i + \frac{1}{2})/E_i] - \hat{\sigma}^2/2}{1 + (O_i + \frac{1}{2}) \hat{\sigma}^2}$$

where $\hat{\phi}$ and $\hat{\sigma}^2$ are priori mean and variance,

$$\hat{\phi} = \frac{1}{n} \sum_{i=1}^n b_i = \bar{b}$$

$$\hat{\sigma}^2 = \frac{1}{n} \left\{ \hat{\sigma}^2 \sum_{i=1}^n \left[1 + \hat{\theta}^2 (O_i + 1/2) \right]^{-1} + \sum_{i=1}^n (b_i - \hat{\phi})^2 \right\}$$

and b_i is updated by the formula until convergence. Therefore, the estimation of θ_i is $\hat{\theta}_i = \exp\{\hat{\beta}_i\}$.

In Fig. 4, EBLN represents the risk estimator of the Log-Normal model, the estimated result is the combination of logarithmic local estimation and ϕ . The estimation values computed by R programming language are shown in Table 3 and comparison graph shown in Fig. 4.

3.4 Statistical Model of EB Estimator of Moments

Marshall (1991) proposed a new empirical Bayes (EB) estimator by adopting the method of moment, which assumes that the relative risk factor θ_i has a common prior mean μ and variance α^2 [5], as shown below:

$$\hat{\theta} = \hat{\mu} + C_i(SMR_i - \hat{\mu}) = (1 - C_i)\hat{\mu} + C_i SMR_i$$

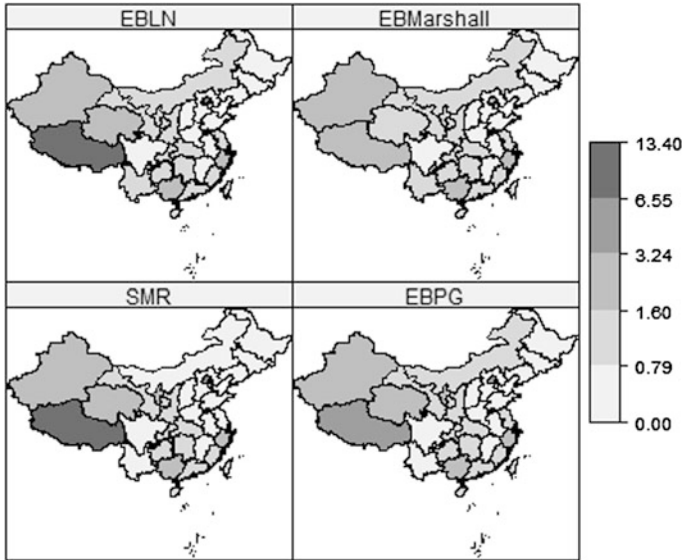


Fig. 4 Comparison of the risks estimated by different models

where

$$\hat{\mu} = \frac{\sum_{i=1}^n O_i}{\sum_{i=1}^n E_i} \quad C_i = \frac{S^2 - \hat{\mu}/\bar{E}}{S^2 - \hat{\mu}/\bar{E} + \hat{\mu}/E_i}$$

in which \bar{E} represents the average value of E_i and S^2 usually represents the unbiased estimation of the variance SMR_i . And when $S^2 < \hat{\mu}/\bar{E}$, $\hat{\theta}_i = \hat{\mu}$. The EB estimator will produce negative estimates of relative risk. The contraction of the estimates largely depends on the value of E_i . High value of E_i suggests that SMR_i is a reliable estimation, C_i is close to 1 and SMR_i is given higher weight; otherwise, a priori estimate of $\hat{\mu}$ will be accounted for higher weight, because the credibility of SMR_i is very low and need to be reestimated [7].

Table 3 shows the estimated values of each model calculated by R programming language. Figure 4 shows the comparison of the risks estimated by different models, where SMR represents standard mortality, EBPG, EBLN and EBMarshall respectively represent the models of Poisson-Gamma, log-normal and MarshallEBPG.

Judging by Table 3 and Fig. 4, similar estimation is obtained by each model. By comparing the SMR maps generated by various methods, we can see extremums (max or min) are shifted to global average to a great extent. In order to compare the variability of the estimates produced by various methods, we draw boxplots of the results, as shown in Fig. 5, where the estimators of SMR, EBPG, EBLN and EBMarshall are compared. EBMarshall also shows the trend toward global average, however unobvious, because only part of the information is applied. It clearly

Table 3 The estimated values of each model calculated by R programming language

Region	Observation	Population	Expectation	SMR	EBPG	EBLN	EBMarshall
Beijing	0.287725	1977.5	0.06098727	4.71778797	3.757501409	3.899449584	2.998605442
Tianjin	0.03035	1323.75	0.040825233	0.743412773	0.901985273	0.91267613	0.922295078
Hebei	0.10315	7189.25	0.221720725	0.46522489	0.532389973	0.590427012	0.581246978
Liaoning	0.05665	4372	0.134835068	0.420142927	0.530848545	0.618649109	0.602438569
...
Tibet	0.1242	301.75	0.009306149	13.34601514	4.7133693	7.098812423	2.863520149
Ningxia	0.044625	636	0.019614616	2.275089099	1.690041185	1.60645435	1.381860347
Xinjiang	0.1426	2196.5	0.06774136	2.105065502	1.873859044	1.807058858	1.640332719

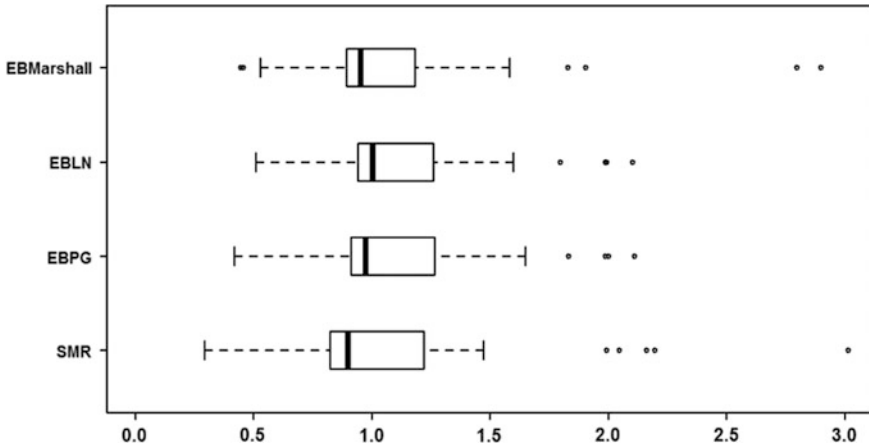


Fig. 5 Comparison of the estimation of SMR and EB

shows that SMR has the highest variability, while the other three all converged to global average, approximately 1. Therefore, EB estimator based on log-normal model seems to be the most stable and reliable method.

4 Conclusions and Future Work

In this work, by the analysis on morbidity data of influenza A (H1N1) across China’s administrative regions from 2009 to 2012, a comparative study was carried out among four different estimators SMR, EBPg, EBLN and EBMarsHall as risk model to explore and make improvements for the problems of risk model and pattern of survival distribution in spacial disease analysis. We compared the variability of different method through SMR maps. Similar estimation is obtained by each model and extremums (max or min) are largely transferred to global average. EBMarsHall also shows the trend toward global average, however unobvious, because only part of the information is applied. It clearly shows that SMR has the highest variability, while the other three all converged to global average, approximately 1. Therefore, EB estimator based on log-normal model seems to be the most stable and reliable method. Because the risk for these models is being smoothing estimated globally, we must consider the size and range of data, thus to take adjacent areas into account when estimating the risk seems more reasonable. We should also consider the need of different scientific fields for the choice of model.

The future research work is to inspect spatial autocorrelation and aggregation, to study hierarchical Bayesian model and CAR model for the application of spacial structure, and by introducing the approach of evolutionary computation to achieve more efficient noise smoothing and deeper research on noise detection.

The study on spacial variability of disease morbidity is helpful in detecting epidemic area and forewarning the pathophoresis of prospective epidemic disease, with great theoretical and practical significance.

Acknowledgments (1) Funding Project of Science and Technology Research and Development in Hebei North University (Grant No. ZD201301). (2) Major Scientific Research Projects in Higher School in Hebei Province (Grant No. ZD20131085).

Bibliography

1. Ying Q, Chen K (2012) Progress of spatial analysis techniques in tuberculosis research. *Dis Surveill* 27(4):330–334
2. Peng B, Zhang Y, Hu D, Luo K, Wang R (2007) Use of space Analysis technology to explore the spatial patterns of TB. *Chin Health Stat* 24(3):229–231
3. Li Z, Zhao W, Xie X (2013) PoissonLog_normal regression model evaluation. *Kunming University (Nat Sci Ed)* 38(4):102–108
4. Zhang G, Liu C, Ma X (2006) Bayes sequential estimation of lognormal population distribution parameters. *Stat Decis* 11:7–8
5. Bivand RS, Pebesma EJ, Gómez-Rubio V *Applied spatial data analysis with R*. ISBN: 978-1-4614-7617-7
6. Lv W, Wu Y, Ma H (2007) Lognormal distribution parameter estimation based on the EM algorithm. *Stat Decis* 21–23
7. Shi Y, Yang Z (1995) Linear regression coefficient EB consensus estimate of convergence rate. *Pure Appl Math* 11(2):15–20
8. Wang L (1999) An approximate method for three-parameter distribution parameter estimation log_normal. *Stat Principle* 18(2):40–43
9. Xie J (2008) RANDOM TESTING Poisson-gamma model coefficients based on longitudinal data. *Series A Coll Univ Appl Math Newspaper* 181–187
10. Li J (2006) The concept of spatial scales and Logic. *Remote Sens* 76–77
11. Chen H, Fang Y (2010) Network Traffic Gaussian mixture model-based clustering analysis. *East Chin Unive Technol (Nat Sci)* 255–260
12. Wang B, Wei Y, Sun C (2008) Poisson distribution and negative binomial distribution in the risk management. *Tianshui Normal Univ Rep* 28(5):23–24
13. Song H (2013) Management mechanism cloud user-oriented service requirements. *University of Science and Technology of China*
14. Shi H, Ji Y (2013) Multiple normal distribution parameter under semi-order restriction Bayes estimation and equivalence test. *Jilin Univ Newspapers (Sci Ed)* 51(1):1–8
15. Cuesta H (U.S) *Practical data analysis*
16. You Y (2013) EB estimates and convergence rate failure rate of exponential distribution. *Luoyang Normal University Rep* 32(5):9–12
17. Zhang M, Yue L (2014) Hybrid system reliability simulation Monte Kano and EB estimation. *Shenyang Univ Technol* 33(3):26–31
18. Zhao Z (2014) Estimation based on a small domain data binomial EB. *Guizhou Sci* 30–33
19. Chen F, Yang S (1999) Corresponding analysis and its application in a variety of diseases clustering analysis. *Chin Health Stat* 16(2):26–31
20. Yin F, Li X, Feng Z, Ma J (2009) Web-based reporting system and temporal clustering of infections detected simulated real-time monitoring and early warning. *Mod Prev Med* 36(12):2204–2207

21. Yang W, Li Z, Lan Y, Wang J, Ma J, Jin L, Sun Q, Lv W, Lai S, Laio Y, Hu W (2011) Chinese outbreak automatic detection and rapid response system is based on Internet. *Monit Syst* 67–71
22. Jiang W, Shen X, Zong F (2014) Multivariate normal spatial scan statistics model in detecting the strongest aggregation of endemic disease. *Shanghai Univ Newspaper (Nat Sci)* 20 (3):274–280
23. Feng J, Wu X, Li S, Zhou X (2011) Statistical analysis of spatial and related software applications in infectious disease research. *J Chin Schistosome Dis Prev* 23(2):217–220
24. Liao Y, Wang J, Yang W, Li Z, Jin L, Lai S, Zheng X (2012) Infectious disease detection method of multi-dimensional clustering. *Geogr Sci* 67(4):135–443
25. Qi X, Zhou Y, Hu Y, Wang L, Ge H, Zhuang D, Yang GH (2010) Apply GIS to detect spatial clustering of gastrointestinal cancer mortality. *Geography* 29(1):181–187

An Algorithm for Image Denoising Based on Adaptive Total Variation

Guo Xiaoling, Yang Jie and Zhang Xiao

Abstract Although the traditional TV (Total Variation) model owns excellent image denoising ability, there are staircase effect problems for TV model. In this article, two detection operators for staircase effect problem are proposed. The staircase effect problem can be solved effectively by introducing two operators into traditional TV model. On the basis, it proposes an adaptive total variation model for image denoising. When dealing with image edge, it can still use the traditional TV model. Its purpose is to maintain the advantages in edge protection for TV model. When it is in the smooth area of image, linear diffusion is used to avoid the staircase effect.

Keywords Image denoising · Staircase effect · Detection operator · Total variation

1 Introduction

In the field of image processing, image denoising technology has been the focus of the study. In recent years, the image denoising methods based on partial differential equation had made great breakthrough. In 1990, Perona and Malik proposed the image denoising method based on anisotropic diffusion, called PM model [1]. The model used Laplace operator instead of the traditional nonlinear operator. Though it had achieved good denoising effect, there were serious step effect problems. In order to remedy the defects of the PM model, You and Kaveh proposed a four order partial differential equation, named Y-K model. But the Y-K model had brought the new “dot effect” problems [2]. And the partial differential equation of higher order increases the computational complexity greatly. In 1992, Rudin, Osher and Fatemi proposed the image denoising methods of total variation, called TV model [3].

G. Xiaoling (✉) · Y. Jie · Z. Xiao
School of Information Science and Engineering, Hebei North University,
Zhangjiakou 075000, Hebei, China
e-mail: 175666832@qq.com

The TV model was a functional minimization problem substantially. It can control the image to diffusion in the gradient direction orthogonal. The TV model made a qualitative leap compared with the PM model and the Y-K model. Although the TV model showed strong advantage in the image denoising, it also appeared serious staircase effect problems in the smooth area.

2 Traditional TV Model

Traditional TV(Total Variation) model was proposed by Rudin, Osher and Fatemi. It was also known as the ROF model. And it was a image denoising model based on partial differential equation [4]. The equation is as follows:

$$E_{TV} = \int_{\Omega} \left(|\nabla u| + \frac{1}{2} \lambda (u - u_0)^2 \right) dx dy \quad (1)$$

This equation can be further expanded into the form of the following:

$$\frac{\partial u}{\partial t} = \frac{1}{\sqrt{(u_x)^2 + (u_y)^2}} \frac{u_{xx}(u_y)^2 - 2u_x u_y u_{xy} + u_{yy}(u_x)^2}{(u_x)^2 + (u_y)^2} - \lambda(u - u_0) \quad (2)$$

$\frac{u_{xx}u_y^2 - 2u_x u_y u_{xy} + u_{yy}u_x^2}{u_x^2 + u_y^2}$ is the second order derivative along the tangent direction of the isophotes of image. $u(x, y)$ and $1/\sqrt{u_x^2 + u_y^2}$ is the diffusion coefficient. It can be found by the above formula that the TV model can only diffuse in the gradient orthogonal direction. It will cause the serious staircase effect in image smooth regions inevitably.

3 ATV (Adaptive Total Variation) Model

In order to overcome the staircase effect problem of the traditional TV model, two detection operator which is created from the steerable filter are proposed.

3.1 Steerable Filter

Oriented filters are useful in many early vision and image processing tasks, such as texture analysis, edge detection, image data compression, motion analysis, and image enhancement [5–8]. Here the steerable filters are designed in quartering pairs to allow adaptive control over phase as well as orientation. There are four

applications below: Orientation and phase analysis, angularly adaptive filtering, edge detection and shape from shading [9]. Edge detection is the key factor for image restoration. Here, the steerable energy $E(\theta)$ can detect edges effectively by using the n th derivative of a Gaussian and its Hilbert transform [10].

$$E_n(\theta) = [G_n^{\theta}]^2 + [H_n^{\theta}]^2 \tag{3}$$

In this article, $n = 2$, $0^\circ \leq \theta \leq 360^\circ$ and G is a Gaussian function.

Thus, a steerable quartering pair based on the frequency response of the second derivative of the Gaussian G_2 and its Hilbert H_2 is designed as the following functions:

$$\begin{aligned} G_2^\theta(x, y) &= k_1(\theta) * G_2^0(x, y) + k_2(\theta) * G_2^{\frac{\pi}{3}}(x, y) + k_3(\theta) * G_2^{\frac{2\pi}{3}}(x, y) \\ H_2^\theta(x, y) &= l_1(\theta) * H_2^0(x, y) + l_2(\theta) * H_2^{\frac{\pi}{3}}(x, y) + l_3(\theta) * H_2^{\frac{2\pi}{3}}(x, y) + l_4(\theta) * H_2^{\frac{3\pi}{4}}(x, y) \end{aligned} \tag{4}$$

where $k_j(\theta)$, $l_j(\theta)$ are the interpolation functions of G_2 and H_2 , they are

$$\begin{aligned} k_j(\theta) &= \frac{1}{3} * [1 + 2 * \cos(2 * (\theta - \theta_j))] \\ l_j(\theta) &= \frac{1}{4} * [2 * \cos(2 * (\theta - \theta_j)) + 2 * \cos(3 * (\theta - \theta_j))] \end{aligned} \tag{5}$$

Image information features of each direction can be extracted based on principle steerable filter. Figure 1 is the result that the simple geometry are dealed after fixed direction filtering and steerable filtering. Among them, (a) is as the original image, (b) is for Canny operator, (c) is for 0° fixed direction filtering, (d) is for 90° fixed direction filtering, and the final picture (e) is filtered through a steerable filter. Although figure (b) after Canny operator treatment can identify image edge information greatly after, the square has appeared double edge. It makes a big deviation in the details. Obviously, the details and features will always be lost in a certain single direction, which is not beneficial to the analysis of the image, such as Figure (c) and (d). In contrast, (e) has almost no loss of detail after the steerable filter is. It shows all the edge features of the original image excellently. It can be seen from this experiment that the steerable filter has great advantages in edge detection.



Fig. 1 Comparison chart for Different algorithms. **a** Original image, **b** Canny operator, **c** 0° fixed direction filtering, **d** 90° fixed direction filtering, **e** steerable filter

3.2 Detection Operator

The spectral power $E(\theta)$ of the steerable filter is the orientation strength along a particular direction by the squared output of a quartering pair of band pass filters steered. The responses of the same pixel are different by different phases. Here the max value by the phase is needed which is the main direction. As the maximum, where

$$m_{(x,y)} = \max(E_{(x,y)}(\theta)), \quad 0^\circ \leq \theta \leq 360^\circ \quad (6)$$

Based on the new indicator m from the steerable filter, we present a natural way to improve the TV model. The STV model is presented as follows:

$$AE_{TV} = \int_{\Omega} \left(|\nabla u|^{q(\bar{m})} + \frac{1}{2} \lambda(\bar{m})(u - u_0)^2 \right) dx dy \quad (7)$$

where the functions $q(\bar{m})$ and $\lambda(\bar{m})$ are as follows [11]:

$$q(\bar{m}) = 1 + \sqrt{\bar{m}}, \quad \lambda(\bar{m}) = k * \sqrt{\bar{m}}, \quad \bar{m} \in (0, 1], \quad k > 0 \quad (8)$$

In this way, the TV model can achieve adaptive by changing their coefficient. And when $\bar{m} \in (0, 1]$, $q(\bar{m}) \in [1, 2)$, $\lambda(\bar{m}) \in (0, k]$ when $q(\bar{m}) \rightarrow 1$, $\lambda(\bar{m}) \rightarrow k$. This model is close to TV model; when $q(\bar{m}) \rightarrow 2$, $\lambda(\bar{m}) \rightarrow 0$. At this time, the model is close to the least squares method.

Thus, it can achieve good denoising ability and edge preserving by using the TV model in image edge region. And it can effectively solves the problem of the step effect by using the least squares method in the smooth region.

4 Comparative Analysis and Experimental Results

There are a large number of “false edge” in the ramp images. This kind of image will have a relatively serious staircase effect in the process of denoising [12]. Therefore, the ramp image is a typical example for staircase effect testing. Here select a typical ramp images as the original image. Figure 2 is a comparison diagram of using two algorithms for image denoising.

The figure (a) is the original image, figure (b) is Noisy image after adding Gauss white noise, figure (c) is the denoised image for TV model diagram. It is easy to see that the image has been very clear. But there are obvious step effect in the image. figure (d) is the denoised image for adaptive TV model. It not only avoid the staircase effect partly, but also has strong denoising ability. It is one of the most ideal method.

PSNR (Peak Signal to Noise Ratio, PSNR) is the most popular objective evaluation method. It is widely used for measuring image quality. The bigger of the

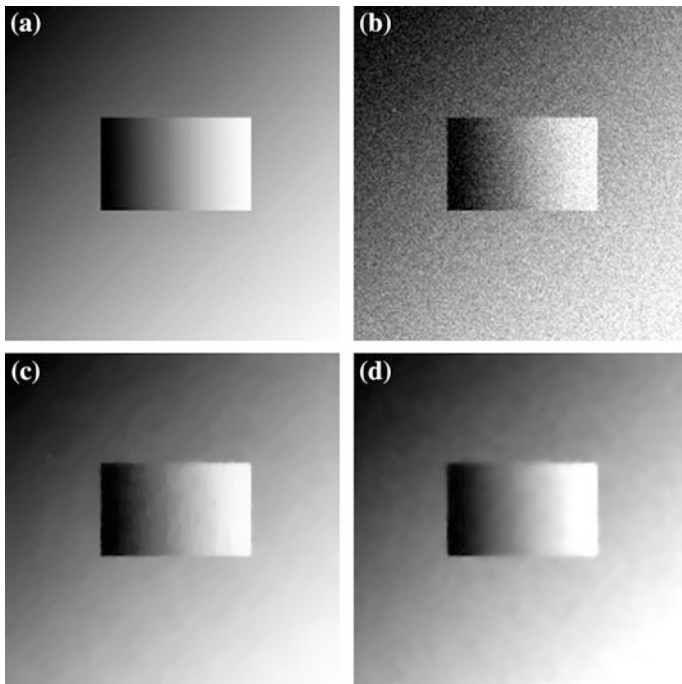


Fig. 2 The comparison diagram of using two algorithms for image denoising. **a** Original image, **b** Noisy image ($\sigma = 20$), **c** TV model, **d** ATV model

PSNR value, the better of the image quality, it mean less distortion. MSSIM (mean structure similarity, MSSIM) is also an evaluation method of image quality. It can evaluate the content similarity degree of two images. Table 1 gives the MSSIM and PSNR value for Fig. 1. It is not difficult to find that the MSSIM and PSNR for ATV model are higher than the traditional TV model.

The noise has little effect to image when the Gauss white noise $\sigma = 20$. Table 2 gives a comparison of MSSIM and PSNR for TV model and ATV when the image

Table 1 MSSIM and PSNR for TV model and ATV model

Model	MSSIM	PSNR	Iterations
TV	0.901	29.001	55
ATV	0.915	29.430	80

Table 2 MSSIM and PSNR of different noise value for TV model and ATV model

Noise	Model	MSSIM	PSNR	Iterations
$\sigma = 20$	TV	0.901	29.001	55
	ATV	0.915	29.430	80
$\sigma = 30$	TV	0.893	29.123	85
	ATV	0.900	29.334	100
$\sigma = 40$	TV	0.898	28.968	90
	ATV	0.899	28.812	100

is in different noise environment. Experiments show that, although in strong noise environment, the adaptive total variation model not only has great image denoising ability, but also deal step effect excellently.

5 Conclusion

The results show that the proposed new method achieves a great balance in denoising ability and avoiding the staircase effect after a number of experiments. However, it has some limitations in selecting parameters for some partial differential equations in this paper. Therefore, it is very necessary to conduct more experiments and select more accurate parameter values in the follow-up study.

References

1. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. *Pattern Anal Mach Intell* 12(7):629–639
2. You Y, Kaveh M (2000) Fourth order partial differential equations for noise removal. *Image Process* 9(10):1723–1730
3. Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noiseremoval algorithms. *Physica D* 60:259–268
4. Xu J (2006) Iterative regularization and nonlinear inverse scale space methods in image restoration. University of California, Los Angeles
5. Kass M, Witkin A (1987) Analyzing oriented patterns. *Comp Vision Graph Image Process* 37:362–385
6. Knutsson H, Granlund GH (1983) Texture analysis using two-dimensional quadrature filters. In: IEEE computer society workshop on computer architecture for pattern analysis and image database management, pp 206–213
7. Knutsson H, Wilson R, Granlund GH (1983) Anisotropic nonstationary image estimation and its applications: Part 1—restoration of noisy images. *IEEE Trans Commun* 31(3):388–397
8. Zucker SW (1985) Early orientation selection: tangent fields and the dimensionality of their support. *Comp Vision Graph Images Process* 32:74–103
9. Freeman WT, Adelson EH (1991) The design and use of steerable filters. *IEEE Trans Pattern Anal Mach Intell* 13(9):891–906
10. Yokono JJ, Poggio T (2004) Oriented filters for object recognition: an empirical study. In: IEEE international conference on automatic face and gesture recognition processing, pp 755–760
11. Guo X, Wu R (2014) Application research on adaptive total variation image denoising. *J Hebei North Univ (Nat Sci Ed)* 30(5):21–24
12. Chan T, Esedpglu S, Park F et al (2005) Recent developments in total variation image restoration. Technique report, UCLA

Social Events Detection and Tracking Based on Microblog

Guiliang Feng, Yiping Lu, Jing Qin and Xiao Zhang

Abstract With the popularity of microblog, more and more people like to use microblog to speak, to communicate feelings, sharing anecdotes, microblog has increasingly become the important platform for people to share information. Social events, which are concerned by a considerable amount of people in the society, will inevitably be reflected in the microblog data, and this often has the characteristics of timeliness, high speed, thus it become a good place to start for the social event detection and analysis. Social event tracking and detection is important for improving the level of social governance, improving the ability of the enterprise brand management and enhancing the level of anticipation and intervention of the problem. The study is based on SAS text analysis and natural language processing technology, through many experiments, finally found the effective ways of event detection and tracking. Combining event mining of main body, auxiliary information exhibition and trend tracking, this subject constructs a preliminary prototype system.

Keywords Microblog · SAS · Data mining

G. Feng · Y. Lu · X. Zhang (✉)
Hebei North University, Zhangjiakou, Hebei, China
e-mail: 780117251@qq.com

G. Feng
e-mail: 6838710@qq.com

Y. Lu
e-mail: 475494052@qq.com

J. Qin
HeBei University of Architecture, Zhangjiakou, Hebei, China
e-mail: 21475243@qq.com

1 Background

This study mainly solves the serious problem of data noise. For example, analysis the difference of word frequency, high frequency words often represent a higher visibility. However, after sorting the every day's high-frequency words, we found that the noise is very serious. Take the data of September 25, 2014 as an example, the Top 30 words are as shown in Table 1.

From Table 1 we can see that Top 30 vocabularies are short, commonly used words, it can not bring truly valuable information. The second problem is that the amount of vocabulary is huge, the data of total include not repeated words on September 25, 2014 are above 245,700, which makes the method based on filter (for example, according to the word frequency filtering) not applicable, because every word frequency section contains a lot of vocabularies, contains a lot of noise. Further, after decomposing the event factors, we can believe event consists of subject and its affiliated information, ancillary information including time, place, cause, process, result and influence and so on. However, the subject is the event stakeholders, and starting from the main body, by retrieving can ultimately find other elements. Subject is given priority to with noun type words and events, including name, organization, product, etc. Therefore, the following analysis is on the basis of nouns type words. As shown in Table 2 is September 25, 2014, the Top 30 noun words.

We note that, according to the word order of noun, the noise is still very serious. And the number of nouns is still not small, on September 25, the none-repeated noun words is total 19,234, it is still difficult to effectively handle the problem based on the filtering method. The root of the problem is, many nouns is daily used, and is not associated with certain events.

Then, if we limit the content of the noun in the entity, the specific things that exist in the real world, whether can solve this problem? Many words in the table above, such as "people", "love" and "will", have not any corresponding objects in

Table 1 Top 30 Chinese words

Rank	Words	Frequency	Rank	Words	Frequency	Rank	Words	Frequency
1	的	560,296	11	人	83,808	21	去	49,770
2	我	327,444	12	好	74,402	22	很	49,328
3	了	299,614	13	就	70,280	23	会	48,973
4	一	186,693	14	都	66,915	24	吧	46,948
5	你	171,561	15	说	57,750	25	和	44,289
6	是	161,631	16	要	57,099	26	给	43,776
7	不	154,839	17	啊	57,028	27	微	43,443
8	个	102,308	18	这	54,626	28	来	41,963
9	在	100,956	19	想	52,122	29	爰	41,746
10	有	89,258	20	也	50,457	30	到	41,446

Table 2 Top 30 Chinese noun words

Rank	Words	Frequency	Rank	Words	Frequency	Rank	Words	Frequency
1	人	44,321	11	时候	12,511	21	城市	9531
2	我的	31,161	12	游戏	12,076	22	自己的	9436
3	泪	28,986	13	时	11,490	23	月	9311
4	会	20,396	14	图片	11,431	24	选项	9251
5	你的	18,449	15	天	10,933	25	家	9089
6	心	18,407	16	网	10,725	26	人生	8988
7	今天	14,018	17	勋章	10,719	27	阿	8503
8	爱	13,481	18	时间	10,287	28	年	8387
9	抓狂	13,173	19	一个人	9994	29	手机	8378
10	一下	13,100	20	活动	9607	30	点	8202

Table 3 Top 30 Chinese entity words

Rank	Words	Frequency	Rank	Words	Frequency	Rank	Words	Frequency
1	哈哈	13,771	11	新浪	7026	21	视频	3467
2	中国	13,345	12	鞋	6137	22	水果	3305
3	门	13,325	13	苹果	5985	23	同学	3215
4	包	12,004	14	一点	5354	24	德克萨斯	3178
5	一天	9438	15	茶	4969	25	9月	2916
6	妈妈	8898	16	爸爸	4528	26	烟	2808
7	酒	8790	17	老师	4200	27	美国	2770
8	手机	8596	18	儿子	3926	28	上海	2761
9	三国	8498	19	测试	3506	29	奥迪	2667
10	音乐	8143	20	北京	3496	30	十年	2437

the real world, While entities including person names, place names, organization names, name or title, etc. Table 3 is the Top 30 entities on September 25.

We can get some clues from the study above; however, data quality is low. What are the reasons? We found that many entities are still often used in our daily life, the randomness of their presence is obvious. For the further consideration, the root of the problem is “random”, what we are looking for is nonrandom and with burst characteristics entities.

2 Program

2.1 Based on the Frequency or Rankings Change Detection Program

How to automatically filter out random, daily entities, and then detect the random and sudden situation? The answer is to put the time into consideration. Assuming

Table 4 Top 7 Chinese words (ranking from the changes)

Rank	Words	Rank	Words	Rank	Words
1	仁川	1	仁川	1	仁川
2	25日	2	25日	2	25日
3	亚运会	3	亚运会	3	亚运会
4	西藏	4	西藏	4	西藏
5	长寿岛	5	长寿岛	5	长寿岛
6	叙利亚	6	叙利亚	6	叙利亚
7	星期四	7	星期四	7	星期四

the random and daily entities within each day (or other time period) is similar to the emergence of the rules, if the information in the next period of time minus the last time period of information, it should be able to effectively remove the noise. So really meaningful index is no longer a frequency, but the change of frequency; If the frequency changes are volatile, then we can consider the change of the frequency ratio, the change of the “sort”. Relevant indicators calculation is as follows:

$$\begin{aligned}
 Freq_Offset_d &= Freq_d - Freq_{d-1} \\
 Freq_Offset_Ratio_d &= \frac{Freq_Offset_d}{Freq_{d-1}} \\
 Rank_Offset_d &= Rank_{d-1} - Rank_d
 \end{aligned}$$

$Freq_d$ represents on the events frequency of the day, $Freq_{d-1}$ is on behalf of the previous day’s events frequency, $Rank_d$ and $Rank_{d-1}$ are respectively represent the rank in the day and the day before.

Table 4 shows the Top 7 vocabulary in frequency variation, rate of frequency variation, ranking changes on September 25.

It can be seen that the results have been improved greatly, events which has been found include “incheon Asian games”, “longevous and longevity island” and so on.

Table 5 shows the Top 15 vocabulary in frequency variation, rate of frequency variation, ranking changes on September 25.

2.2 Joint Analysis Based on Clustering and Denoising

As mentioned earlier, a very important challenge in this study is abundant noise; Many microblogs are not talking about the events which we are interested in, but in the release of some daily information. So is there a way automate to filter out the noise? Can noise filtering and data mining analysis be simultaneously? This is another topic we are interested in. Although removing noise and data analysis at the same time sounds run counter to the usual analysis process, usually the data is completed before to analysis module is analyzed. However, data reduction often requires a lot of manpower, for example to analyze what characteristics do the noise

Table 5 Top 15 Chinese noun words(ranking according changes)

Rank	Words	Rank	Words	Rank	Words
1	西藏	1	仁川	1	仁川
2	阳澄湖	2	亚运会	2	亚运会
3	俞灏明	3	俞灏明	3	长寿岛
4	马英九	4	马英九	4	男足
5	叙利亚	5	清华	5	清华
6	3D	6	3D	6	3D
7	仁川	7	西藏	7	西藏
8	亚运会	8	副市长	8	副市长
9	男足	9	男足	9	好声音
10	风电	10	风电	10	风电
11	军校	11	军校	11	俞灏明
12	长寿岛	12	长寿岛	12	马英九
13	中校	13	博士	13	博士
14	清华	14	叙利亚	14	克隆
15	博士	15	丁玲	15	丁玲

data has, due to big noise data volume, even occupied more than half of the total data, therefore, filter cost of it is too high, this mode has become difficult to be achieved.

We try to combine clustering and denoising, the method is as follows: first of all, all data clustering, there must be a number of clustering of all are belongs to the obvious noise category; Then we will filters out the document of these categories, the remaining are some data with slightly higher quality, then iterate the process. As you progress through the iteration, the moisture in the data is “crowd out” bit by bit, the rest part is valuable and this time it will improve the quality of clustering, then it will realize the process of the event subject excavation.

As shown in Table 1 is iterative analysis flow figure. Among them “the first iteration”, “the second iteration” and “the third iteration” are SAS code node, their functions are filtering out some categories of noise in all documents from the previous iteration output, generate new data sets. The SAS code is as follows (Fig. 1).

```
libname mylib "C:\microblog\Testing Microblog\Workspaces
\EMWS1";
proc sql;
create table mylib.cluster_filter as
select t1.var1, t1.TextCluster4_cluster_
from mylib.textcluster4_train t1
where t1.TextCluster4_cluster_ not = 1 and t1.var1 not = "
order by TextCluster4_cluster_;
quit;
```

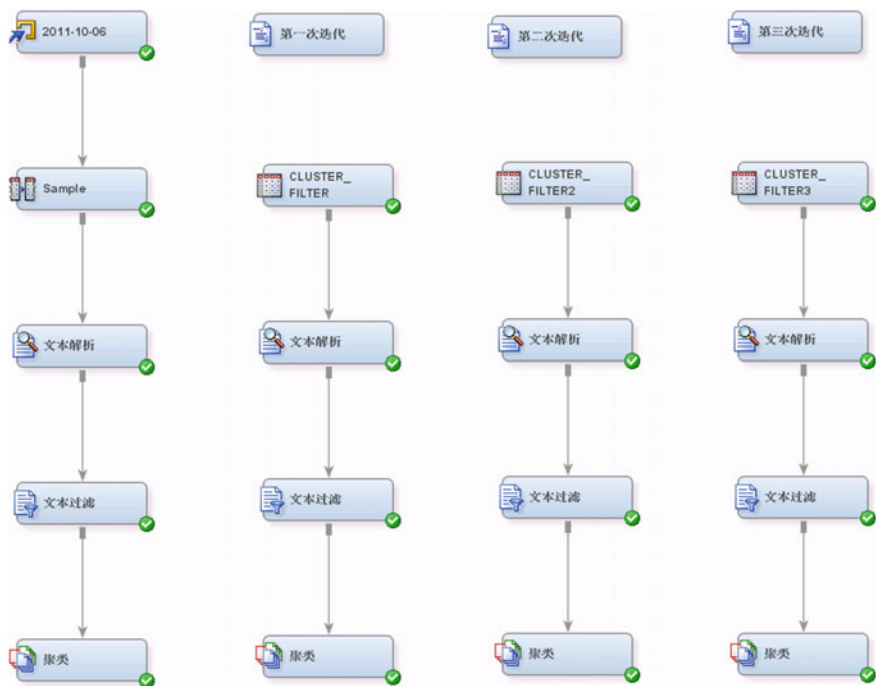



Fig. 1 Iterative analysis flow figure

In each round of iteration, “text analysis”, “text filtering” and “cluster” node configuration and the initial iteration are exactly the same configuration.

“Text analysis” is used to do the natural language processing to the input text, including word segmentation, part-of-speech tagging, entity extraction, etc., thus transform the unstructured data into structured data. The output word is a document frequency matrix. The representation is a typical sparse matrix storage method.

“Text filtering” node configuration mainly completes two tasks, namely the calculation of statistical indicators and filtering the vocabulary in statistical significance, After this step, sparse matrix from simple document—word frequency matrix is transformed into matrix with statistical indicators.

In both text parsing nodes output and the output of text filtering, the size of the matrix is very large, the line number is the number for the collection of words in the document, the column number is the total number of documents. Direct operation of the matrix will have two problems, one is a very large amount of calculation, the second is the moderating effect of the noise information. So before clustering, usually use dimension reduction to the matrix. Text Miner use eigenvalue decomposition (SVD) as a method for dimension reduction, its principle can be seen from citation [1]. The Text Miner supports two different types of clustering algorithm, which means the maximum expected algorithm and hierarchical clustering algorithm. Because this study wants to adopt “partition type” clustering

聚类ID	描述词	频数	百分比
1	版主 边堆地区人民 大女星 大一分 地歌少女 地歌少女 电脑今天 恶梦醒醒醒醒 富有才情 高明架 更多希望 关爱儿童 群伦理大...	104362	60%
2	博睿 茶 存精 耶手 哈哈 耶原 好礼 红樱伊 耶电 金冠店百强 精英 快捷入口 鹿康 鹿康记 美国 牛仔 女装 神掌 ...	9492	5%
3	安因 表妹 博睿 大圣诞 德国 法国 夫人 福州 姑娘 姑娘 很多东西 很多朋友 活力宝贝 金刚 精彩内幕 老干 竞猜 美女老师 命运交...	4103	2%
4	ceo 笔记本电脑 聪明才 二部资源 辉煌 加国 雅盘 建讯 南克 鑫源壹肆 牛精 苹果 苹果产品 苹果创始人 苹果定义 苹果发布会 苹...	4080	2%
5	舞式 郑瑞 彩虹杯 单执 赵重机 单崇 高主 电动钢琴 顶级钢琴 佛山 佛山所 个儿半夜 会计 杭州 厂商 杭州工程 快乐岛 拉手阿 栏...	2421	1%
6	阿姨 爸爸 爸爸 翠依林 曹佳 衬衫 次郎 大巴 大哥 大牌 大盘 泰希 队长 儿子 哥哥 国足 教练 杭州 姐姐 老公...	8123	5%
7	包 单原 德官方旗舰 东方神秘 风水起 妙 风水女皇 付出才 够了 博逸赛 醉斯 欢乐高 活动盛世 江北 江西北 酒 酒醉手舞 力度才能 米...	3788	2%
8	百度 北京 别克 别克 车模 成都 大姨姐 榕榕 郭德的 国美 朝国美女 辉煌 基情 昆明 刘诗诗 南京 尼玛 七雄 汽车 火灾又...	6749	4%
9	白羊座 电子产品 凤凰 改变世界 冠军 果粉 很多人 华尔街 季军 今天天气 巨蟹 空调 美国 鹿康 苹果公司 全明星 人都 日本 射...	6323	4%
10	ctv 爸爸 爸爸妈妈 白紫发 伯伯 不断她 彩霞 菜刀都 德克萨斯 德克萨斯扑克 德克萨斯扑克 地鼓力 电脑游戏 董董 饭岛 饭...	3215	2%
11	百分百 品牌 规格 大半夜 大碟 大歌都 大明 大明星 大牌 大哥 代表 芬香 父亲 红酒 价值球员 咖啡 开始游戏 昆仑 连连看 美的...	9452	5%
12	博文 彩虹 长沙 创新 东京 短信 关键词 广州 滚石 韩国 好朋友 很多时候 会长 拉手网 伦敦 美腿 面包 模特 纽约 欧美...	11828	7%

Fig. 2 The results of the initial clustering

聚类ID	描述词	百分比
1	百度 百分百 北京 别克 车模 大明 大明星 电子产品 改变世界 郭德的 辉煌 会长 基情 礼品 面包 七雄 资产 赛车 上海 沈阳 ...	9%
2	城堡 法国 芬香 佛山 红旗 华尔街 金门 今天天气 开始游戏 老师 美国 美国人 美女老师 南海 男明星 美国 苹果公司 苹果公...	7%
3	包 单原 德官方旗舰 短款 多酒 风街女皇 付出才 博逸赛 醉斯 欢乐高 活动盛世 江北 江西北 酒 酒醉手舞 力度才能 基尼黑 ...	6%
4	白紫发 博瑞战 兵器暴君 兵器屠 兵器黄金旗帜 兵器勇士 兵器血球 不断她 菜刀都 火影 大剑 地鼓力 电脑游戏 过...	2%
5	阿姨 爸爸 爸爸妈妈 爸爸 品牌 表妹 彩霞 大半夜 大碟 德克萨斯 德克萨斯扑克 德克萨斯扑克 游戏 儿子 父亲 公公 姑娘 湖...	11%
6	才发 衬衫 成都 大巴 大哥 到时候 福州 哈哈 杭州 和你 很多时候 很多感情 华语 江边 姐姐 金刚 精英 老妈 连连看 莫文...	12%
7	白羊座 饼干 蔡徐林 茶 大的人气 方便面 冠军 戒指 郭虎 武艺 黑眼睛 洪辰 化妆品 季军 金壮 巨蟹 快男 刘忻 绿茶 鹿鹿 鹿鹿...	5%
8	813875941689 比基尼 超短裙 美女性感 舞蹈 陈年旧事 陈奕迅 大爆炸 动漫星球 范冰冰 凤凰 凤凰视频 功能强大 国美 哈利波...	2%
9	琼瑶 彩虹杯 次工资 大結局 岛主 点问题 个儿半夜 古味 会计 空间日志 快乐岛 栏版本自由 力功能 跑酷 雷耀 门 绵羊 宵 热...	4%
10	创新 存精 短信 关键词 海量资源 好礼 好运指数 晋族 晋族劲章 老朋友 领悟 鹿康 鹿康记 美国 免费空间 摩天 盘新浪...	6%
11	博文 彩虹 大家都 大哥 代表 东京 董董 榕榕 姑娘 广州 韩国 好朋友 很多人 拉手网 伦敦 美腿 南京 纽约 苹果旗舰店 妻子 ...	20%
12	大牌 福建 感觉都 耶手 红酒 红樱伊 加国 建筑 静心 乐器名称 乐曲 乐曲记忆 刘德华 逻辑数学 明星终极 三亚 时后 私人照...	5%
13	ceo 笔记本电脑 春水堂 春水堂情趣用品 顶尖人才 个人魅力 诡异图标 辉煌 想潮总流 德盛 街机 可爱架 老哥 一路 品质 品质...	5%
14	博睿 大特盘 德国 队长 凤凰 哥哥 国足 黄河 活动劲章 耶电 耶电黄金国 金冠店百强 精英 快捷入口 猎猫 命运...	7%

Fig. 3 Clustering results after the first iteration

(that is, each document is divided into at most one clustering), so the maximum expected algorithm is accepted.

The iterative clustering results before the first round are shown in the Fig. 2. It is interesting to note that the largest percentage of a category, that number is “1” is noise data clustering, so in the first round of iterations, we will delete all the microblogs.

As shown in Fig. 3 is the result of after the first iteration.

At this point, there are two changes, one is the size of each cluster is more uniform, the other is the number of clustering increased 2, display algorithm can find more topics. At this time we found 4, 5, 14 three clustering are associated with games, advertising or daily life, it’s not what we want, so it can be removed in the second iteration. The second round of iterative clustering results are shown in Fig. 4.

At this point we can still get 14 clustering, although the total number of microblogs has reduced a lot. We will continue to delete some undesired in poly (such as 3, 4, 7, 8, 14, get the results as shown in Fig. 5.

At this time, most clustering gets more meaning.

- Cluster 1: motor or something related
- Cluster 2: sports events

聚类ID	描述词	频数	百分比
1	白羊座 城市 大赛借机车 冠军 季军 加国 台值球员 建筑 巨壁 快男 昆仑 群众 岳阳 面包 魔境 魔境 糯米网 汽车 球员 射手座 ...	2747	5%
2	不少功夫 成都 大萧东 歌手 富桐家 贵阳 红樱罗 黄朝明 欠石 久石让 快乐心声 刘德华 龙梅子 明天才 明星崇拜 男歌手 声音乐 私 ...	1790	3%
3	有锋 大咖那 大牌 大牌 犬等 鱼主 短笛 斯科 舒礼 很多时候 红酒 快乐岛 狂欢本 自由 力劲魔 雅康 雅康记 美国 门 南海 牛拉 ...	7165	13%
4	心越雷 饼干 茶 木江 大巴 大席 弟弟 方便面 感觉 哈哈 哈 江边 静心 咖啡 老陈 乐昌名桥 乐曲 乐曲记忆 林子祥 绿茶 逻辑数学 ...	4187	8%
5	唱碟 彩虹 原依林 代表 冯冰冰 高洁 歌手 梅梅 关键词 海洋 海涵王 华语 剪辑 剪辑软件 李宇春 梁静茹 刘诗诗 莫文蔚 木兰 内衣 ...	3560	6%
6	ceo 妮娜 创新 芬香 改变世界 红楼 棋盘 开始游戏 库克 领袖 牛顿 苹果 苹果手机 苹果创始人 苹果公司 人人都 上博海关 时间浪 ...	6764	12%
7	百分百 才发 长沙 杉都 大明 大明星 罗时隆 梁石 凉纸 和尔 基情 精灵 连连看 刘忻 牛仔 蓉手 全明星 人气王 沈阳 ...	6101	11%
8	大儿 儿聊 古古 黄德 金茂 蓝城 林雨 琼血日记 小金鱼 新年 新年 新市民 叶苏静 探短 聊阳 大爱 小公主 小蘑菇 小鱼 松江 ...	854	2%
9	辣油 包 潮剧 潮人 大捧捧 大神盘 单屏 德雷方 魏航 动漫星 炫 炫款 凤尚女鱼 付出才 国美 哈利波特 海逸赛 游版 美国美女 粉丝 ...	1921	6%
10	白酒 次会 点味 多酒 福建 翻国明星 翻国明星 加水 江北 江北 白酒 酒精手腕 酒都 开心早晨 力度才能 茅台 基尼果 基尼果牌 酒节 观 ...	3222	3%
11	百度 部分地区 测试 常务委员 会 摩根社 社 嘉岭长 高法 个人资料 官员 官员财产 海涵王 熊立大 亨 好多资源 好朋友 辉庭 会长 廖 ...	2155	4%
12	北京 曹植 车根 大连 大连 东莞 工具 站她 广东 广州 郭德的 很多人 金朋 拉手网 伦敦 很冲 南京 纽约 苹果旗舰店 青岛 ...	9068	16%
13	爱情公寓 表弟 长时间 凤姐 贵公司 很简单 顺序 快乐大本营 老泽 雷女 美女老泽 尼玛 大老泽 潘石屹 彭湃 廖亮女 苹果公司 苹果 ...	2042	4%
14	博文 电影源 菲 俄罗斯 法国 夫人 福州 翻国 杭州 华尔街 记录器 记者 短组 今天天气 轮胎 美国 美国人 美颜 男明星 欧美 欧洲 ...	3995	7%

Fig. 4 Clustering results after the second iteration

聚类ID	描述词	频数	百分比
1	1387591689 翻国美女生活 珊瑚 约会软件 大爆料 大赛借机车 倒卖 东北 东北 东北 汽车 动漫星 炫 借机车美女 哈利波特 好朋友 爆料地 机手美女魔头 加国 建筑 江苏 翻国 ...	2155	6%
2	阿联酋 北越河 越在 德雷方 魏航 多酒 芬香 福建 翻国明星 翻国明星 台值球员 江北 江北 白酒 酒精手腕 开始游戏 昆仑 群众 基尼果 基尼果牌 游轮 ...	3449	10%
3	阿星 百首 开过 部分物品 一百位 爱情那 六折那 第一 曹志 过程 港口 俩金屋 五金百首 米米 吹手和 自手 耶 精英手机 曹敬怡 ...	1429	4%
4	4次炒 蛋卷 大虾 苏夏 蛋卷 凤肉粉 广东 广州 国美 翻国美女 翻国明星 翻国明星 翻国明星 翻国明星 翻国明星 翻国明星 翻国明星 翻国明星 翻国明星 ...	2690	9%
5	笔记本电脑 彩虹 翻国 翻国 才 大等 代表 队员 二 等 芬香 改变世界 华山 妮娜 会长 魏盘 库克 嘉露露 炫 炫 潘子 七雄 时间浪 耀 ...	2679	8%
6	ceo 翻国 苹果 不少 苹果 傅小子 原立人 才 个人魅力 郭生 假期软件 原料 蔡志雄 黄露露 牛顿 女儿 苹果 苹果才 苹果产品 苹果台 苹果官网 ...	3085	9%
7	翻国 冠军 红歌 欠石 欠石 快男 咖啡 ...	2629	11%
8	6台 北京 北京本 北京时间 测试 别克 大太阳 凤尚女鱼 服装 付出才 个人资料 郭德的 郭新 廖亮子 活动盛世 游版史 兰州 类型 翻国 内 ...	4357	13%
9	白羊座 百度 蓝星 单屏 成都 东莞 翻国明星 冠军 边山 会员 鱼先耀 季军 巨壁 昆明 拉羊网 龙凤胎 伦敦 岳阳 门 面包 ...	2071	6%
10	辣油 工具 很多资源 菲 ...	3537	11%
11	博文 翻国 电子产品 东方 多季 翻国 ...	3979	12%

Fig. 5 Clustering results after the third iteration

- Cluster 3: shopping
- Cluster 4: entertainment event correlation
- Cluster 5: IT related
- Cluster 6: IT related (apple CEO)
- Cluster 7: entertainment event correlation
- Cluster 8: shopping
- Cluster 9: entertainment event correlation
- Cluster 10: entertainment and IT
- Cluster 11: entertainment and IT

Of course, this process also can continue, such as can continue to delete the clustering associated with shopping.

2.3 The Comparison of These Two Methods

First of all, based on the method of frequency analysis, this product, it can effectively reflect the change of the data, so as to filter out the noise information; Clustering and denoising of conjoint analysis method, mainly for static data analysis, we can find the main content, but does not consider the factor of time.

Secondly, based on the frequency method to analyze the relatively more events, but the method based on clustering, often can find only a small amount of most meaningful event.

Again, the method based on frequency is convenient for us to further analyze the trend of events, but the method based on clustering, since each analysis are independent static data based on a point time, and therefore it cannot be directly used in the analysis of the trend.

Again, the method based on clustering, its advantage is easier for us to find some event correlation among the subjects, such as the relevance among the words “apple”, “jobs” and “CEO”, although this found often require several modified node configuration, and it is difficult to complete automation (for new data). While the method based on frequency cannot be directly found these knowledge.

Finally, the method based on clustering, with high computing complexity and computing time is significantly longer than the frequency method.

The most contents of this report in the rest, including online demo, are all based on the method of frequency analysis; However, we can design the system architecture that we still want to be able to combine the two, this is the important direction of future efforts.

3 The Event Ancillary Information

As mentioned earlier, events in addition to be the principal information, there are a lot of ancillary information, including time, place, cause, process and result from many aspects, such as, influence and the information spreading in multiple microblog event related information. How can you set out from event subject to obtain the ancillary information? This study provides two information presentation—vocabulary cloud and search.

Vocabulary cloud is a certain events include all the words general information in all microblog, according to the different frequency of vocabulary, different sizes will be shown.

Search is to deliver the event subject to SAS search solution, which can display and event subject related microblog information.

4 Trend Analysis

The role of event trend analysis are manifold. First, after an event is detected, the user may want to follow up the progress of the event, the top event propagation may not accord with the time of the incident detection. Second, the trend analysis is helpful to obtain periodic characteristics and life cycle of events. “星期一”, for example, in some date will be matched for the event, from the 20 days trends are shown in Fig. 6, the incident has obvious periodicity, and the periodic

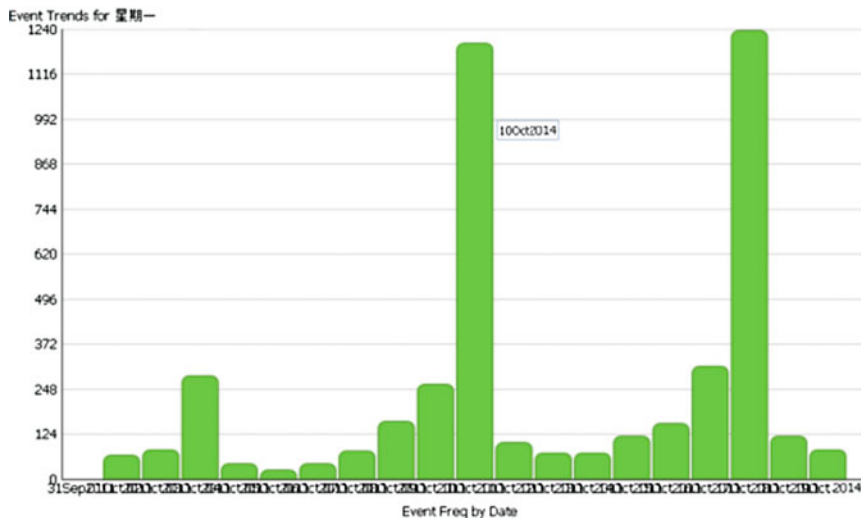


Fig.. 6 20 days trend chart for the event “星期一”

characteristics of the future can be classified as important basis for the event filtering and classification. Trend analysis results show by dhtmlxChart library [2].

5 System Architecture

System architecture, which is composed of four layers: Data Integration, realize the Data Integration, cleaning, and filtering, Data contains a large amount of redundancy, repetition of microblogs accounts for over 40 % of the total microblogs, so the data deduplication is a necessity; Analysis Layer, achieves the extraction of the main event, sorting, recommendations, and establish the index, to prepare for event affiliated information display; Data Layer: the storage events related index of Data and documentation; User Interface: provide event display, query, report forms, etc.

6 Conclusion and Future Work

This study of the work has a certain application prospect in the field of business, including brand management, social event monitoring and treatment (public sector), early warning to a commodity defect, etc., it can help customers insight into trends early, plan ahead in the market competition and corporate governance, etc.

To sum up, the work and contributions of this study are as follows:

Decomposition model, the event is decomposed into the core of the subject and peripheral events affiliated information, the framework allows us to decomposition analysis of events; Presents two mining model of the main events, one is unsupervised, frequency variation analysis model without human intervention, the other is based on a semi-supervised clustering and collaborative denoising model; Two models can be combined in the future in order to obtain better event subject analysis results; Designed and implemented a variety of ancillary information display, including word cloud, retrieval, and so on; Event trend analysis show platform is realized.

In the interest of time, this study has many shortcomings, It can be improved more in the future. Including: strengthen the data cleaning, deduplication and spam filtering; Integration of a semi-supervised and unsupervised analysis results, construct a unified event store; Lexical analysis, the term analysis and entity integrating lexical analysis, and thus screening better event list; Improve the efficiency of event trend calculation.

References

1. Singular Value Decomposition. http://en.wikipedia.org/wiki/Singular_value_decomposition
2. dhtmlxFigure library. <http://dhtmlx.com/docs/products/dhtmlxFigure/index.shtml>

An Optimization of the Delay Scheduling Algorithm for Real-Time Video Stream Processing

Hongbin Yang, Jianhua Guo, Chao Liang, Zhou Lei
and Changsheng Wang

Abstract We used Spark as a platform for large-scale, real-time intelligent video stream analysis. We observed that the default task scheduling algorithm of Spark was not efficient for scheduling image frame data processing tasks, incurring problems such as poor data locality, high network traffic, low utilization of computing resources, etc. This paper investigates why Spark's default task scheduling algorithm is not suitable for real-time video stream processing. Further, we present a new real-time task scheduling algorithm that leverages the notion of data locality. This algorithm schedules tasks based on data locality and information collected at runtime, including task execution time and workload of each node. Experiments show that our proposed algorithm increases data locality and CPU utilization while reducing network traffic and latency.

Keywords Spark · Task scheduling · Data locality

H. Yang (✉) · J. Guo (✉) · Z. Lei (✉)
School of Computer Engineering and Science, Shanghai University,
Shanghai 200072, People's Republic of China
e-mail: hbyoungshu@staff.shu.edu.cn

J. Guo
e-mail: gjhkael@shu.edu.cn

Z. Lei
e-mail: Leiz@shu.edu.cn

C. Liang (✉)
Mobile Business Group, Lenovo Inc, Beijing, China
e-mail: liangchao@Lenove.com

C. Wang (✉)
Information Management Center, Ordnance Engineering College,
Shijiazhuang, China

1 Introduction

Spark is an increasingly popular parallel computing framework developed by the AMP Lab at UC Berkeley [1]. Similar to Hadoop, Spark is based on the Map/Reduce algorithm. However, unlike Hadoop, Spark saves the intermediate output and results of a job in memory, eliminating the need to read and write data to HDFS [2]. Spark is typically better suited for data mining and machine learning applications, which often need to perform MapReduce tasks in multiple iterations. Spark is a distributed memory computing system written in Scala, and is based on an abstract model of RDD (Resilient Distributed Dataset) [1, 3]. Spark uses the master/slave architecture model. The master node is responsible for managing and coordinating the tasks of the entire cluster, and multiple slave nodes are responsible for performing specific tasks.

Spark Streaming [4] is a streaming data processing platform which is built on the Spark platform. Spark Streaming splits the streaming data based on time, and then submits the split data to the underlying Spark platform for batch processing.

The resources in a Spark cluster are allocated based on worker units, and the workers use the executors to perform thread-level tasks. All tasks in Spark will be eventually transformed into two kinds of task, `finalResult` and `shuffle`. These tasks are the smallest scheduling units, and the executor manages a thread pool and is responsible for executing these tasks. The number of tasks each executor executes at the same time determines the degree of parallelism that exists in the system.

At the core of video intelligent analysis is analyzing images frame by frame. During video monitoring, environmental change may cause the same algorithm for different frames to take different amounts of computing time. As an example, consider the OpenCV machine vision library. The face detection and face recognition algorithms [5] take time that varies widely for processing different image frames. The more faces that exist in an image, the longer the algorithm executes.

In a large cluster, task locality can improve throughput. This is because the network bandwidth is typically much lower than disk bandwidth [6, 7]. In intelligent video analysis, each frame of a high definition camera typically has, after decoding, data volume of several megabytes. Each camera produces twenty to thirty frames per second, and there may be multiple transmissions in the network. This will create great pressure to the whole cluster network, which in turn slows down system computation.

This paper is focused on how to speed up video streaming real-time intelligent analysis by leveraging the notion of data locality to reduce network traffic. The default scheduling algorithm in Spark simply allocates the next task to an idle work node for execution. If the data needed for the next task does not reside in the work node, the data needs to be transmitted through the network to the work node. This process can take a lot of IO resources and may have a significant impact on the system throughput.

The default scheduling algorithm of Spark is the delay scheduling algorithm [8] is developed based on the following idea: if a task did not finish within a specified threshold, then the scheduling level is downgraded for the subsequent tasks.

Furthermore, the degradation is not reversible, which causes some tasks that should be executed locally to be placed on other nodes for execution. For example, currently there are ten tasks, for some reason, this ten task level is downgraded, in fact the ten tasks can be completed in locally, but the scheduler think that executed locally could not finish so many tasks, then scheduled five tasks to other nodes task queue for execution.

This paper is aimed to address the above problem. That is, occasionally a long task execution may affect the scheduling of other tasks. The main idea of this paper is collect statistics about the execution time of every task belonging to each worker node within the time window that Spark Streaming uses to convert flow data into the batch before execution. Note that the execution time of a latest task is more important than that of an early task. Thus, we use a weighted approach to calculate the number of tasks that a worker node could complete within the time window. If some assigned tasks have not completed, then the scheduling level is downgraded so that the data can be taken to other nodes for execution.

The remainder of the paper is organized as follows. Section 2 discusses the related work and points out the problems that exist in the default scheduling algorithm. Section 3 presents our scheduling algorithm. Section 4 discusses the experimental results. Section 5 provides the concluding remarks.

2 Related Work

Scheduling for parallel computing in large-scale clusters has been investigated by many researchers [9]. Most of the existing parallel scheduling algorithms were demonstrated in Hadoop [10]. For task scheduling, Spark uses the delay scheduling algorithm which takes into account data locality. In this algorithm, a job is divided into a set of tasks. Each task is associated with the data block that needs to be processed by the task. The tasks are submitted to the task scheduler. Spark uses TaskSetManager to manage the task of each job and control the task's executive level. The executive level of a task determines whether the task is performed locally or transferred to other nodes for execution. Task level conversion is determined by the time interval, which is set statically and cannot be changed at runtime. Spark has very good throughput using delay scheduling, because of it take into the data locality but some mechanisms have problems for example if a task execution time is too long, leading to the next task downgrade, in fact, if the tasks is not to downgrade, these following tasks can also be completed locally within prescribed time. After the job is submitted, TaskSetManager is used to manage tasks for each job. There are four task levels, including PROCESS_LOCAL, NODE_LOCAL, RACK_LOCAL and ANY. These four task levels indicate the next task will be performed locally (i.e., in the same virtual work node), or be transmitted to a virtual work node of the same physical node, or be transmitted to another physical node within the same rack, or be transmitted to another rack, respectively. In our practice we found that the existing scheduling algorithm generates a lot of NODE_LOCAL

tasks. Although the default scheduler can reduce data transfer between hosts, `NODE_LOCAL` tasks may increase data transmission between virtual work nodes when a physical node has several virtual work nodes. To improve throughput, delay scheduling algorithms work as follows. The scheduling mechanism is triggered at the following two points, i.e., when a task is submitted, and when the compute container nodes have completed a task. The scheduling process determines the location of a task execution and uses the most recent delay to determine the scheduling task level. The formula is as follows:

$$\begin{aligned} & \text{If}((\text{Current scheduling time} - \text{The last time the task start time}) \\ & > \text{Window Time} \ \&\& \ \text{There have next task level}) \quad \text{Task downgrade} \end{aligned} \quad (1)$$

If a `PROCESS_LOCAL` task runs for a very long time, it will be converted into a `NODE_LOCAL` task. The reason is that the default scheduling algorithm only checks the current task execution time. If the task's execution time exceeds the prescribed threshold, which has a default value in Spark, then the following task will be downgraded and does not take into account the situation of the previous task execution. Therefore, a task whose execution time fluctuates can cause the next task to be transferred to another node for execution. We found that in practice, `TaskSetManager` will be affected by the weakest computing node since it could not complete the number tasks that expected. If the delay increases it will continue to affect the subsequent arrivals of tasks, causing tasks at the beginning of the scheduling to be downgraded. For example, in face recognition, if the first image has a lot of faces, then this image would take more time than the prescribed threshold, which causes subsequent images to be downgraded even though these images do not even have a face.

In intelligent video analysis, a task can be computationally intensive, and thus can take long to complete. Further, task execution time can vary significantly, e.g., due to different sizes of data blocks. In the process of “convergence import”, based on the location of the receiver, location-aware distribution rules have data scattered on different nodes. In this case, the time delay task scheduling mechanism will break data locality because task execution level downgrade than some data would assigned to other node for computing causing the data block on the same host to be transmitted between computing containers. The delay mechanism of Spark has a bias for `NODE_LOCAL` node locality. This bias is consistent with its simple handling of large data sets to reduce network IO operations. Video analysis is a time-consuming computation. In order to meet real-time requirements of surveillance video streams the computing units should use the frame not the gop which have a lot of frame. Based on this unit building block, in a stream processing time window, it will generate a lot of data blocks within the system. If these data blocks are dealt with by `NODE_LOCAL` level tasks, it will cause the local host to perform many network operations. Each block of data from one container to another container is fetched from computer memory to the network and then the network back to computer memory. Therefore, we need to avoid transmission of data blocks between nodes, encouraging data executed locally.

The default delay scheduling mechanism is based on the idea that if the tasks delay scheduling execution time is reached the threshold value, the tasks be converted to the next task level. For example, consider that in the `NODE_LOCAL` task level, there is a data block located on the host A for the task named `Taska`. Without conversion to the `RACK_LOCAL` level, the scheduling algorithm would not be able to assign `Taska` to other hosts for execution (such as host B). In this approach, the task scheduler does not re-schedule `Taska` until the next scheduling decision is made. Consider a set of tasks ranging over the four different scheduling levels, including `PROCESS_LOCAL`, `NODE_LOCAL`, `RACK_LOCAL` and `ANY`. Assume that the waiting time due to scheduling delay for each task level is 2000 ms, and each level corresponds to a set of calculation containers, denoted by w_p , w_n , w_r and w_a , where $w_p \subseteq w_n \subseteq w_r \subseteq w_a$. The delay scheduler takes into account the level of each task and the corresponding delay waiting time and assign tasks to the appropriate level execution. If the level of the current task is `PROCESS_LOCAL`, this task will be assigned to w_p worker node for computing. When the task execution time exceeds the threshold time, the following task will be assigned to w_n worker node for computing. For surveillance video streams real-time analysis, such a mechanism would further increase data processing delay.

Another factor is the level of conversion based on a time interval that is set statically. It is difficult to adapt to changes that may occur during job execution. This approach is simple, but relies on experience to make decisions, which can be difficult in practice and results in poor performance. Real-time processing of video data handles a large amount of data and requires a lot of computing resources, if the execution time of a task results in a downgrade of the next task, it could affect the overall system throughput, thereby increasing delay time.

3 Design and Implementation

We choose a sliding time window, denoted by W_t , and record the execution time of each task within the time window. The execution time of task i within the time window is denoted as T_{Ei} . Assume that there are n tasks within the time window. The more recent tasks are given higher significance in the following formula, which computes the weighted sum of the execution time of each task within the time window.

$$F_{wt}(T_{E1}, T_{E2}, \dots, T_{En}) = \sum_{i=0}^n w_i * T_{Ei} \quad (2)$$

In the above formula w_i is the weight for the execution time of task i . For example, w_1 is the weight for the execution time of the first task in the time window, and w_n is the weight for the execution time of the latest task. Note that $w_1 < w_2 < \dots < w_n$ and $\sum_{i=0}^n w_i = 1$. F_{wt} is the weighted total execution time within the time

window, and is also known as the moving average time cost. F_{wt} is used to evaluate the computing capacity of a container node.

We use a structure called TaskFieldInfo to record some important information about each task execution, including task ID, execution time, execution position, and the data block feature. The specific format is as follows.

$$\text{TaskFieldInfo} = (\text{taskId}, \text{executionTime}, \text{position}, \text{feature}) \quad (3)$$

In TaskFieldInfo, position is implemented by concatenating the container number and host name; the block feature is used to represent surveillance video streams that need to be processed, such as IP surveillance video streams. The task information record is used to keep important information that is needed by the scheduling algorithm.

In practice, each data stream generated during the sliding window time produces a corresponding job. The jobs are decomposed into a set of tasks, denoted TaskC. The task set for the i -th job is denoted as task_i , where $\text{task}_i \in \text{TaskC}$. Each task has a task descriptor that includes task ID, data blocks that need to be processed, and the location information of the data blocks. Specifically, the format of each task descriptor is as follows:

$$\text{task}_i = (\text{taskId}, \text{data}, \text{position}) \quad (4)$$

When the task set is presented to TaskManager, the tasks are placed in a pending task queue to be processed, denoted to TaskQ. It is an ordered queue, based on the position of each task, which records the hostname and computing container.

To ensure the tasks completed within the time window the assignment of the tasks needs to satisfy the following two constraints:

- The task level, namely PROCESS_LOCAL, NODE_LOCAL, RACK_LOCAL or ANY, is used to determine the local priority of task allocation;
- The tasks need to be completed within a time limit. That is, the following restrictions need to be satisfied:

$$\sum executionTime_i \leq WINDOW_TIME \quad (5)$$

$\sum executionTime_i$ Represents the total execution time for the first n tasks, and $WINDOW_TIME$ represents the length of the stream processing time window.

The task scheduler works as follows. First, the scheduler tentatively assigns several tasks within the time window for the computing container. Then it computes the weighted average of the execution time. At this point, the scheduler dynamically evaluates the handling capacity of each computing container. Next, the task scheduler assigns tasks to different containers based on the weighted average of the execution time and the time limit, i.e., $WINDOW_TIME$. The scheduler begins

with tasks of `PROCESS_LOCAL` level for computing container task assignment task. If no assigned tasks are completed, then the subsequent tasks will be changed to `NODE_LOCAL` tasks. After each task completion, the scheduler will report on the execution status, and if the task is successfully completed, it will update the weighted total execution time.

The pseudo-code for the new task scheduling algorithm is presented below:

```

Algorithm: Scheduling Algorithm
Begin
1Sort tasks order by hostname, executorId and set tasks
level to PROCESS_LOCAL;
2While (task in tasks) do
3  taskLevel = getTaskLevel(task);
4  positionInfo ← getPositionInfo(task);
5  candidateExecutors = getCandidateExecutor(
taskLevel, positionInfo);
6  While (candidate in candidateExecutors) do
7    pendingTaskList =
pendingTasksForExecutor(candidateExecutor);
8    estimateTotalTime =
statisticEstimateExeTime(pendingTaskList, candidate); //
According to the task list and candidate position esti-
mate the execution time
9    If (estimateTotalTime + estimateTimeForTask(task))
<= WINDOW_TIME then
10     add task to pendingTasksForExecuor; break;
11   End If
12 End While
13 If task not assigned
14   set task level to NextSchedulerLevel
15 End If
16End While
End

```

4 Experiment

4.1 Setup

Work nodes include 5 ThinkServer RD640 servers, each of which is configured with a CPU Xeon E5-2620 2.1 GHz 24 core processor, 32 GB of DDR3 memory, a hard drive of 1 TB, and an Ethernet card of four gigabits. The master node is a Lenovo T4900V desktop consisting of a 3.6 GHz CPU Intel Core i7, DDR3 16 GB

DDR3 memory, 1 TB hard drive, and a gigabit Ethernet card. The work nodes and master node are connected using a Cisco 24 Gb Switch.

Each machine has several software packages installed, including the operating system Ubuntu 12.04 LTS, spark-1.0.2, JDK1.7, FFmpeg2.1.5, OpenCV2.4.8, ganglia3.6.0.

Tables 1 and 2 show the detailed configuration of each machine. The test program is a face detection and recognition program written using the opencv library implementation.

4.2 Experimental Results

We recorded runtime statistics about the tasks for each of the four task scheduling level, including PROCESS_LOCAL, NODE_LOCAL, RACK_LOCAL and ANY. The following commands are used to analyze the log:

```
grep "Starting task" /tmp/driver_running.log | grep -c "PROCESS"
grep "Starting task" /tmp/driver_running.log | grep -c "NODE"
grep "Starting task" /tmp/driver_running.log | grep -c "RACK"
grep "Starting task" /tmp/driver_running.log | grep -c "ANY"
```

We tested two runtime environments. One is Spark1.2.0 without making any changes. We will refer to this environment as Spark Original. The other is Spark1.2.0 that uses our optimized task scheduling algorithm. We will refer to this environment as Spark Improved.

Table 1 Master node configuration

Hardware configuration		Operating system	Basic software configuration
CPU type	Intel(R) Core(TM) i7-4790S CPU @ 3.60 GHz	Ubuntu12.04 64	Spark-1.0.2, JDK1.7, Ganglia3.6.0 data collection of components, control components, web components
The number of CPU cores	8		
Memory	16 GB		
IP	192.168.1.93		

Table 2 Slave node configuration

Hardware configuration		Operating system	Basic software configuration
CPU type	Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10 GHz	Ubuntu12.04 64	Spark-1.0.2, JDK1.7, FFmpeg2.1.5, OpenCV2.4.8, Ganglia3.6.0 data collection of components
The number of CPU cores	24		
Memory	32 GB		
IP	192.168.1.107–192.168.1.111		

As shown in Fig. 1, with Spark Original, the majority of the tasks are at the NODE_LOCAL level, and the number of tasks at the PROCESS_LOCAL level is negligible. After the examination, we found that the default task level of Spark is set to NODE_LOCAL. This explains why it would cause a lot of data exchange between computing containers. In the improved optimize scheduling algorithm, we set task level to PROCESS_LOCAL, and the number of tasks that execute locally is increased by 20 %. The performance of the real-time task scheduling strategy based on the local nature of the data block is very good, reduces the IO data transmission, increasing system performance.

Figure 2 shows the cluster utilization rate of CPU of the two scheduling algorithms with the same camera data. In this paper, we use the ganglia [11] to monitor cluster resources and collect runtime data during our experiment. As shown in

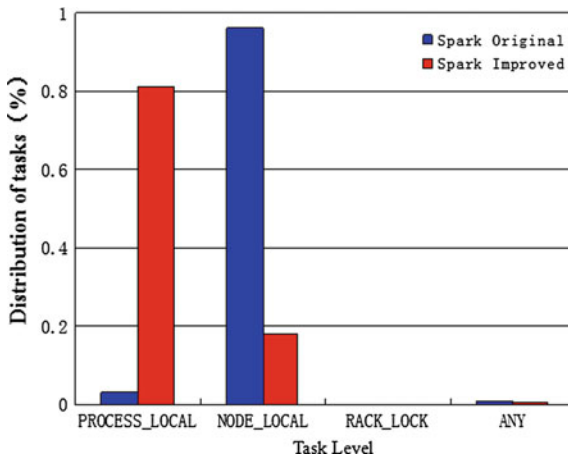


Fig. 1 Task distribution at different scheduling levels with spark original and spark improved

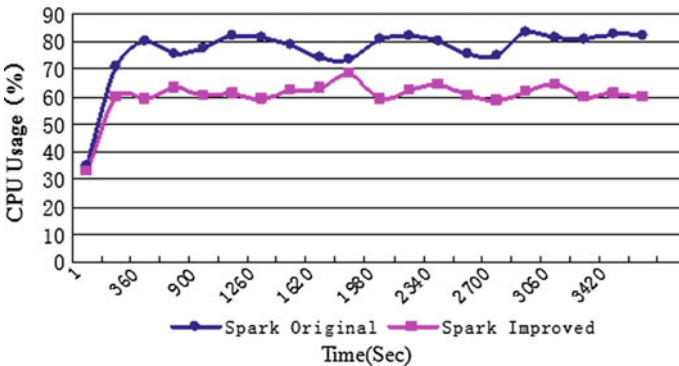


Fig. 2 Cluster CPU usage

Fig. 2 our scheduling algorithm has allowed CPU utilization to be increased by 10 %. Because most of the tasks does not require network transmission, i.e., they are executed locally, then cpu utilization is significantly increased. So the experimental results suggest that our proposed scheduling algorithm is more efficient for large-scale, real-time video streaming analysis.

5 Conclusions and Future Work

When using Spark to perform real-time analysis of large-scale video streaming, the default scheduling algorithm often shows low resource utilization rates, large amounts of network transmission, and higher real-time analysis delay. To address this problem, we present a new scheduling algorithm that exploits the notion of data locality to reduce the amount of network traffic, improve the utilization rate of CPU, and reduce the real-time analysis delay. The experimental results show that the new scheduling algorithm works well for video stream real-time analysis.

The new algorithm uses a weighted average to dynamically estimate task execution time. How to choose the weight of each task is a challenging problem and it depends on experience and domain knowledge. In the future we will investigate how to improve the accuracy of task execution time estimates, which can further improve the utilization of computing resources.

References

1. Zaharia M, Chowdhury M, Franklin MJ et al (2010) Spark: cluster computing with working sets. In: Proceedings of the 2nd USENIX conference on hot topics in cloud computing, pp 10–10
2. Kapil BS, Kamath SS (2013) Resource aware scheduling in Hadoop for heterogeneous workloads based on load estimation. In: 2013 Fourth international conference on computing, communications and networking technologies (ICCCNT). IEEE, pp 1–5
3. Zaharia M, Chowdhury M, Das T et al (2012) Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing. In: Proceedings of the 9th USENIX conference on networked systems design and implementation. USENIX Association, pp 2–2
4. Zaharia M, Das T, Li H et al (2012) Discretized streams: an efficient and fault-tolerant model for stream processing on large clusters. In: Proceedings of the 4th USENIX conference on hot topics in cloud computing. USENIX Association, pp 10–10
5. Lin SH (2000) An introduction to face recognition technology. *Informing Sci* 3(1):1–8
6. Dean J, Ghemawat S (2008) MapReduce: simplified data processing on large clusters. *Commun ACM* 51(1):107–113
7. Zaharia M, Borthakur D, Sen Sarma J, et al (2010) Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling. In: Proceedings of the 5th European conference on Computer systems. ACM, pp 265–278
8. Zaharia M, Konwinski A, Joseph AD, et al (2008) Improving mapreduce performance in heterogeneous environments. *OSDI* 8(4):7

9. Bezerra A, Hernández P, Espinosa A, et al (2013) Job scheduling for optimizing data locality in Hadoop clusters. In: Proceedings of the 20th European MPI users' group meeting. ACM, pp 271–276
10. Tian C, Zhou H, He Y, et al (2009) A dynamic mapreduce scheduler for heterogeneous workloads. In: Eighth international conference on grid and cooperative computing, GCC'09. IEEE, pp 218–224
11. Massie ML, Chun BN, Culler DE (2004) The ganglia distributed monitoring system: design, implementation, and experience. *Parallel Comput* 30(7):817–840

Microblogging Recruitment Information Mining

Jing Qin, Yiping Lu, Shuo Feng and Guiliang Feng

Abstract As microblog is becoming more and more popular, people not only begin to use microblog, getting and sharing information at any time, but also begin to do more based on microblog, such as using microblog for recruitment. Microblog recruitment has gradually become a kind of fashion, at the same time, also there are a lot of people using personal microblog to release information about the job or focus on hiring dynamics in real-time. This study classifies the recruitment information based on the text analysis tools SAS, extracts the specific information such as recruitment position, recruiters, hiring requirement, salary, work place, contact information, so that the job seekers can understand industry supply and demand dynamics, grasp the recruitment information in time easily.

Keywords Microblog · Information mining · SAS

J. Qin
Hebei University of Architecture, Zhangjiakou 075000, Hebei, China

Y. Lu · G. Feng (✉)
Hebei North University, Zhangjiakou, China
e-mail: 21475243@qq.com

S. Feng
Hebei Chinese Nanche Group Company, Shijiazhuang, China

1 Introduction

Compared with other social contact networking tools, microblog has the characteristics of in-time and fast. A lot of companies open official microblogs for recruiting work. In the candidate groups, microblog has become an important tool for job-hunting, job seekers can master the enterprise dynamics by focusing on their interested enterprises, it can save the trouble to search every enterprise website, also need not to find HR's contact distressing of related enterprise, microblog's rapid feedback narrow the distance between recruiters and job seekers. However, people are paying more and more attention to the microblogging recruitment today, besides focusing on a certain enterprise we are interested in, it is not easy for us to master the in-time accurate recruitment information from vast amounts of microblog data. There are a lot of irrelevant information through a simple keyword search, it produces very big interference to access the effectively information. If there is a tool that can accurately obtain the recruitment information and also search the information according to the relevant recruitment characteristics, such as salary, employers, job requirements and position, meanwhile it can control the rapidly updated recruitment information of microblog in-time, it will provide great help for job seekers.

2 The Mining Method

2.1 *Data Filtering Based on Keyword Search*

By observing the description characteristics of the recruitment information in microblog, we can find that the data which contains "urgent recruit, hire, recruit, positions, titles, job descriptions, job information, job requirements", and other general keywords are real recruitment information, we can extract microblog data of these keywords through the Python script, therefore as the first step to filter a large amount of data. Data filtration Program Code.

```

import re
import os
import codecs
list=["急招", "急聘", "诚招", "诚聘", "职位名称", "职位描述", "职位信息",
"职位要求", "职责名称"]
flist=os.listdir('extracted')
i=0
for fname in flist:
    if fname.endswith('.xml'):
        print 'processing file:%s' %fname
        fd=open('extract\\'+fname,'rb')
        content=fd.read()
        results=re.findall('<text>(.*?)</text>',content)
        if results:
            for result in results:
                text=result.strip().replace('\n',"").replace('\r',"")
                for zhaopin in list:
                    idx=text.find(zhaopin)
                    if idx !=-1:
                        f=open('weibo\\'+str(i)+'!.txt','wb')
                        f.write(text)
                        f.close()
                        i=i+1
            fd.close()

```

2.2 *Create Rule Model According to the Position Classification*

When classifying according to the position, we mainly refer to “Zhaopin Website”, jobs can be divided into IT, finance, real estate, service and management, machinery, marketing, production, sales, pharmaceutical, consulting, etc. There are total 11 classes. Then according to the characteristics of every kind of position, write classification rules. Through data matching results of classification rules, we can do the secondary filtration through keyword data fetching in Data filtration Program As shown in Fig. 1, the left column for this project divided into 11 kinds of position; Right column are the job matching rules for classification of types. Analysis of the classification and concept extraction results.

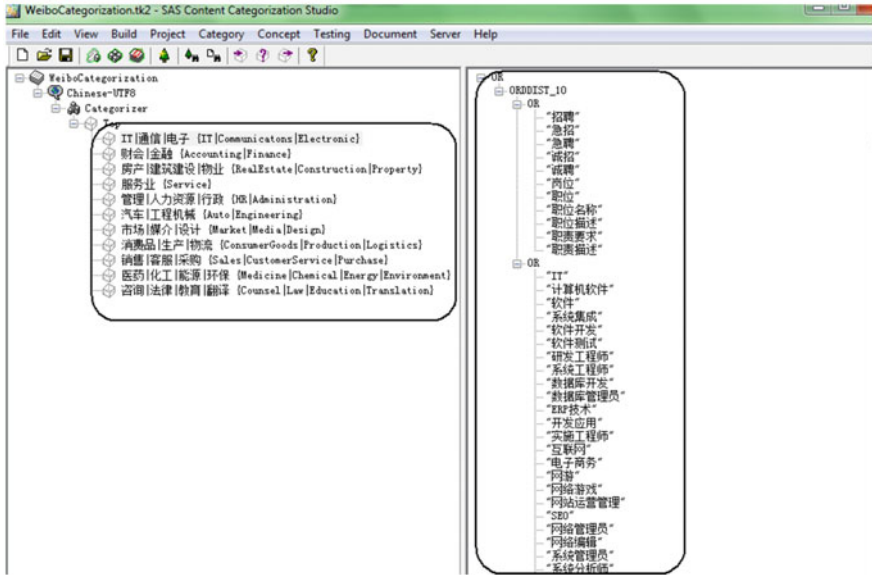


Fig. 1 Recruitment information classification and rules

Through the analysis of the data in 2.2, we can draw the following results report. Figure 2 is the results of the classification of 2.2.

Process the above two report through Python scripts, forming the Flex UI procedures required format XML report, after processing the results as follows in Fig. 3.

Based on the above four parts of the work, we can display microblog position distribution and the details of the relevant situation by Flex development interface.

```

./weibo/6417.txt:Top/Accounting|Finance; Top/Service
./weibo/3720.txt:Top/HR|Administration
./weibo/2707.txt:Top/IT|Communicatons|Electronic; Top/Market|Media|Design
./weibo/2876.txt:Top/IT|Communicatons|Electronic
./weibo/1405.txt:Top/Consumer Goods|Production|Logistics; Top/Medicine|Chemical|Energy|Environment
./weibo/3537.txt:Top/IT|Communicatons|Electronic
./weibo/3927.txt:Top/Sales|CustomerService|Purchase
./weibo/5299.txt:
./weibo/4090.txt:Top/Service
./weibo/5677.txt:

```

Fig. 2 Classification results report

Fig. 3 The classification results of XML format report

```

<?xml version="1.0" encoding="UTF-8"?>
-<xml>
  <category value="585" name="IT|通信|电子" id="5"/>
  <category value="239" name="财会|金融" id="0"/>
  <category value="194" name="房产|建筑建设|物业" id="8"/>
  <category value="546" name="服务业" id="10"/>
  <category value="558" name="管理|人力资源|行政" id="4"/>
  <category value="93" name="汽车|工程机械" id="1"/>
  <category value="845" name="市场|媒介|设计" id="6"/>
  <category value="462" name="消费品|生产|物流" id="2"/>
  <category value="137" name="医药|化工|能源|环保" id="7"/>
  <category value="750" name="销售|客服|采购" id="9"/>
  <category value="419" name="咨询|法律|教育|翻译" id="3"/>
</xml>

```

3 Software and Data

3.1 Software

50 million microblog data filtering, classification and the analysis of concept extraction are finished by using Python scripts. The extraction of concept of industry classification and recruitment information use SAS Content Categorization and SAS Concept Creation. The results show to use the FlexUI application development.

3.2 Used Data

50 million microblog corpora provided in the competition are used to carry out the experiments and results display. 5000 microblog data based on keyword search fetching, is used to write the classification rules and concepts extraction rules.

4 The System Architecture

The system architecture showed by Fig. 4 consists of four layers: Data Preprocessing, Data Preprocessing; The Model Building: modeling; Data Processing: Data Processing; User Interface: results show.

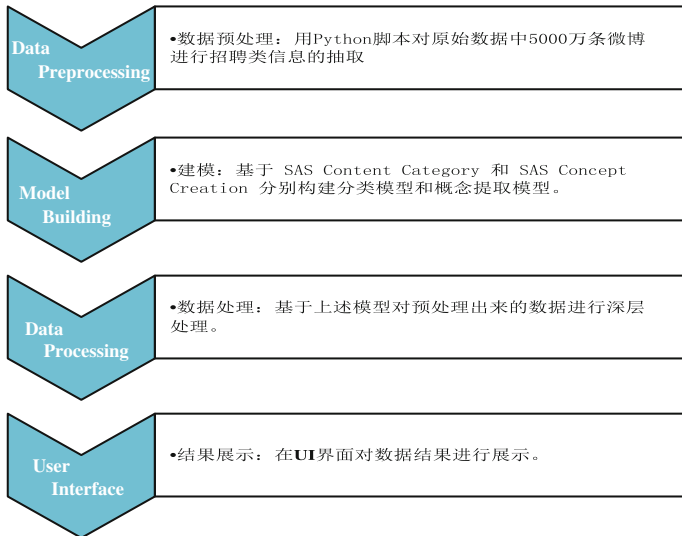


Fig. 4 System architecture diagram

5 Operation Steps

Click on the link, enter the position distribution of the total pie chart, as shown in Fig. 5.

Click on any industry, classification will pop up to the industry’s specific job information, as shown in Fig. 6.

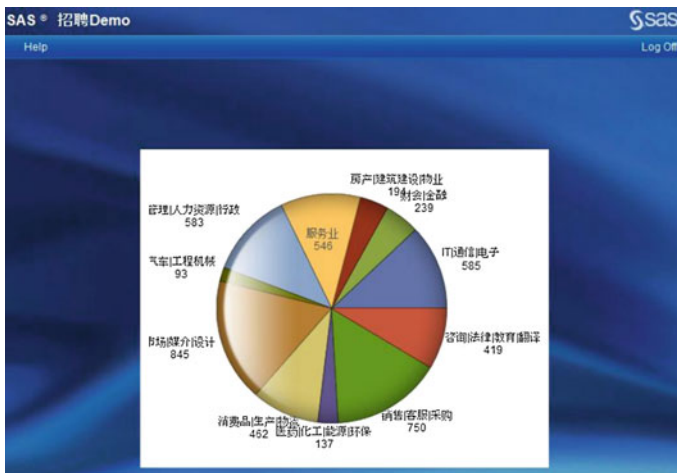


Fig. 5 Job profile

职位名称	招聘单位	招聘要求	工作地点	待遇	招聘对象	联系方式
冶金工程师-工程师	中国重型机械总公司		北京,北京-海淀区	面议	社会人才	
机械工程师-工程师	上海三木实业有限...		上海	面议	社会人才	
橡胶工程师-工程师	潍坊市亿隆汽车部...		浙江-温州-乐清...	面议	社会人才	
机械工程师-工程师		3年,3年以上,应届...	湖北,湖北省-湖北省...	面议	社会人才	
模具工程师-工程师...	上海金隆制品有...		上海,上海-松江区		社会人才	
焊工	湖南华康新材料有...		湖南,湖南-郴州-郴...	面议	应届毕业生,应届生...	
模具工程师-工程师...	上海金隆制品有...		上海,上海-松江区		社会人才	
维修工						电话: 18007736191

Fig. 6 Recruitment information concept extraction

6 Conclusion and Future Work

6.1 The Project Summary

Raw data provided by the competition is 50 million posts, we use Python script to extract the 9604 recruitment information related data according to the 12 keywords “now hiring”, “help wanted” or “sincerely recruitment”, “position name” and “job description”, “post information”, “job requirements”, accounting for 0.019208 % of the total number of all the microblogs, equivalent to about 0.02 %. after processing extracted 9604 microblogs based on classification model, information issued by 4853 microblogs are related to the industry. Jobs of different industry shown in microblogs and the percentage of the total number are respectively as follows.

Proportion of various industries in Table 1 values with a scatter diagram as shown in Fig. 7.

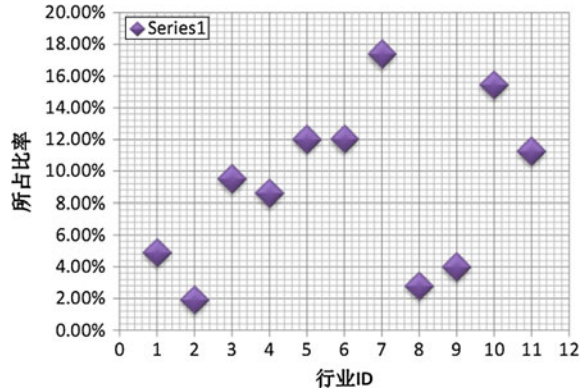
In Fig. 7 it shows, in microblog, the recruitment information is most associated with the “market, media and design” but the least position information related to the “car, construction machinery”. Proportion of various industries from high to low in the order:

1. Market|media|design.
2. Sales|customer service|procurement.
3. IT|communication|electronics.
4. Management|human resources|administration.
5. Service industry.
6. Production|Production|logistics.

Table 1 Careers distribution ratios for various sectors

Industry	ID	Recruiting number	Industry number	Percentage
Accounting finance	1	4853	239	4.92
Automotive construction machinery	2	4853	93	1.92
Consumer goods production logistics	3	4853	462	9.52
The advisory law education translation	4	4853	419	8.63
Management human resources administration	5	4853	583	12.02
IT communications electronics	6	4853	585	12.05
Market media design	7	4853	845	17.41
Pharmaceutical chemical energy environment	8	4853	137	2.82
Real estate construction property	9	4853	194	4.00
Sales support buy	10	4853	750	15.45
Services	11	4853	546	11.26

Fig. 7 Job distribution ratio scatter diagram



7. Advisory|law education|translation.
8. Accounting|financial.
9. The housing|building construction|property.
10. Medicine|chemical energy|environmental protection.
11. Automotive|construction machinery.

4853 microblogs are filtered based on classification model, then process its seven concepts using the concept extraction model, including: recruitment position, recruitment company, recruitment requirements, working place, salary, recruitment objects (points of social talents, fresh graduates, interns) and contact information. In the UI display screen, concept results which not mentioned in microblogs are automatically empty.

Due to the microblog content is limited by words, recruitment information based on this platform release is relatively simple. According to the result of concept extraction, we can have a perceptual knowledge of the extracted seven concepts. According to each of the concepts, recruitment information content can draw its importance degree from high to low:

1. Job position.
2. The job requirement.
3. The recruitment company.
4. Salary.
5. Contact.
6. Working location.
7. Recruiting object.

6.2 *Limitations and Future Work*

Factors that affect the results and its improved method:

Question 1: Method of extracting the data from the original data is not accurate enough. Out of consideration for time and the hardware and software resources, the project extracts recruitment information according to the keywords and uses the method in Python scripts. As the standard of selecting keywords is both can accurately represent the recruitment information and try to avoid ambiguity, so we used the 12 keywords mentioned in part 6.1, instead of the word “recruiting”, the most concrete representative (could cause the ambiguity such as the recruitment/recruitment website/recruitment centre for skills/information/comments, etc.). That are a direct result of this method which is recruiting information extraction is not completely, but more accurate.

Method: It will parse the raw data into a separate chapter microblog, Content Categorization model based on SAS Content Categorization Server to precisely filter and classify different industries from 50 million microblogs.

Question 2: For the sake of time, just dig the recruitment information, but no recruitment information is included in the project job.

Method: Additional information on the establishment of job classification rule model and concept extraction model.

Question 3: Demo show effect tends to static, impossible to relevant search and trend report generation. Due to the project participants is limited by a professional skills (not software development background), the project adopt the demo show effect tends to be static.

Method: Write corresponding program to achieve searching function for recruitment positions, recruitment company, working place and other specific information, and generate related trends report. And more importantly, able to provide in-time monitoring microblog for new recruitment and job information, and provide job matching and recommendation to recruiters and job-seekers accordingly.

Acknowledgments (1) Fund Project: QN2014203 “Smart car air conditioning energy conservation and emissions reduction optimization research”. (2) Fund Project: QN2014203 “Smart car air conditioning energy conservation and emissions reduction optimization research”.

References

1. Dessler G (2010) Human resource management. Prentice Hall, New Jersey
2. Kleiman LS (2013) Human resource management: managerial tool for competitive advantage. Atomic Dog, Cincinnati
3. Barrick MR, Ryan AM, Schmitt N (2012) Personality and work reconsidering the role of personality in organizations. Pfeiffer Wiley, New Jersey

4. Aamodt MG, Carr K (1998) Relationship between recruitment source and employee behavior. Paper presented at the annual meeting of the international personnel management association—assessment council, Las Vegas, NV
5. Zottoli MA, Wanous JP (2013) Recruitment source research: current status and future direction. *Hum Resour Manage Rev* 10(4):353–382
6. Weddle P (2015) Finding a job on the web: fast facts on job search tools and techniques. ISBN 978-1928734345
7. Hunter JE, Schmidt FL (2004) *Methods of meta-analysis: correcting error and bias in research findings*. Sage Publications, Newbury Park

Community Trust Recommendation Based on Probability Matrix Factorization

Xunfeng Li and Weimin Li

Abstract With increasing of using smart phones and the social network, the numerous and confused data makes people could hardly easily get what they really want. Although the recommendation systems have been vigorously searched for decades and successfully applied in the business world, they have faced old and new challenges now in the social network. In this paper, we propose a Community Trust model based on Probability Matrix Factorization (CT-PMF) and by using the ratings and networks we make predictions about users' no-rated items. Extensive experiments have been conducted to evaluate our model on the real data set obtained from Dianping. The experimental results demonstrate that CT-PMF outperforms competitor methods, in terms of predicting the missing rating.

Keywords Social network · Recommendation systems · Community trust · Probability matrix factorization

1 Introduction

Recommendation systems take use of people's purchase records and cookies, and recommend users purchase or other products they may be interested in. Web 2.0 users generate and disseminate large amounts of data, including some special data about users themselves, which make the recommendation systems turn to personalized recommendation while the traditional ones usually give the same recom-

X. Li · W. Li (✉)

School of Computer Engineering and Technology, Shanghai University,
Shanghai, China
e-mail: wml@shu.edu.cn

X. Li

e-mail: hqi11wn@gmail.com

X. Li · W. Li

Shanghai Key Laboratory of Computer Software Evaluating and Testing,
Shanghai University, Shanghai, China

mentation to all the users. It is a difficulty for people to get what they really want among the large amount of data. Many personalized recommendation algorithms are proposed to mine information that may be of users' interest and that of the users' attributes.

Collaborating filtering algorithm [1] is one of the most successful recommendation methods and it has been successfully applied in the field of commercial. However, traditional collaborative filtering methods only use the rating matrix, ignoring any other information, such as users' social network as well as the review of the users. Matrix factorization (MF) [2] and probability matrix factorization (PMF) [3] are two famous methods in collaborative filtering. MF was firstly proposed in Netflix competition and because of its high precision and high efficiency, it becomes a popular technology. PMF plays better than MF because PMF adds regularization to the loss function.

Recently, many researchers have turned to make use of the above information and relations. Xu et al. [4] adopted the LDA method and by analyzing the user reviews to the items, he got the latent communities of the users and the categories of the items. SBPR model [5] which is proposed by Tong Zhao takes the users' relations in the social network into account, improving the recommendation with both a warm-start and a cold-start. Jamali and Ester [6] also used the social network and MF method to analyze the ratings and relationships. Due to his use of both direct and indirect relations, the recommendation of his model works well when cope with cold start. In the real world, one's view to the items is susceptible to the people around, especially to close friends and close relatives, and he tends to accept their opinions or advice. On social networks, it also works and people are affected to opinions spread by their following. When recommending items to new users with little useful data in their purchase records, recommendation systems are often unprepared or lost. This is because the systems can hardly get some information available. But the social network can be infinite and people's friends, those they trust, their friends' friends are all included in the network.

Social networks make the recommendation algorithm could get some useful information, but most of the existing methods do not consider the social networks. The information, which enables the recommendation systems to cope with the rating matrix sparse, is necessary to new users on the network (cold-start users). We can get this information through users' behavior on online social platforms. However, it is universally acknowledged that the social network connections online always contain a great many weak ties [7] and a lot of noise, while, clan trusts exit in strong connections. But most of the time the taking social networks into account methods treat the contacts of users as trusts directly. However, in the social networks a relation usually not means a trust relation. Just because of an approval behavior, a weak relationship, we should not say that it is reliable. In reality, individuals tend to trust and rely on people around them: relatives, colleague, schoolmates, etc. This is because they have the same properties, such as community property.

Here we introduce the concept of community to recommendation. The connections in a community usually are strong ties and people in a community

establish trust relationships. In this paper, we make use of the ratings generated online, classify the users into communities and then, in perspective of the trust relationship between users and their following, quantify the trust of them. We propose a community trust model based on probability matrix factorization and by using the ratings and the networks we make a prediction about the users' unrated items. Experiments on real datasets show that our community trust model CTPMF outperforms the traditional methods, MF and SocialMF, in terms of ratings predicting accuracy.

2 Community Trust Model Based on Probability Matrix Factorization (CT-PMF)

We propose a new model which models the realistic society trust relations of users on the social networks. In our model, firstly, we should obtain communities of the users. To get the communities, we can purify the information from the locations of the users, their purchase records, the occupations of the users, etc. We can also get the communities by means of classification algorithms. In our work, we adopt K-means cluster algorithm and consider the ratings data of different users as input. And we give each of the users one community attribute c . In generally, users in the some community have similar preferences and their trust relationships are strong.

$$V_{u_1u_2} = C_{u_1u_2}t_{u_1u_2} \quad (1)$$

where $V_{u_1u_2}$ represents the trust value of u_1 to u_2 and $t_{u_1u_2}$ is an indicator function. If u_1 have a relation to u_2 in Matrix $T_{N \times N}$, $t_{u_1u_2}$ is 1; if not, it is 0.

The social network is $T_{N \times N}$. Usually the matrix $T_{N \times N}$ is not symmetrical. For example, user u_1 is following user u_2 , it doesn't mean that user u_2 is following user u_1 , too; user u_1 agrees to the review of user u_2 , it doesn't mean that user u_2 is agree to user u_1 , or user u_1 may have never written a review on the network. And user u_1 agrees to user u_2 , user u_2 agrees to user u_3 , generally, we should not spread the relationship, believing that user u_1 agrees to user u_3 . It is not difficult to draw that V and T have the same dimension and we assumption that:

$$C_{u_1u_2} = \begin{cases} \alpha, & u_1, u_2 \in \text{the same } c \\ \beta, & \text{others} \end{cases} \quad (2)$$

where C is a value matrix and $C_{u_1u_2}$ is an element in it. The value of α and β in the formula can be obtained through practical machine learning.

In our model, the idea of probability matrix factorization technique is adopted to learn the user latent factor matrix P and the item latent factor matrix Q . Matrix P is a $K \times N$ matrix and Q is $K \times M$. The rating matrix R is similar to the product of P and Q :

$$R \approx Q^T P \tag{3}$$

Each line of the matrix P represents a user’s latent factor vector and each line of the matrix Q represents an item latent factor vector. The prediction rating \hat{r}_{ui} of the missing rating in R is:

$$\hat{r}_{ui} = P_u Q_i^T \tag{4}$$

where u is the number of a user and i is the number of an item. P_u is the line u in matrix P and Q_i is the line i in matrix Q .

Of course the prediction is inaccurate. The error can be assumed that it follows a Gaussian distribution, with an expectation 0 and standard deviation σ . Thus, for all the existing ratings in R , they follow the conditional probability:

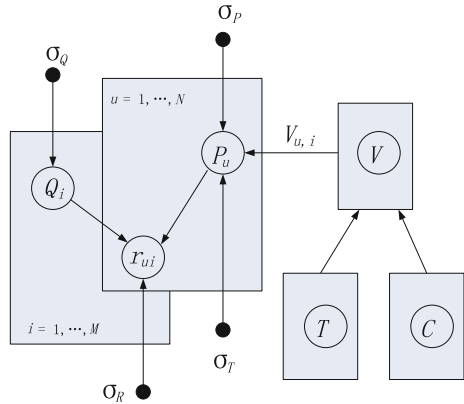
$$p(R|P, Q, \sigma_R^2) = \prod_{u=1}^N \prod_{i=1}^M [G(r_{ui}|PQ_i^T, \sigma_r^2)]^{I_{ui}^R} \tag{5}$$

where $G(x|\mu, \sigma^2)$ represents a Gaussian distribution with mean 0 and variance σ^2 . I_{ui}^R is an indicator function and if user u has rated item i , I_{ui}^R is 1; if not, it is 0. Basic probability matrix factorization needs to add a Gaussian prior with mean 0 to the user’s latent factor and the item’s latent factor. Different from the basic probability matrix factorization, we also need to add a Gaussian prior to the user’s latent factor with the influence of trust.

$$p(Q|\sigma_Q^2) = \prod_{i=1}^M G(P_u | \sum_{i \in T} V_{u,i}, \sigma_T^2 I) \tag{6}$$

Our proposed model can be represented in Fig. 1. In our model we use $U = \{u_1, u_2, \dots, u_N\}$ to represent the users set and N is the number of users;

Fig. 1 The proposed model CT-PMF



$I = \{i_1, i_2, \dots, i_M\}$ is the items set and M is the number of items. We use R represents the user-item rating matrix and often it is a sparse matrix. r_{ui} is the rating score of user u to item i and it is a integer range from 1 to 5. r_{ui} is decided by P_u , Q_i and a Gaussian prior σ_R . P_u is decided by Gaussian priors σ_R , σ_P and the community trust value which is decided by the community he belongs to and the relations on social networks.

According to Fig. 1, after a Bayesian inference we can get the prior probability of our model, and probability equation is as follows (in log):

$$\begin{aligned} \ln p(P, Q|R, T, V, \sigma_R^2, \sigma_T^2, \sigma_P^2, \sigma_Q^2) &= -\frac{1}{\sigma_R^2} \sum_{u=1}^N \sum_{i=1}^M I_{u,i}^R (r_{ui} - P_u^T Q_i)^2 \\ &\quad - \frac{1}{\sigma_T^2} \sum_{u=1}^N (P_u - \sum_{i \in T} V_{u,i} P_i)^T (P_u - \sum_{i \in T} V_{u,i} P_i) \\ &\quad - \frac{1}{\sigma_P^2} \sum_{u=1}^N \|P\|^2 - \frac{1}{\sigma_Q^2} \sum_{i=1}^M \|Q\|^2 - \frac{1}{2} \ln \sigma_R^2 \sum_{u=1}^N \sum_{i=1}^M I_{u,i}^R \\ &\quad - \frac{1}{2} (NK \ln \sigma_P^2 + MK \ln \sigma_Q^2 + NK \ln \sigma_T^2) + C \end{aligned} \quad (7)$$

where $\sum_{i=1}^m \|P\|^2$ is the sum of squares of the data in user latent factor matrix. Similarly, $\sum_{i=1}^n \|Q\|^2$ is the sum of squares of the data in item latent factor matrix. C is a constant which means nothing to the model. Maximum of the posteriori probability is equivalent to minimize the loss function below:

$$\begin{aligned} Loss(R, V, T, P, Q) &= \frac{1}{2} \sum_{u=1}^N \sum_{i=1}^m I_{u,i}^R (r_{ui} - P_u^T Q_i)^2 \\ &\quad + \frac{\gamma_V}{2} \sum_{u=1}^N (P_u - \sum_{i \in T} V_{u,i} P_i)^T (P_u - \sum_{i \in T} V_{u,i} P_i) \quad (8) \\ &\quad + \frac{\gamma_P}{2} \sum_{u=1}^N \|P\|^2 + \frac{\gamma_Q}{2} \sum_{i=1}^M \|Q\|^2 \end{aligned}$$

where $\gamma_P = \sigma_R^2 / \sigma_P^2$, $\gamma_Q = \sigma_R^2 / \sigma_Q^2$, $\gamma_V = \sigma_R^2 / \sigma_T^2$.

To solve our model we could use the Alternative Least Squares method (ALS) as well as the Stochastic Gradient Descent (SGD) method. Both of these two methods need to take the partial derivatives of P_u and Q_i with respect to the loss function:

$$\left\{ \begin{array}{l} \frac{\partial Loss}{\partial P_u} = -(r_{ui} - P_u^T Q_i) \sum_{i=1}^M I_{u,i}^R Q_i + \gamma_V \left(P_u - \sum_{i \in T} V_{u,i} P_i \right) \\ \quad + \gamma_V \sum_{u \in T^T} V_{i,u} \left(P_i - \sum_{i \in T} V_{i,m} P_m \right) + \gamma_P P_u \\ \frac{\partial Loss}{\partial Q_i} = -(r_{ui} - P_u^T Q_i) \sum_{i=1}^M I_{u,i}^R Q_i + \gamma_Q Q_i \end{array} \right. \quad (9)$$

Then, adjust the parameter along the decline speed of direction and update the two latent factor matrixes. The update rule can be represented with the following formula. v is the learning rate of the model.

$$\left\{ \begin{array}{l} P_u = P_u \leftarrow v \frac{\partial Loss}{\partial P_u} \\ Q_i = Q_i \leftarrow v \frac{\partial Loss}{\partial Q_i} \end{array} \right. \quad (10)$$

3 Experiment

3.1 Dataset and Experiment Setup

The dataset we use is collected for a week from Dianping, the online website. The collected user data rates at least once and the items (hotels) are scored at least one time. The datasets contains 45,109 users, 3383 items, 92,290 ratings as well as 134,560 relationships between the users. The dataset is very sparse with the sparsity of $1 - 92290 / (45109 * 3383) = 0.06\%$, while the sparsity of MovieLens datasets and Netflix dataset are 4.5 and 1.2%. So, we carry on a data processing and retain the users who at least rate 5 times. The processed dataset information is as shown in Table 1.

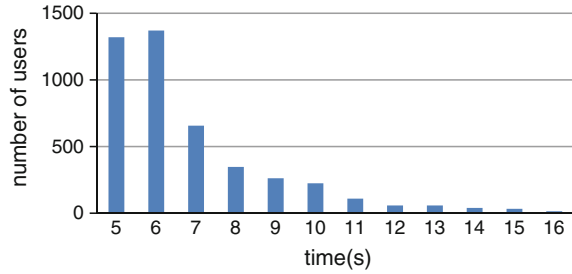
According to Fig. 2, the statistic suggests that the frequency of users' rating follows power-law distribution.

In order to verify the effectiveness of our proposed model, we conduct comparison with the MF model, which is popular in recommendation systems. But the

Table 1 Statistics of dataset

Category	Number
Users	4689
Items	2608
Ratings	36,106
Social relations	17,382
Average relations every user	3.7
Average ratings every item	13.8
Average ratings every user	7.7

Fig. 2 The frequency of users rate items



MF model ignores the social relations, we compare with SocialMF [6], which takes the social relations into account. In our experiment, we do not concern about relation spread.

Recommendation systems, especially the rating-based ones, are often evaluated by the prediction accuracy. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are common evaluation methods in the field of Recommendation. Smaller MAE value indicates higher prediction accuracy.

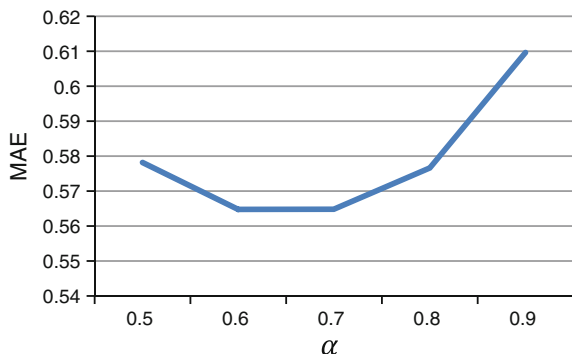
$$MAE = \frac{\sum_{u=1}^N \sum_{i=1}^m Test_{ui} |r_{ui} - \hat{r}_{ui}|}{\sum_{u=1}^N \sum_{i=1}^m Test_{ui}}, RMSE = \sqrt{\frac{\sum_{u=1}^N \sum_{i=1}^m Test_{ui} (r_{ui} - \hat{r}_{ui})^2}{\sum_{u=1}^N \sum_{i=1}^m Test_{ui}}} \tag{11}$$

If r_{ui} is in the test set, $Test_{ui} = 1$ and otherwise, $Test_{ui} = 0$. In our experiment, we use these two indicators to measure the prediction of the recommendation methods.

Each row of the rating matrix R can be seen as a user’s latent feature vector. We use K-means algorithm classifies the users into communities and every user has one community number. We make $\beta = 1 - \alpha$ and set the community trust parameter $\alpha = 0.5, 0.6, 0.7, 0.8, 0.9$. Results are shown below.

From Fig. 3 we can conclude that the community trust does have an effect on the result of the experiment, and when $\alpha = 0.6$, MAE obtained the minimum, the most accurate prediction results.

Fig. 3 The influence of α to MAE



In order to verify the true prediction accuracy our method performs, we use the 10 cross-validation, divide the data into 10 portions and take turns take one of them as a test data set. We put the results of 10 times average as the result of the experiments.

In the experiments, we set the model parameter as follows: $\gamma_P = 0.01$, $\gamma_Q = 0.01$, $\gamma_V = 0.01$.

3.2 Results on Rating Prediction

As can be seen from Figs. 4 and 5, our proposed method, compared with MF and SocialMF, outperforms the baselines as the curve of our proposed model is always under the other two curves, indicating that our proposed method achieves higher prediction accuracy. We can find that our method convergence earliest at the fifteenth interaction while SocialMF is at nearly 25.

Fig. 4 MAE curve variation with iterations

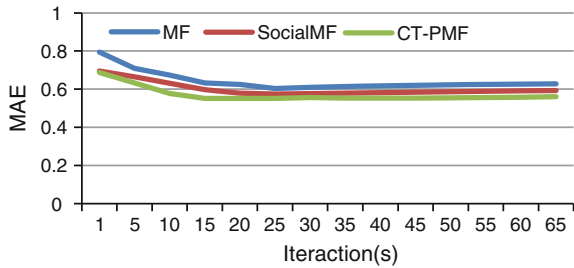
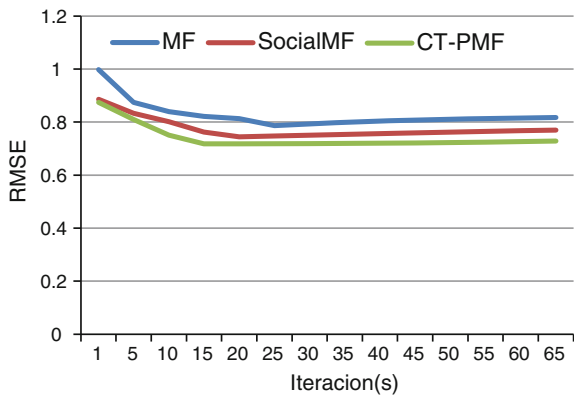


Fig. 5 RMSE curve variation with iterations



4 Conclusion

In this paper, based on the observation that a user and his trust following select similar services, we take users' social network and their community into account and propose a new model CT-PMF for predicting users' preference from the social data. By incorporating the ratings generated by users we unearth the users' hidden communities and by making use of these communities, we enhanced the performance of recommendation system. Extensive experiments have been conducted to evaluate our proposal on the real data set obtained from Dianping. The experimental results demonstrate that our model CT-PMF outperforms MF and SocialMF, in terms of predicting the missing rating.

References

1. Marlin B, Zemel RS, Roweis S et al (2012) Collaborative filtering and the missing at random assumption. arXiv preprint [arXiv:1206.5267](https://arxiv.org/abs/1206.5267)
2. Purushotham S, Liu Y, Kuo CCJ (2012) Collaborative topic regression with social matrix factorization for recommendation systems. arXiv preprint [arXiv:1206.4684](https://arxiv.org/abs/1206.4684)
3. Shan H, Kattge J, Reich P et al (2012) Gap filling in the plant kingdom—trait prediction using hierarchical probabilistic matrix factorization. arXiv preprint [arXiv:1206.6439](https://arxiv.org/abs/1206.6439)
4. Xu Y, Lam W, Lin T (2014) Collaborative filtering incorporating review text and co-clusters of hidden user communities and item groups. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management. ACM, pp 251–260
5. Zhao T, McAuley J, King I (2014) Leveraging social connections to improve personalized ranking for collaborative filtering. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management. ACM, pp 261–270
6. Jamali M, Ester M (2010) A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the fourth ACM conference on recommender systems. ACM, pp 135–142
7. Levin DZ, Cross R (2004) The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. *Manage Sci* 50(11):1477–1490

Robust Markov Random Field Model for Image Segmentation

Taisong Xiong and Yuanyuan Huang

Abstract Finite mixture model (FMM) obtains good results when it is applied to image segmentation under non-noise condition. But it cannot obtain satisfied segmentation results when the image is degraded by noise. The main reason is that it regards the relationship of pixels are statistical independent and the position information has no effect for image segmentation. However, in fact, the spatial relationship between these pixels play an important role for image segmentation. Markov Random Field (MRF) considers the spatial information of pixels and has been widely applied image segmentation. In this paper, combine FMM and MRF, a robust MRF model is proposed. The proposed model effectively captures the spatial information of pixels and is applied to color real world image segmentation. Visual and quantitative experimental results demonstrate the robustness, effectiveness and correctness of proposed model.

Keywords Finite mixture model · Markov random field · Image segmentation · EM algorithm

1 Introduction

Finite mixture model (FMM) is one flexible statistic tool and successfully applied to many fields [1]. It obtains better results when it is applied to image segmentation [2]. The FMM is referred to as Gaussian Mixture model (GMM) if its component function is Gaussian distribution. The FMM cannot obtain satisfied results when the

T. Xiong

College of Applied Mathematics, Chengdu University of Information Technology,
Chengdu 610225, People's Republic of China
e-mail: xiongtaisong@gmail.com

Y. Huang (✉)

Department of Network Engineering, Chengdu University of Information Technology,
Chengdu 610225, People's Republic of China
e-mail: iyyhuang@hotmail.com

images are degraded by noise [3]. The main reason is that FMM regards the pixel is independent of each other. However, the probability belonging to the same class is higher if the distance of these pixels is nearer [4]. In fact, the spatial information plays a very important role in image process.

To obtain better image process results, Markov random field (MRF) considers context-dependent relationship and correlated features of pixels was proposed in [5, 6]. MRF models have been successfully applied to many fields, such as image restoration, edge detection, image segmentation [7]. At the same time, some mixture models based on MRF have been proposed for image segmentation [8–10]. The main distinction between the GMM and the mixture models based on MRF is the representation of their context mixture coefficient. In GMM, the context mixture coefficient π_k is independent of the pixel. But the context mixture coefficient π_{nk} in the mixture model based on MRF is closely related to the pixel. The context mixture coefficient π_{nk} reflects the spatial relationship of the pixel x_n and its neighborhood system. Many variants of mixture model based on MRF have been proposed. In general, many different representations or priori distributions are imposed on π_{nk} [3, 11]. To infer the parameters of these models, EM algorithm is in general adopted for most models [12, 13]. But some models cannot obtain closed form solutions or their computational complexity is very high.

Based on these aforementioned models and inspired by [11], a mixture model based on MRF is proposed in this paper. In the proposed model, the spatial relationship between the pixel and its neighborhood system is reflected by the posterior probability of the pixels of its neighborhood system. It effectively captures the spatial between these pixels. Some numerical experimental results are obtained conducted on real world color images to demonstrate the effectiveness, robustness and correctness of the proposed model compared with some other mixture models.

The remainder of this paper is organized as follows. A theoretical background about MRF is introduced in brief in Sect. 2. The description of the proposed model is given in Sect. 3 in detail. In Sect. 4, some experiments conducted on real world color images are given to demonstrate the performance of the proposed model. At last, some conclusions are presented in Sect. 5.

2 The Theoretical Background

In this section, the theoretical background of mixture model based on MRF is introduced in brief. The representation a pixel x_n 's GMM is written as follows [14].

$$f(x_n) = \sum_{k=1}^K \pi_k N(x_n | \mu_k, \Sigma_k). \quad (1)$$

where the Gaussian distribution is given by

$$N(x_n|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{(D/2)} |\Sigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right\}. \quad (2)$$

where Σ_k and μ_k represent the covariance matrix and mean of Gaussian distribution, respectively. It can be seen from (1) that the context mixture parameter π_k is independent of pixel x_n . At the same time, π_k must satisfy the following constraints.

$$0 \leq \pi_k \leq 1, k = 1, 2, \dots, K. \quad (3)$$

While the representation of mixture model based on MRF is written as follows [3].

$$f(x_n) = \sum_{k=1}^K \pi_{nk} N(x_n|\mu_k, \Sigma_k). \quad (4)$$

Similarly, the context mixture parameter π_{nk} also should satisfy the following constraints.

$$0 \leq \pi_{nk} \leq 1, n = 1, 2, \dots, N, k = 1, 2, \dots, K. \quad (5)$$

Compared FMM with mixture model based on MRF, it is very obvious to demonstrate that the mixture model based on MRF considers the spatial relationship between the pixels. The all observations are regarded as statistical independent. The joint conditional density of observations $X = (x_1, x_2, \dots, x_N)$ can be written as follows [11].

$$P(X|\Pi, \Theta) = \prod_{n=1}^N f(x_n) = \prod_{n=1}^N \sum_{k=1}^K \pi_{nk} N(x_n|\mu_k, \Sigma_k) \quad (6)$$

To capture the spatial relationship between the pixels, MRF prior is imposed on Π [7].

$$P(\Pi) = Z^{-1} \exp\left\{-\frac{1}{T} U(\Pi)\right\}. \quad (7)$$

where parameter Z and T represent a normalizing constant and a smooth constant, respectively. The function $U(\Pi)$ is used to reflect the spacial information of pixels. In [15], $U(\Pi)$ is defined as

$$U(\Pi) = - \sum_{x_i \in V} \left\{ t_{z_i} v_i + \frac{q}{2} \sum_{j \in N(i)} V_{ij}(z_i, z_j) \right\}. \quad (8)$$

For more details about (8), the reader should refer to [15]. In [4], the $U(\Pi)$ is written as following.

$$U(\Pi) = \beta \sum_{n=1}^N \sum_{m \in \partial_n} \sum_{k=1}^K (\pi_{nk} - \pi_{mk})^2. \quad (9)$$

3 The Proposed Model

Inspired by [11], in this paper, we proposed a mixture model based on MRF. In the proposed model, we defined a factor about an observation x_n . Its definition is given by

$$F_{nk} = \exp \beta \left(z_{nk} + \frac{1}{N_n} \sum_{m \in \partial_n} z_{mk} \right) \quad (10)$$

where z_{nk} denotes the posterior probability and ∂_n is the neighborhood system of x_n . Throughout this paper, a second neighborhood system is chosen for the proposed model. N_n is the number of neighbor in the neighborhood system ∂_n . Parameter β is a smooth factor. The factor only depends on the posterior probability of a pixel.

Then, a smoothing prior $U(\Pi)$ is given by

$$U(\Pi) = - \sum_{n=1}^N \sum_{k=1}^K G_{nk} \log(\pi_{nk}). \quad (11)$$

According to the smoothing prior, we obtain a MRF distribution whose definition is given by

$$\phi(\Pi) = \frac{1}{Z} \exp \left\{ -\frac{1}{T} U(\Pi) \right\} = \frac{1}{Z} \exp \left\{ -\frac{1}{T} \sum_{n=1}^N \sum_{k=1}^K G_{nk} \right\}. \quad (12)$$

When the MRF distribution is determined, the log-likelihood function can be written as follows

$$L(X|\Pi, \Theta) = \sum_{n=1}^N \log \left\{ \sum_{k=1}^K \pi_{nk} p(x_n|\theta_k) \right\} - \log Z - \frac{1}{T} \sum_{n=1}^N \sum_{k=1}^K G_{nk} \quad (13)$$

where the component function $p(x_n|\theta_k)$ is in general chose Gaussian distribution. Its definition is given by in (2). When the all data is determined, maximizing (13) can be rewritten as follows.

$$J(X|II, \Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \{\log \pi_{nk} + \log p(x_n|\theta_k)\} - \log Z - \frac{1}{T} \sum_{n=1}^N \sum_{k=1}^K G_{nk} \quad (14)$$

where z_{nk} represent the conditional expectation values. Its value can be obtained by the following equation

$$z_{nk} = \frac{\pi_{nk} p(x_n|\theta_k)}{\sum_{j=1}^K \pi_{nj} p(x_n|\theta_j)}. \quad (15)$$

To optimize the parameter set $\{II, \Theta\}$, we maximize the objective function (14). Similar to the MRF in [4, 9, 10], the values of Z and T are both set to one. According to (14), we obtain a new objective function which is given by

$$J(X|II, \Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \{\log \pi_{nk} + \log p(x_n|\theta_k)\} - \sum_{n=1}^N \sum_{k=1}^K G_{nk}. \quad (16)$$

At the same time, according to the Gaussian distribution given in (2), the objective function can be rewritten as follows.

$$J(X|II, \Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \left\{ \log \pi_{nk} - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_k| \right. \\ \left. - \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \right\} - \sum_{n=1}^N \sum_{k=1}^K G_{nk} \quad (17)$$

To maximize the objective function given in (17), we adopt the EM algorithm [12, 13] to inference the objective function.

To obtain the solution of $\partial J / \partial \mu_k = 0$, we have

$$\mu_k = \frac{\sum_{n=1}^N z_{nk} x_n}{\sum_{n=1}^N z_{nk}}. \quad (18)$$

Then the solution of $\partial J / \partial \Sigma_k^{-1} = 0$ about Σ_k equals to

$$\Sigma_k = \frac{\sum_{n=1}^N z_{nk} (x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^N z_{nk}}. \quad (19)$$

To obtain the optimization solution of π_{nk} , an important constraint $0 \leq \pi_{nk} \leq 1$ should be considered. Then, we use the Lagrange multiplier to obtain the solution of π_{nk}

$$\pi_{nk} = \frac{z_{nk} + G_{nk}}{1 + \sum_{j=1}^K G_{nj}} \quad (20)$$

The steps of proposed model can be summarized as follows.

- Step 1 Initialize: the values of the means μ_k and covariance matrix Σ_k are determined by K -means, set $\pi_{nk} = \frac{1}{K}$.
- Step 2 E-Step Calculate the Gaussian distribution $p(x_n|\theta_k)$ in (2) using the current parameters. Calculate the conditional expectation value z_{nk} in (15) and factor F_{nk} in (10). Calculate G_{nk} .
- Step 3 M-Step Update the mean μ_k and covariance matrix Σ_k in (18) and (19), respectively. Update context mixture parameter π_{nk} in (20).
- Step 4 Check the convergence of the log-likelihood function according to (13) or according the numbers of iteration. The iteration is end if the convergent conditions are satisfied.

When all parameters are optimized, the label of pixel x_n is determined by the max posterior probability

$$\arg \max_k \{z_{nk}\}. \quad (21)$$

4 Experimental Results

In this section, to evaluate the performance of the proposed model, some models are chosen, such as GMM, the mean field algorithm (MEANF) [16], CA-SVFMM [10], MRFGMM [11]. To quantitatively evaluate the performance of the image segmentation, the probabilistic rand (PR) index [17] is chosen. Its value varies from 0 to 1. The segmentation result is better if the value of PR is higher. A image segmentation database is chosen [18] to evaluate the performance of these models. In this database, there are 500 color real world images.

First, a color image shown in Fig. 1a whose no is 326025 is chosen for visual effectiveness. The image is divided into four parts: background, flower, leopard and grass. The segmentation results obtained by GMM, MEANF, CA-SVFMM, MRFGMM and proposed model are shown in Fig. 1b–f, respectively. It can be seen from the Fig. 1f, the background is correctly segmented from the image by the proposed model. The contour of all objects in the proposed model is more smooth than any other model. At the same time, proposed model obtains the highest PR compared with any other mode. It demonstrates that the proposed model obtains better segmentation result than any other model. It also proves that the effectiveness and robustness of proposed model for image segmentation.

To further verify the correctness and robustness of the proposed model, another color real world image is chosen. The original image shown in Fig. 2a is segmented into four classes. The image consists of sky, house, door and corridor. From the segmentation results shown in Fig. 2b–f, the proposed model and MRFGMM segment the sky very well, the difference between the sky and the other objects is very clear. The sky is wrongly segmented by the other models. However, the

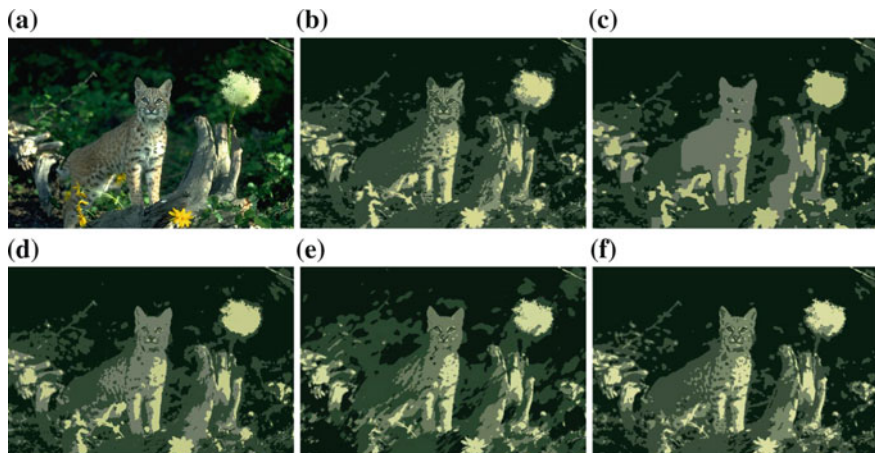


Fig. 1 Color image segmentation (326025). **a** The original image, **b** GMM PR = 0.691, **c** MEANF PR = 0.724, **d** CA-SVFMM PR = 0.715, **e** MRFGMM PR = 0.663, **f** Proposed method PR = 0.752

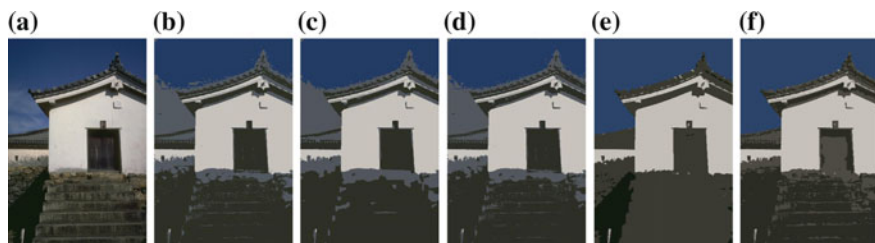


Fig. 2 Color image segmentation (334025). **a** The original image, **b** GMM (PR = 0.777), **c** MEANF (PR = 0.785), **d** CA-SVFMM (PR = 0.781), **e** MRFGMM (PR = 0.808), **f** Proposed method (PR = 0.834)

proposed model can correctly segment the corridor compared with MRFGMM. Furthermore, the proposed model obtains the highest PR value among the all models. The experiment shows that the correctness and robustness of proposed model is more better than any other model.

To quantitatively evaluate the proposed model for image segmentation, 18 color images are chosen to apply to image segmentation. The proposed model is applied to image segmentation compared with GMM, MEANF, CA-SVFMM, MRFGMM. The PR values obtained by these models are given in Table 1. From Table 1, we can see that the PR values obtained by proposed model are all higher than any other model. Of course, its mean is also higher than the other models. These results of image segmentation furthermore prove that the proposed model is more robust and effective than some other model for image segmentation. Therefore, it proves that the spatial relationship is effectively captured in our model.

Table 1 Comparison of image segmentation results based on Berkeley images: PR index

Image	K	GMM	MEANF	CA-SVFMM	MRFGMM	Proposed model
15062	5	0.796	0.816	0.814	0.787	0.831
28083	5	0.863	0.859	0.866	0.903	0.911
33044	2	0.564	0.567	0.557	0.639	0.635
43051	3	0.802	0.793	0.803	0.824	0.829
103029	3	0.554	0.577	0.564	0.579	0.593
69007	4	0.808	0.792	0.810	0.851	0.838
117025	3	0.793	0.832	0.811	0.835	0.835
2018	5	0.768	0.788	0.773	0.786	0.783
220003	5	0.732	0.746	0.730	0.734	0.742
372019	4	0.768	0.768	0.768	0.788	0.783
226033	5	0.722	0.716	0.726	0.719	0.737
232076	2	0.593	0.590	0.597	0.559	0.617
250047	2	0.708	0.709	0.709	0.710	0.721
296028	5	0.785	0.771	0.787	0.807	0.807
306051	4	0.742	0.745	0.755	0.742	0.783
326025	4	0.691	0.724	0.715	0.663	0.752
344010	3	0.726	0.727	0.724	0.749	0.751
334025	4	0.777	0.785	0.781	0.808	0.834
Mean	–	0.733	0.739	0.738	0.749	0.766

5 Conclusions

In this paper, we proposed a mixture model based on Markov random field for image segmentation. In the proposed model, the spatial relationship between is based on the posterior probability of its neighborhood system. The algorithm to infer the parameters of the proposed model is EM algorithm. The visual and quantitative segmentation results all demonstrate that the proposed model is more effective, robust and corrective compared with some other models. One limitation of the proposed model is that the value of parameter β is same throughout this paper. It may obtain better segmentation results for different images using different values of β .

Acknowledgment This work is partially supported by the National Natural Science Foundation of China (No. 61303126), and the Applied Basic Research Program of Sichuan Province (No. 2014JY0168) and the Scientific Research Foundation of the Education Department of Sichuan Province (No. 14ZA0178) and Foundation of Chengdu University of Information Technology (No. KYTZ201426) and (No. KYTZ201419).

References

1. McLachlan G, Peel D (2000) Finite mixture models. Wiley, New York
2. Richard S (2010) Computer vision: algorithms and applications. Springer, Berlin
3. Thanh MN, Wu QM, Ahuja S (2010) An extension of the standard mixture model for image segmentation. *IEEE Trans Neural NetW* 21:1326–1338
4. Sanjay GS, Hebert TJ (1998) Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. *IEEE Trans Image Process* 7:1014–1028
5. Geman S (1984) Geman D Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6:721–741
6. Besag J (1986) On the statistical analysis of dirty pictures. *J R Stat Soc Ser B* 48(3):259–302
7. Li SZ (2008) Markov random field modeling in image analysis, 3rd edn. Springer, Berlin
8. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans Med Imaging* 20:45–57
9. Blekas K, Likas A, Galatsanos N, Lagaris I (2005) A spatially constrained mixture model for image segmentation. *IEEE Trans Neural Netw*
10. Nikou C, Galatsanos N, Likas A (2007) A class-adaptive spatially variant mixture model for image segmentation. *IEEE Trans Image Process*
11. Nguyen TM, Jonathan Wu QM (2013) Fast and robust spatially constrained gaussian mixture model for image segmentation. *IEEE Trans Circ Syst Vid*, pp 621–635
12. Dempster P, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via EM algorithm. *J Stat Soc* 39:1–38
13. McLachian GJ, Krishnan K (2008) The EM algorithm and extensions. Wiley-Interscience, New Jersey
14. Bishop C (2006) Pattern recognition and machine learning. Springer, Berlin
15. Scherrer B, Forbes F, Garbay C (2009) Distributed local MRF models for tissue and structure brain segmentation. *IEEE Trans Med Imaging* 28(8):1278–1295
16. Choi HS, Haynor DR, Kim Y (1991) Partial volume tissue classification of multichannel magnetic resonance images: A mixed model. *IEEE Trans Med Imag* 10(3):395–407
17. Unnikrishnan R, Pantofaru C, Hebert M (2007) Toward objective evaluation of image segmentation algorithms. *IEEE Trans Pattern Anal Mach Intell* 29:929–944
18. Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. *IEEE Trans Pattern Anal Mach Intell* 33(5):898–916

Community Clustering Based on Weighted Informative Graph

Yi Xu, Yingning Gao and Weimin Li

Abstract Community clustering means the vertices in networks are often used to cluster into tightly-knit group with a high density of within-group edges and a lower density of between-group edges. However, most community clustering algorithms do not involve the node attributes and relationship, and these approaches lead to inaccuracy clustering. In this paper, we propose two algorithms which involve both node attributes and link structure in social networks based on Girvan-Newman algorithm (GN) and Weighted Informative Graph (WIG). The related experimental results verify the effectiveness of our proposed algorithms.

Keywords Node attributes · Relationship · Weighted informative graph (WIG) · Clustering algorithm

1 Introduction

In recent years, with the rapid development of Internet technology, the social networking service has become a new hot spot. Many social networking websites have collected a lot of communication dataset of users, which urgently need effective methods to analyze and mine. In order to mine more valuable hidden

Y. Xu · Y. Gao · W. Li (✉)

School of Computer Engineering and Technology, Shanghai University,
Shanghai, China
e-mail: wml@shu.edu.cn

Y. Xu
e-mail: fraudxyiy@gmail.com

Y. Gao
e-mail: gaoyingning512@gmail.com

W. Li
Shanghai Key Laboratory of Computer Software Evaluating and Testing,
Shanghai, China

knowledge from social networks, clustering algorithm will be attached more attention.

The traditional clustering algorithm mainly includes three categories. (a) Optimization based algorithms, such as Spectral methods, Kernighan-Lin Algorithm [1], Fast Newman Algorithm [2] and CNM Algorithm [3] and so on. This category needs to define the objective function so that it can be converted into the problem of optimization. (b) Heuristic Algorithm, such as GN Algorithm whose heuristic rule is that edge betweenness in the same community is more than the edge betweenness between two communities. (c) Other community clustering algorithm, such as Random Walk based Similarity [4], SCAN algorithm based on edge density and Clustering Centrality [5].

On the basis of the existing community network clustering algorithms and study on algorithms which combine object attribute values and link structure, two kinds of improved algorithms are proposed mainly based on following aspects.

- (1) Measuring the community clustering quality by the modularity Q through greedy strategy and local optimization is a significant community clustering way. It is mainly based on two kinds of thoughts. (a) The optimization bases on the value of objective function, such as CNM algorithm. (b) Clustering the network community through the hierarchical clustering algorithm or others, measuring the community clustering quality by the modularity is one of the good clustering methods, such as GN Algorithm. After considering these two points, we design a new index to measure the quality of clusters.
- (2) Currently, most of the clustering algorithms separate the node attributes and relationship to cluster the complex graph. But in fact, both node attributes and relationship affect the result of clustering. Therefore, the Weighted Informative Graph is introduced to solve the problem of merging the attributes and relationship.

The second section of this paper will introduce two proposed algorithms in detail; the forth will evaluate the experiment and analysis; the last part will conclude the algorithms and state the outlook for future work.

2 Community Clustering Based on WIG Algorithm

2.1 Preprocessing

Informative graph Informative graph is a representation of a graph containing both node attributes and relationship. Suppose a node has m attributes, then a node can be stood for an m -dimensional vector. Suppose an undirected graph containing n nodes is $G = (V, E)$, $V = V_1 \cup \dots \cup V_n$, and the attributes belong to node $V_i (1 \leq i \leq n)$ are represented by vector $A(V_i) = (a_{i1}, \dots, a_{im})$. So, the informative graph can be expressed as $IG = (V, E, AS)$, $\forall A(V_i) \in AS$.

In this way, the attributes of n nodes can be expressed by a $n \times m$ matrix. Then, a $n \times n$ similarity matrix can be got according to similarity measurement [6]. We use Jaccard coefficient $sim(i, j) = \sum_{k=1}^m a_{ik}a_{jk} / (\sum_{k=1}^m a_{ik}^2 + \sum_{k=1}^m a_{jk}^2 - \sum_{k=1}^m a_{ik}a_{jk})$, where $sim(i, j)$ stands for the similarity between node i and node j . In social network if node i and j are adjacent nodes, we convert $sim(i, j)$ to the weight w of the edge, and if not, then adjacent nodes' similarity is transformed into the correlation degree d .

$$\begin{cases} w_{ij} = W(sim(i, j)), e(i, j) \in E \\ d_{ij} = D(sim(i, j)), e(i, j) \notin E \end{cases} \quad (1)$$

WIG Through formula (1), we convert the node attributes into the cost of the edge and the correlation degree of nonadjacent points. As of now, we get the WIG of the network. As for different network, we can choose different function W and D .

Then, the steps of the preprocessing are listed as follows:

1. For a connected graph with n vertices, calculate each pair of the similarity and get the similarity matrix.
2. If $e(i, j) \in E$, the weight of the edge is $w_{ij} = 1 - sim(i, j)$, else $d_{ij} = 1 - sim(i, j)$.

After all above are done, we get the WIG. That is to say, we finish our preprocessing.

2.2 GN Algorithm Based on User's Social Information (GN-USI)

The GN-USI algorithm's steps for community detection are summarized below:

1. Get the WIG.
2. Calculate the betweenness of all existing edges in the network.
3. Remove the edge with highest betweenness.
4. Recalculate the betweenness of all edges affected by the removal.
5. Steps 3 and 4 repeated until for any subgraph G_i satisfies $M(G_i) \leq \varepsilon$.

Edge Betweenness Edge betweenness of an edge is the number of shortest paths between pairs of nodes that run along it [7]. If there is more than one shortest path between a pair of nodes, each path is assigned equal weight such that the total weight of all the paths is equal to unity. Here, we use e_{uv} to stand for the edge with a pair of nodes u and v , and $g_{ij} = g_{ji}$ means the number of shortest path between node i and j . Then, we define $g_{ij}(e_{uv})$ as the number of shortest path through e_{uv} .

$$C_B(e_{uv}) = \sum_{i \in V} \sum_{j \in V} \frac{g_{ij}(e_{uv})}{g_{ij}}, i \neq u \neq v \neq j, i < j \quad (2)$$

In GN-USI, the cost of shortest path is calculated by the node dissimilarity, so in WIG, the function W is $w_{ij} = 1 - sim(i, j)$. That is to say, if two nodes' similarity is higher, the weight of the edge is lower.

CMD In this paper, we define the Cluster Mixed Degree (CMD) because both the attributes difference and relationship have much impact on the result of clustering.

$$M(G) = \frac{1}{\log_x(1 - \theta)} \cdot \frac{\sum_{i \in V} \sum_{j \in V} (1 - sim(i, j))}{|e'|}, e(i, j) \notin E, 1 - \theta < \alpha < 1 \quad (3)$$

where G stands for the undirected graph, the CMD of the graph G is $M(G)$, θ means the density of the graph and $\theta = 2e/(n * (n - 1))$, and $|e'|$ expresses the pair of the non-adjacent nodes, $sim(i, j)$ is the similarity between node i and j . Obviously, the smaller $M(G)$ is, the purer cluster G is. If there is no edge e satisfy $e(i, j) \notin E$ that means G is a complete graph, $M(G)$ equals to 0. Then, we will introduce the process of GN-USI in the form of pseudo code (Tables 1 and 2).

Table 1 Main process of GN-USI

Input: 1.Cluster Mixed Degree as threshold value \mathcal{E}	2. WIG
Output: The set of the group called as ResultGroupSet	

```

1: ResultGroupSet =  $\emptyset$ 
2: CutEdgeSet = GetCutEdgeSet( $G'$ )
3: SubgraphSet= GetSubgraphSetByCutEdges( $G'$ , CutEdgeSet)
4: While SubgraphSet !=  $\emptyset$  do
5:    $G_i(V_i, E_i) = \text{SubgraphSet}.GetOneSubgraph()$ 
6:   if  $M(G_i) \leq \mathcal{E}$  then
7:     ResultGroupSet.Insert( $G_i$ )
8:     SubgraphSet.Delete( $G_i$ )
9:   else
10:    CutEdgeSet = GetCutEdgeSet( $G_i$ )
11:    SubgraphSet = SubgraphSet  $\cup$  GetSubgraphSet
        ByCutEdges( $G_i$ , CutEdgeSet)
12:  end if
13: end while
14: output ResultGroups  $G_1', \dots, G_k'$  in ResultGroupSet

```

Table 2 GetCutEdgeSet of GN-USI

Input: Connected graph $G(V,E)$
Output: CutEdgeSet based on the edge betweenness

```

1: CutEdgeSet =  $\varnothing$ 
2: while  $G - \text{CutEdgeSet}$  is connected do
3:   Calculate the weighted shortest paths between all
   pairwise nodes
4:   for each  $e_i \in E$  do
5:     GetEdgeBetweenness( $e_i$ )
6:   end for
7:    $\text{maxEdge} = \text{MAX}_{\text{btw}}(E)$ 
8:    $\text{CutEdgeSet} = \text{CutEdgeSet} \cup \text{maxEdge}$ 
9: end while

```

Table 3 GetSubgraphSetByCutEdges of GN-USI

Input: Connected graph G and CutEdgeSet
Output: SubgraphSet

```

1: SubgraphSet =  $\varnothing$ 
2:  $G = G - \text{CutEdgeSet}$ 
3: while  $G \neq \varnothing$  do
4:    $n' = \text{any node in } G$ 
5:   Find a connected subgraph  $G'$  using the BFS approach
   from the start node  $n'$ 
6:    $G = G - G'$ 
7:    $\text{SubgraphSet.Insert}(G')$ 
8: end while

```

The third part is about removing the CutEdgeSet from graph G and getting SubgraphSet (Table 3).

Through the three parts, we can cluster the graph into some groups. The time complexity of GN-USI is nearly equal to $O(nmm' + m'n^2 \log n)$, m' is the number of the edge removed. And the space complexity is $O(n + m + n^2)$ because we need to store the whole graph which cost $O(n + m)$ and similarity matrix which costs $O(n^2)$.

2.3 GN Algorithm Based on Cut Edge Coefficient (GN-CEC)

After analysis the time complexity of the GN-USI, we found the high time complexity caused by the step of calculating edge betweenness. To solve this problem, we define cut edge coefficient to replace the edge betweenness. Cut edge coefficient is different from the edge betweenness, it only takes the edge of the local information in account to reduce the computational complexity. The cut edge coefficient of the edge e_{uv} is listed as below.

$$C(e_{uv}) = \frac{\sum_{z \in \Gamma(u)} \beta \text{sim}(z, v) + \sum_{z \in \Gamma(v)} \beta \text{sim}(z, u)}{\sum_{z \in \Gamma(u) \cup \Gamma(v)} \beta \text{sim}(z, v) + \sum_{z \in \Gamma(u) \cup \Gamma(v)} \beta \text{sim}(z, u)} \quad (4)$$

where $C(e_{uv})$ is the edge coefficient between node u and v , $\Gamma(u)$ means the set of adjacent nodes except node v , $\text{sim}(u, v)$ is the similarity between node u and v , β is a coefficient, $\begin{cases} \beta = \mu_1, e(i, j) \in E \\ \beta = \mu_2, e(i, j) \notin E \end{cases}$. So, cut edge coefficient only need to take the similarity with nodes which are adjacent into account.

The main process of GN-CEC is the same as the main process in GN-USI. That is to say, the algorithm 1 and 3 can be used directly without modification (Table 4).

After calculating cut edge coefficient, we will describe the process of GetCutEdgeSet. In the process of this algorithm, removal of verge edge is not allowed for resulting in the result of group containing a lot of isolated nodes (Table 5).

From the process listed above, the time complexity of calculating all edge's cut edge coefficient is $O(tm)$, t is approximately equal to twice average nodes degree. It can be concluded that $O(tm)$ is far less than $O(nm + n^2 \log n)$ in calculating the edge betweenness. And another advantage is that after removing an edge, GN-CEC do not need to calculate all the edge again, which is better than GN-USI. In the following operation, the time complexity of calculating cut edge coefficient is approximately $O(t)$. Therefore, the time complexity for GN-CEC is $O(tm + tm')$. As for space complexity, in order to store all edge's cut edge coefficient, we need $O(n^2)$ space more. So the space complexity for GN-CEC is $O(n + m + 2n^2)$.

3 Evaluation Experiment and Analysis

Our research is based on social network services and we choose facebook and Sina Micro-blog as our data source. To avoid invalid experiment, we remove the isolated nodes in the preprocessing.

In order to visualize the experiment results, we extract a network containing 50 users and choose the province, city, and user descriptions as the user's attributes from Sina Micro-blog (Table 6).

Table 4 Calculate the cut edge coefficient of the edge in a graph

Input:	1. Connected graph $G(V,E)$	2. edge e to be calculated
	3. the coefficient μ_1 and μ_2	
Output:	Cut edge coefficient	


```

1: node i and j are two vertices of edge e
2: adjNodeiSet = all adjacent nodes of node i without j
   in G
3: adjNodejSet = all adjacent nodes of node j without i
   in G
4: sumOfDiffSet = 0
5: sumOfSameSet = 0
6: for each  $n_i \in \text{adjNodeiSet}$  do
7:   if  $n_i \in \text{adjNodejSet}$  then
8:     sumOfDiffSet +=  $\mu_1 \cdot \text{sim}(n_i, \text{node } j)$ 
9:   else
10:    sumOfDiffSet +=  $\mu_2 \cdot \text{sim}(n_i, \text{node } j)$ 
11:    sumOfSameSet +=  $\mu_1 \cdot \text{sim}(n_i, \text{node } i)$ 
12:   end if
13: end for
14: for each  $n_i \in \text{adjNodejSet}$  do
15:   if  $n_i \in \text{adjNodeiSet}$  then
16:    sumOfDiffSet +=  $\mu_1 \cdot \text{sim}(n_i, \text{node } i)$ 
17:   else
18:    sumOfDiffSet +=  $\mu_2 \cdot \text{sim}(n_j, \text{node } i)$ 
19:    sumOfSameSet +=  $\mu_1 \cdot \text{sim}(n_j, \text{node } j)$ 
20:   end if
21: end for
22:  $C(e_{ij}) = \text{sumOfDiffSet} / (\text{sumOfDiffSet} + \text{sumOfSameSet})$ 

```

Because users in social network services don't contain the information which organization they belong to, result assessment can only through the intuitive evaluation, but not quantitative evaluation. In order to evaluate accuracy, the result of clusters should be the only result. On the basis of this, we choose the co authorship network data. In this network, we extract 181 nodes and 574 edges. Each node stands for an author. And in order to control the scale of the network, we add

Table 5 GetCutEdgeSet of GN-CEC

Input: Connected graph $G(V,E)$
Output: CutEdgeSet based on the cut edge coefficient

```

1: CutEdgeSet =  $\emptyset$ 
2: while G is connected do
3:   minCutEdgeCoefficient = 1
4:   for each  $e_i \in E$  do
5:     if  $C(e_i) < \text{minCutEdgeCoefficient} \ \&\& \ !\text{IsVer}$ 
       geEdge( $e_i$ ) then
6:       minCutEdgeCoefficient =  $C(e_i)$ 
7:       minEdge =  $e_i$ 
8:     end if
9:   end for
10:  CutEdgeSet = CutEdgeSet  $\cup$  minEdge
11:  G = G - minEdge
12:  node i and j are two vertices of minEdge
13:  reCalculateCutEdgeCoefficientEdgeSet = all edges
       have node i or j
14:  for each  $e_i \in \text{reCalculateCutEdgeCoefficientEdgeSet}$ 
15:     $C(e_i) = \text{CalculateCutEdgeCoefficient}(G, e_i, \mu_1, \mu_2)$ 
16:  end for
17: end while

```

Table 6 The algorithm results of different threshold value ε

	GN-USI		GN-CEC	
	$\varepsilon = 0.25$	$\varepsilon = 0.05$	$\varepsilon = 0.25$	$\varepsilon = 0.05$
Number of the clusters	3	17	27	7
The average nodes in a cluster	3	17	29	8
The maximum nodes in a cluster	6	8	13	5
The minimum nodes in a cluster	6	8	14	5

an edge when there is more twice cooperation between two authors. Meanwhile, we choose major, school and research direction as our node attributes. Because of the scale is small, we classify the nodes in an artificial way. The number of four categories is 60, 45, 36, 40.

In order to evaluate the accuracy, we adjust the termination condition in the three algorithms. From Figs. 1, and 2, we can know the accuracy of GN is 68.5 %, the accuracy of GN-USI is 77.9 %, but GN-CEC is 60.2 %. Compare GN with GN-USI, we can conclude that merging the node attribute and relationship will get a higher accuracy.

Then, we extract dataset from facebook dataset and Sina Micro-Blog dataset. The scale varies from 100 to 1000. The network is randomly generated for because we just want to compare the efficiency of the algorithm. Because the number of iterations is different in different algorithm, we choose run time in a turn to compare. Figure 3 shows the result of the experiment. From the figure, it verifies our calculation of time complexity. The time complexity of GN-USI is higher than GN. And the efficiency of GN-CEC is higher than GN.

Fig. 1 The recall ratio of four clusters

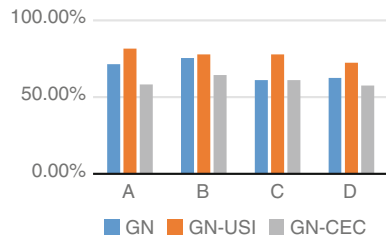


Fig. 2 The accuracy of these three algorithms

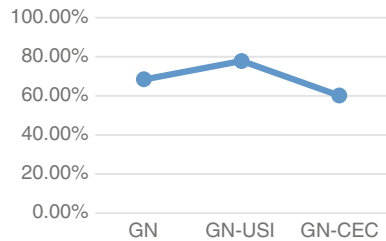
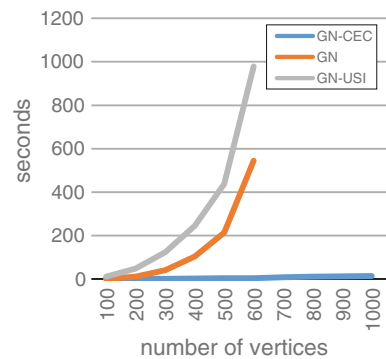


Fig. 3 The average run time in a turn



4 Conclusion

In order to merge the node attributes and relationship, WIG is proposed. It solves the problem effectively. Next, we define CMD to measure the purity of clusters. The index takes both attribute differences and relationship sparseness into account. Finally, on the foundation of the WIG and CMD, we improve GN algorithm and propose GN-USI which achieve a higher percentage of accuracy but the a little higher complexity. As for the complexity, it's not suitable for the large dataset. Therefore, we propose GN-CEC. The experiment proves that the GN-CEC has lower complexity, but also lower accuracy. It is valuable for dealing with the large dataset.

There also exist some improvements for these two algorithms, such as the complexity of calculating the betweenness of edge, how to calculate the similarity effectively and so on. In our future work, the thought of WIG will be used in other algorithms. And the clustering algorithms will have a higher percentage of accuracy.

References

1. Kernighan BW, Lin S (1970) An efficient heuristic procedure for partitioning graphs. *Bell Syst Tech J* 49(2):291–307
2. Newman MEJ (2004) Fast Algorithm for detecting community structure in networks. *Phys Rev E* 69:066133
3. Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large network. *Phys Rev E* 70(6):066111
4. Pascal P, Matthieu L (2006) Computing communities in large networks using random walks. *J Graph Algorithms Appl* 10(2):191–218
5. Xu X, Yuruk N, Feng Z et al (2007) SCAN: a structural clustering algorithm for networks. *KDD 2007*, pp 824–833
6. Dang TA, Viennet E (2012) Community detection based on structural and attribute similarities. 978-1-61208-176
7. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *PNAS* 99(12):7821–7826

A Data Clustering Algorithm Using Cuckoo Search

Mingru Zhao, Hengliang Tang, Jian Guo and Yuan Sun

Abstract In this paper, we present a novel algorithm for performing k-means clustering using cuckoo search. A pending problem of K-Means clustering algorithm is that the performance is affected by the original cluster centers. In this paper the K-Means algorithm is improved by cuckoo search and the initial cluster centers are generated by cuckoo search. The experiments and comparisons with the classical K-Means algorithm indicate that the improved k-mean clustering algorithm has obvious advantages on execution time.

Keywords Cuckoo search algorithm · K-Means · Clustering · Levy flight

1 Introduction

Clustering is the process of partitioning or grouping a given set of patterns into different clusters. This is done such that patterns in the same cluster are alike and patterns belonging to two different clusters are different. Clustering is a main task of exploratory data mining, and a common technique used in neural networks, AI, and statistics. Different algorithms for clustering have been proposed. K-means algorithm and its different variations are among those algorithms. The k-means method has been shown to be effective in producing good clustering results for many practical applications.

M. Zhao (✉) · H. Tang
Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology,
College of Computer Science and Technology, Beijing University of Technology,
Beijing 100124, China
e-mail: zhaomingru@126.com

M. Zhao · H. Tang · J. Guo · Y. Sun
Beijing Key Laboratory of Intelligent Logistics System,
Beijing Wuzi University, Beijing 101149, China

The k-means algorithm is well known for its efficiency in clustering large data sets and K-Means partitions data items into k clusters with each cluster which is represented by a single center point. K-Means clustering algorithm groups data items into a predefined number of clusters, based on Euclidean distance as similarity measure. The purpose of K-Means is to find k cluster centers. It starts with a random initial cluster centers and keeps reassigning the data items in the dataset to cluster centers based on the similarity between the data object and the cluster center. The reassignment procedure will not stop until a convergence criterion is met, for instance, the algorithm reaches the fixed iteration number or the cluster result does not change after a certain number of iterations [1, 2].

However, a direct algorithm of k-means method requires time proportional to the product of number of patterns and number of clusters per iteration. This is computationally very expensive especially for large datasets. It is necessary to employ some other global optimal searching algorithm for generating these initial cluster centers.

The cuckoo search (CS) algorithm is a biologically-inspired algorithm motivated by a social analogy that can be used to find an optimal, or near optimal solution to an optimization problem [3]. The CS algorithm can be used to generate good initial cluster centers for the K-Means. Therefore, in order to improve the efficiency of K-Means algorithm and reduce the impact of initial center points, this paper proposed a modified K-Means algorithm based on CS. The experiments show that the new algorithm is more efficient than K-Means.

2 Overview of Cuckoo Search Algorithm

The cuckoo search algorithm is an evolutionary algorithm first introduced by Yang and Deb (2009), inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds. It is a population-based search procedure used as an optimization tool, in solving complex optimization problems. Cuckoos lay their eggs in the nests of other host birds with incredible abilities like selecting the recently spawned nests and eliminating existing eggs that enhance hatching probability of their eggs. The host bird takes care of the eggs presuming that the eggs are its own. However, some of host birds are able to combat with this parasitic behavior of Cuckoos, and throw out the identified alien eggs or build their new nests in new locations. Each egg in a nest represents a solution, and a Cuckoo's egg represents a new solution.

When generating a new solution Levy flight is performed. During the search process, there are mainly three principle rules. The first rule is that each time, each cuckoo can only lay one egg (solution), which will be dumped in a randomly chosen nest. The second rule is that the best nests with the best eggs will be retained for the next generations. The third rule is that, during the whole search process, the number of available host nests is a constant number, and the host bird will find the

egg laid by a cuckoo with a probability. When it happens, the laid egg will be thrown away or the host bird will abandon the nest to build a new nest.

CS algorithm is a simple algorithm, and its code is given in 2010 by Yang and Deb. The initial population of the nests with the size n , which are generated first and they are randomly distributed over the search space. The randomly chosen initial solutions of design variables are defined in the search space by the lower and upper boundaries [4].

The new nest, for example n -th, is generated according to the following law:

$$X_i^{t+1} = X_i^t + \alpha \oplus Levy \lambda \quad (1)$$

CS algorithm is a simple algorithm, and its code is given in 2010 by Yang and Deb. The initial population of the nests with the size n , which are generated first and they are randomly distributed over the search space. The randomly chosen initial solutions of design variables are defined in the search space by the lower and upper boundaries.

Where α is the step size whose value depends on the optimization problem, and t is the current generation. Step size is multiplied by the random numbers with levy distribution, and such random motion is called levy flight.

A levy flight in which the step-lengths are distributed according to the following probability distribution:

$$Levy \lambda \sim u = t^{-\lambda}, (1 \leq \lambda \leq 3) \quad (2)$$

Accordingly, the consecutive jumps of a cuckoo form a random walk process obeying a power law step-length distribution with a heavy tail. In this way, the process of generating new solutions can be viewed as a stochastic equation for a random walk which also forms a Markov chain whose next location only depends on the current location and the transition probability. Note that in the real world; a cuckoo's egg is more difficult to be found the more similar it is to a host's eggs. So the fitness is related to the difference, and that is the main reason for using a random walk in a biased way with random step sizes [5, 6].

3 CS Clustering

Data clustering is the one of NP problems. CS algorithm is an effective technique for solving optimization problems that works based on probability rules and population. So it is feasible to solve clustering problem using CS.

In this paper, it evaluates the fitness of particles with distance measures, defined as (3). The fitness function (3) depicts the sum of all the intra-cluster distance, in which lower is better.

$$J_{cc} = \sum_{t=1}^K \sum_{x_i \in U_t} d(x_i, c_t) \quad (3)$$

It is assumed that $O = \{o_1, o_2, o_3, \dots, o_s\} \subset R^d$ is the data to be clustered, where O_i is d -dimensional. If the data is clustered into k clusters, then $\{c_1, c_2, \dots, c_k\}$ are used for the cluster centers. The data item O is divided into k clusters and $O = \bigcup_{t=1}^k C_t$, where C_t is a cluster.

J_{cc} is the sum of distance between each data in the clusters and the corresponding cluster center. In the algorithm, each particle maintain its' own clusters centers and the fitness function is to be minimized. In the above functions, $d(x_i, c_t)$ represents the distance between the data item and the cluster center and $d(x_i, c_t)$ is Euclidean distance thus $d(x_i, c_t) = \|x_i - c_t\|$.

The main idea of the K-Means algorithm based on CS is as follows: first, the number of clusters is determined. Second, each egg maintains k cluster centers, and each egg provides a clustering solution. Third, in the iteration, CS algorithm adjusts the location of k clusters centers to minimize the fitness function. The stopping criterion for iteration is that fitness function value is below a threshold value or the maximum iteration is reached.

4 Simulation

The improved K-Means algorithm based on CS is written using Matlab running at the windows operating systems and the memory is 512 M and the CPU frequency is 2.0 GHz. The paper compares the improved K-Means algorithm and the normal K-Means algorithm at the same environment in the respects of execution time. The two algorithms are both applied to the same data items. For comparative experiment, 100 two—dimensional points are generated randomly in this paper, which are shown in Fig. 1.

In this paper, both the two clustering algorithm are used to cluster these 100 two-dimensional points into three clusters by setting the value of $k = 4$. They returned the same cluster results, which are shown in Fig. 2.

According to the experiment, the execution time of the new algorithm is far below the K-Means algorithm appreciably. The results of the experiments are shown in Fig. 3.

Fig. 1 The 100 random two-dimensional points

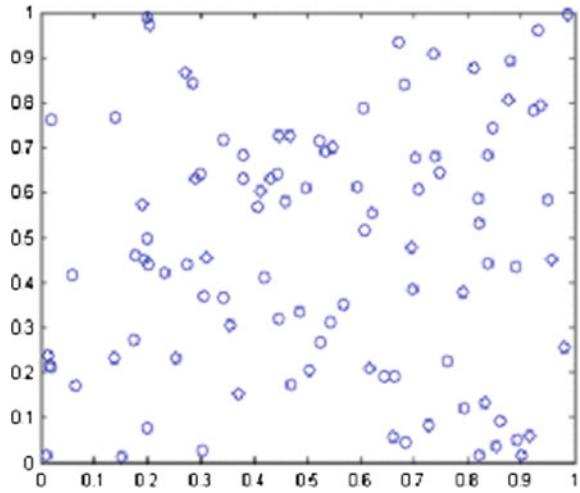


Fig. 2 Clustering results

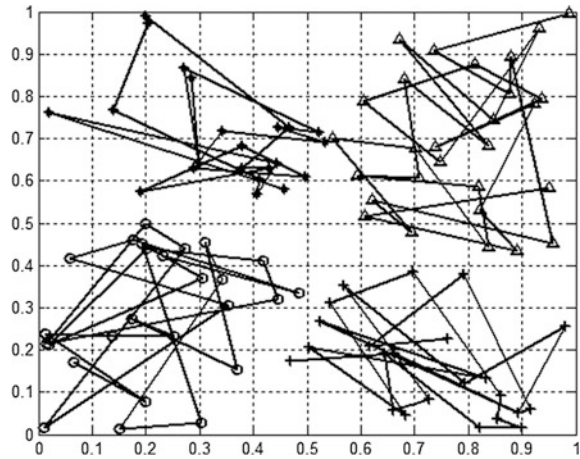
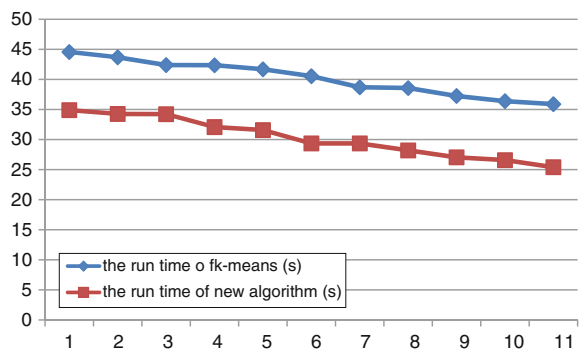


Fig. 3 Run time of two algorithms



5 Conclusions

In this paper, the modified K-Means algorithm has been achieved based on CS, which is a new evolutionary computation technique inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds. The new clustering algorithm and the K-Means have been both used to 100 two-dimensional points data clustering and returned the same results. According to the experiments, the new clustering algorithm using CS is better than the normal K-Means algorithm on the respect of execution time.

Acknowledgments This work was supported by Beijing Higher Education Young Elite Teacher Project (YETP1532); Beijing Excellent Talents funded projects (2013D005009000003). The outstanding talents project supported by Beijing municipal Party Committee Organization Department (No.2013D005009000003).

References

1. Yang XS, Deb S, Karamanoglu M, He X (2012) Cuckoo search for business optimization applications. In: National conference on computing and communication systems (NCCCS), vol 7331, IEEE, pp 1–5
2. Yin XLM (2013) A hybrid cuckoo search via Lévy flights for the permutation flow shop scheduling problem. *Int J Prod Res* 51(16):4732–4754
3. Valian EVE (2013) A cuckoo search algorithm by Lévy flights for solving reliability redundancy allocation problems. *Eng Optim* 45(11):1273–1286
4. Jamil M, Zepernick HJ (2013) Multimodal function optimisation with cuckoo search algorithm. *Int Bio-Inspired Comput* 5(2):73–83
5. Srivastava PR, Sravya C, Ashima S et al (2012) Test sequence optimisation: an intelligent approach via cuckoo search. *Int Bio Inspired Comput* 4(3):139–148
6. Feng Y, Ke J, He Y (2014) An improved hybrid encoding cuckoo search algorithm for 0-1 knapsack problems. *Comput Intell Neurosci* 2014(1):119–150

The Application of Bacteria Swarm Optimization Algorithm in Site Choice of Logistics Center

Mingru Zhao, Hengliang Tang, Jian Guo and Yuan Sun

Abstract A new setting up logistics distribution centers algorithm based on Bacterial Foraging Optimization is proposed in this paper. Logistics distribution centers are the bridges to connect the supplying points and the demanding points; therefore, how to set up the logistics distribution centers is the important problem of a logistics system. Firstly, the logistics centers location model is discussed and then a new setting up logistics distribution centers algorithm based on Bacterial Foraging Optimization is proposed in this paper. To solve discrete space problems, the chemotaxis procedure is modified in the new algorithm. The experiments show that, the proposed algorithm in this paper can return the solution of setting up logistics distribution centers problems.

Keywords Bacteria swarm optimization algorithm · Logistics distribution centers · Discrete space problems · Chemotaxis procedure

1 Introduction

The set choice of logistics center is the center of logistics system. Logistics center plays a big role in the whole logistics system and choosing the logistics center in a economic region with several supply points and several demanding points is very important. A series of location models and algorithms are established for the choosing logistics center problem. But the gravity model approach is mainly used to solve the simple problems of choosing logistics center and linear programming

M. Zhao (✉) · H. Tang
Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology,
College of Computer Science and Technology, Beijing University of Technology,
Beijing 100124, China
e-mail: zhaomingru@126.com

M. Zhao · H. Tang · J. Guo · Y. Sun
Beijing Key Laboratory of Intelligent Logistics System, Beijing Wuzi University,
Beijing 101149, China

technique is used to solve the complicated problems. But the choosing center found by gravity method is very difficult to implement and linear programming technique is strict to the target function. In this paper, a new method of choosing logistics center algorithm based on bacteria swarm optimization is proposed.

2 Overview of Bacteria Swarm Optimization Algorithm

The Bacterial Foraging Algorithm is a new approach of global search techniques and the core idea of the algorithm is to simulate the *E. coli* foraging behavior. Typical bacteria such as *E. coli* propel themselves from a place to a place by rotation of the flagella in one direction. The bacteria's flagella rotates counter clockwise and the bacteria "swims" to moves forward while a clockwise rotation of the flagellum causes the bacterium to randomly "tumbles" itself in a new direction. The microorganism such as the *E. coli* tumbles and swims in its entire lifetime. Alternation between "swim" and "tumble" enables the bacterium to search for nutrients in random directions and escape harmful substance. Ultimately, bacterial chemotaxis is a complex combination of swimming and tumbling that keeps bacteria in places of higher concentrations of nutrients and keeps them from harm.

The typical BFA consists of four procedure, namely, chemotaxis, swarming, reproduction, and elimination and dispersal. We briefly describe each of these processes one by one as follows [1–3].

2.1 Chemotaxis

The chemotaxis process in the Bacterial Foraging Algorithm is attained by swimming and tumbling. If the place is better than before the bacterium will move in a predefined direction (swimming) and otherwise the bacterium will move in an altogether different direction (tumbling). Hence, the bacterium performs the operation modes in its whole lifetime are that of swimming, tumbling or switching between running and tumbling. If the bacteria is places of higher concentrations of nutrients, it will spend more time swimming and less time tumbling [4].

2.2 Swarming

The bacterium moves following the nutrient gradient and the bacterium that has found the optimum path of food should try to attract other bacteria so that they reach the desired place more rapidly [5, 6]. Thus the bacteria release attractant aspartate which lead the bacteria to concentrate into groups hence more bacterial will move to where the place with more nutrient.

Mathematically, swarming process can be represented by

$$J_{cc}(X_i) = \sum_{i=1}^N J_{cc}^i = \sum_{i=1}^N J_a^i + \sum_{i=1}^N J_r^i \tag{1}$$

$$\sum_{i=1}^N J_a^i = \sum_{i=1}^N \left[-d_a \exp(-w_a \sum_{m=1}^v (x_m - x_{mi})^2) \right] \tag{2}$$

$$\sum_{i=1}^N J_r^i = \sum_{i=1}^N \left[h_r \exp(-w_r \sum_{m=1}^v (x_m - x_{mi})^2) \right] \tag{3}$$

where $J_{cc}(X_i)$ is the cost function value to be added to the actual cost function to be minimized to present a time varying cost function. In Chemotaxis process, the bacterial swims and tumbles in order to find the location where the cost function is more minimized [7].

“N” is the total number of bacteria. J_a presents attraction between bacterium, while J_r presents exclusion between bacterium and “v” is the number of parameters to be optimized that are present in each bacterium. There are different coefficients that are to be chosen judiciously such as d_a , w_a , h_r and w_r .

2.3 Reproduction

According to foraging theory, the weak bacterial will be sifted out and the powerful bacterial will remain in the natural selection. The least healthy bacteria will die, and the other healthiest bacteria each will split into two bacteria, which are placed in the same locations. This makes the constant population of bacteria [8].

2.4 Elimination and Dispersal

Because of the consumption of the nutrients, the lives of the bacteria may change either gradually or suddenly. Events may kill or disperse all the bacteria in a region and new bacteria will emerge [9]. Sudden events have the effect of possibly destroying the chemotactic progress, but in contrast, they also assist it, since dispersal may place bacteria near nutrient area. The elimination and dispersal progress may provide some random resolutions. Not all the bacteria are killed in the elimination and dispersal process, the bacteria is randomly killed, and the probability that the bacteria is killed is p_{ed} .

3 Logistics Distribution Center Location Model

In this paper, the distribution center location problem is selected according to the lowest logistics total cost. Total cost includes three parts: distribution transportation cost, management cost and construction cost.

The model is described as follows: there are m positions and choose p locations from the m positions to construct the distribution center providing goods for n allocations and making total cost lowest. The demand of the i th distribution is a_{ij} and the distance from the i th distribution to j th distribution center is d_{ij} and the distance from the factory to j th distribution center is s_j and k is freight rate, therefore the transportation cost of the j th distribution center is computed as follows:

$$f_j = \sum_{i=1}^n a_{ij} \times d_{ij} \times k + s_j \times \sum_{i=1}^n a_{ij} \times k \quad (4)$$

The construction cost of per unit of distribution center j is h_j , and the construction cost of j is:

$$g_j = \sum_{i=1}^n a_{ij} \times h_j \quad (5)$$

The management cost of per unit of distribution center j is e_j , and the management cost of j is:

$$l_j = \sum_{i=1}^n a_{ij} \times e_j \quad (6)$$

Distribution center location is chosen for the lowest total cost of the system, which is shown in formula 4.

$$\min E = \sum_{j=1}^m (f_j + g_j + l_j) \quad (7)$$

In this paper, the logistics center location instance involves n distribution centers and m distribution centers. The purpose of the location is to specify the distribution center for n distribution point, each distribution point from the m distribution center in a distribution center, the total cost of the system is the lowest. Therefore, the location problem of logistics center is $\theta = \{X_1, X_2, \dots, X_n\}$, $X_i \in \{1, 2, 3, \dots, M\}$, $I \in \{1, 2, 3, \dots, n\}$. The purpose of the optimization is to find an optimized numerical sequence $\{X_1, X_2, \dots, X_n\}$ and the total cost of the system is the lowest. The total cost of any numerical sequence can be obtained by using the formula 4, 5, 6, 7.

4 Improved Bacteria Swarm Optimization Algorithm

The logistics center location problem in this paper is discrete, so, it is necessary to make corresponding change the four procedures for discrete problems: chemotaxis, swarming, reproduction, and elimination and dispersal.

Chemotaxis process simulates the movement of an *E. coli* cell through swimming and tumbling. Biologically an *E. coli* bacterium can move in two different ways. It can swim for a period of time in the same direction or it may tumble, and alternate between these two modes for the entire lifetime. Suppose $\theta^i(j, k, l)$ represents i -th bacterium at j -th chemotactic, k -th reproductive and l -th elimination-dispersal step. $C(i)$ is the size of the step taken in the random direction specified by the tumble. Then in computational chemotaxis the movement of the bacterium may be represented by

$$\theta^i(j + 1, k, l) = \theta^i(j, k, l) + c(i) \frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \tag{8}$$

where Δ indicates a vector in the random direction whose elements lie in $[-1, 1]$. In this paper the solution of this problem is $\theta = \{X_1, X_2, \dots, X_n\}$, where $X_i \in \{1, 2, 3, \dots, m\}$, $i \in \{1, 2, 3, \dots, n\}$, so movement of the bacterium in this paper may be represented by

$$\theta^i(j + 1, k, l) = \text{abs}(\text{mod}(\theta^i(j, k, l) + c(i)\text{floor}(\text{rand}() * m + 1), m)) + 1 \tag{9}$$

where the functions in the formula: $\text{abs}()$, $\text{mod}()$, $\text{floor}()$ and $\text{rand}()$ are the built-in functions of MATLAB.

5 Data Set Description

There is a logistics distribution area, which is a square and its range (0, 0) to (100, 100). There are 10 candidate distribution centers, the distribution center coordinates, land prices, unit management cost are shown in Table 1; in the spatial distribution of scattered 30 distribution points, the distribution point coordinates and demand are shown in Table 2; factory coordinates are (50, 45) and the freight rate is 1.

Table 1 The information of candidate distribution centers

Distributions centers	1	2	3	4	5	6	7	8	9	10
Coordinates	(23, 74)	(80, 50)	(83, 90)	(45, 16)	(34, 67)	(95, 95)	(4, 47)	(56, 37)	(10, 10)	(37, 97)
Land price	1.5	1.3	1.0	1.6	1.4	0.9	1.7	1.4	2.0	1.1
Unit management cost	1.2	1.1	1.0	1.3	1.1	1.4	1.25	0.9	1.15	1.0

Table 2 The information of distributions

Distribution	1	2	3	4	5	6	7	8	9	10
Coordinates	(24, 68)	(40, 15)	(70, 60)	(92, 5)	(42, 64)	(87, 45)	(39, 7)	(28, 71)	(55, 20)	(9, 51)
Demands	1.5	0.6	0.5	2.0	0.7	1.4	1.0	2.0	3.0	1.0
Distribution	11	12	13	14	15	16	17	18	19	20
Coordinates	(13, 47)	(34, 78)	(5, 16)	(68, 45)	(15, 16)	(76, 35)	(8, 5)	(33, 62)	(20, 71)	(5, 50)
Demands	0.8	2.1	0.7	0.6	1.3	0.3	4.0	1.7	1.0	2.3
Distribution	21	22	23	24	25	26	27	28	29	30
Coordinates	(6, 46)	(13, 4)	(41, 5)	(68, 29)	(80, 63)	(50, 25)	(78, 47)	(30, 77)	(39, 69)	(7, 20)
Demands	1.2	0.4	1.0	1.1	0.9	0.3	0.4	2.5	1.1	2.0

Table 3 The information of candidate distribution centers

Distributions centers	1	2	3	4	5
Distributions	Null	4, 6, 25, 27	null	2, 7, 23	1, 5, 8, 12, 18, 19, 28, 29
Unit management cost	0	4.7	0	2.6	12.6
Distributions centers	6	7	8	9	10
Distributions	Null	10, 20, 21	3, 9, 11, 14, 15, 16, 24, 26, 30	13, 17, 22	Null
Unit management cost	0	4.5	9.9	5.15	0

6 Conclusions

The algorithm is written using Matlab running at the windows operating systems and the memory is 512 M and the CPU frequency is 2.0 GHz. After program run ten times, the best solution of the proposed algorithm based on bacteria swarm optimization is {5, 4, 8, 2, 5, 2, 4, 5, 8, 7, 8, 5, 9, 8, 8, 9, 5, 5, 7, 7, 9, 4, 8, 2, 8, 2, 5, 5, 8}, the total costs of the best solution is 1.8464e+03. The best solution for the distribution center location scheme is shown in Table 3.

Acknowledgments This work was supported by Beijing Higher Education Young Elite Teacher Project (YETP1532); Beijing Excellent Talents funded projects (2013D005009000003).

References

1. Tian YF, Zhang FY, Yan S (2012) Bacteria foraging optimization algorithm based on particle swarm optimization. *Control Eng China*
2. Wei LI, Wei XU (2013) An improved bacteria foraging optimization algorithm and its application in soft measurement modeling. *Transducer Microsyst Technol* 32(4):149–152
3. Ali ES, Abd-Elazim SM (2012) TCSC damping controller design based on bacteria foraging optimization algorithm for a multimachine power system. *Int J Electr Power Energy Syst* 37 (1):23–30
4. Jung SH (2011) Simple bacteria cooperative optimization with rank-based perturbation. In: *International proceedings of economics development and research*
5. Chu Y, Mi H, Ji Z et al (2010) Fast bacterial swarming algorithm based on particle swarm optimization. *J Data Acquisition Process* 25(4):442–448
6. Daryabeigi E, Moazzami M, Khodabakhshian A et al (2011) A new power system stabilizer design by using smart bacteria foraging algorithm. In: *24th Canadian conference on electrical and computer engineering (CCECE)*, IEEE, USA, pp 000713–000716
7. Chang C, Zhu Y, Hu K et al (2010) Research on smelting ingredient diluting for refined copper strip by bacteria foraging optimization algorithm. In: *International conference on digital manufacturing and automation (ICDMA)*, IEEE, USA, pp 275–278
8. Mo H, Liu L, Geng MA (2014) Magnetotactic bacteria algorithm based on power spectrum for optimization. *Lect Notes Comput Sci* 115–125
9. Gao LF, Zhang XC (2007) Study on logistics distribution center location based on max-min ant system. *Oper Manage* 16(6):42–46

SIDA: An Information Dispersal Based Encryption Algorithm

Zhi-ting Yu, Quan Qian, Rui Zhang and Che-Lun Hung

Abstract High performance encryption is a key means to minimize security risks as protecting private data in cloud or big data environments. In this paper, a new encryption model SIDA is proposed based on the information dispersal and multi-layer encryption. From theoretical analysis and experiments, it shows that SIDA is secure, can not only significantly improve the speed of data encryption and decryption, but also reduce the bandwidth consumption and re-encryption overhead when revoking authority. Taking SIDA4 algorithm as an example, the encryption speed is about 1.6 times of AES. While the overhead of re-encryption when revoking authority, SIDA4 in communication and computation are 1/4 of AES.

Keywords Information dispersal · Information re-encryption · Authority revocation

1 Introduction

With the rapid development of computer and Internet technology, data is increasing exponentially. The emergence of cloud computing gives a new solution that companies or private users can store their data in the cloud. However, when data stored in cloud, users will inevitably be worried about the data securities, i.e. whether the data will be accessed by unauthorized users [1].

Access control and data encryption are normally used to solve the above problems. Access control ensures that only the users who have the appropriate permissions can access the corresponding data. And for data encryption, only users

Z. Yu · Q. Qian (✉) · R. Zhang
School of Computer Engineering and Science, Shanghai University,
Shanghai 200444, China
e-mail: qqian@shu.edu.cn

C.-L. Hung
Department of Computer Science and Communication Engineering,
Providence University, Taichung City 43301, Taiwan

who have decryption keys can understand the data, which requires the users to encrypt their data before uploading to the cloud. However, under some circumstances, the computing, bandwidth consumption and authorization revocation cost for data encryption mechanism is very high. So, developing those encryption algorithms for cloud computing with high security and performance are very necessary.

Concerning about re-encryption, it can be divided into two categories: client-side re-encryption and server-side re-encryption. The client-side re-encryption is that the data owner retrieves the encrypted data, decrypts and re-encrypts it with new keys, and then uploads the newly encrypted data and updates the secret keys. While in server-side re-encryption, the remote server does the data re-encryption, but the server itself does not understand the data. Generally speaking, the client-side re-encryption is more secure, because in the server-side re-encryption, the server is considered to be partly trustworthy. But the bandwidth cost of server-side re-encryption is lower than that of the client-side.

The rest of the paper is organized as follows. Section 2 introduces the related research background. Section 3 presents the SIDA algorithm in detail. Relevant theoretical analysis for the performance and security of SIDA is shown in Sect. 4. Section 5 makes a conclusion and discusses the future work.

2 Related Work

About the client-side re-encryption, Bethencourt et al. proposed a ciphertext policy attribute based encryption algorithm (CP-ABE) [2]. The algorithm associates users private key with a set of property conditions, and if the attribute of a users private key meets the property conditions, then the user has the ability to decrypt the data. Compared with the direct key distribution, CP-ABE is easier to manage the secret keys. However, it is based on asymmetric encryption and the low efficiency limits its use for large scale data encryption. Hong Cheng et al. [3] used symmetric encryption to improve CP-ABE and got a new cryptographic access control scheme named AB-ACCS. When revoking authority, the outdated key must be destroyed and the data owners need to download all data, and then generate new keys to re-encryption them. In [4], Hong Cheng et al. improved the re-encryption model again. They left the secret key management work to the server. Although it reduced some costs, but the data re-encryption work still need to be executed on client.

About server-side re-encryption, there are following typical models. Proxy re-encryption, used in [5, 6], is a ciphertext (secret key) transformation technology and the data are encrypted by asymmetric encryption algorithms. When revoking authority, data owners do not need to download all data for re-encryption and they only need to transmit key transforming parameters to the cloud. And the server will use these parameters to transform the old ciphertext and do re-encryption. Meanwhile, the cloud server also transmits the parameters to users who still have the access authorities, and users can use them to update their own keys. However,

this model use asymmetric encryption, it is not suitable for huge data encryption. Vimercati et al. [7] proposed an over-encryption model. In this model, the data owner encrypts file with secret key $key1$ and transmit the ciphertext to sever, then the server use secret key $key2$ to encrypt the file again. Users who allowed to access the file will receive these two keys. When the data owner want to re-encrypt the file, the server only need to decrypt the file with $key2$, and then use a new key $key3$ to encrypt it and transmit $key3$ to the legal users. So, the re-encryption work has been done without any plaintext and the legal users can use $key1$ and $key3$ to decrypt data. The disadvantage of this method in that it needs two layers of encryption and time overhead is double. Both proxy re-encryption and over-encryption scheme have a prerequisite that the server is trustworthy.

Cepheus proposes a lazy re-encryption model [8]. This model does not re-encrypt the file immediately when authority revoked, and the re-encryption is done when the file contents are changed. As a information redundancy algorithms, information dispersal method [9, 10], proposed by Rabin [11], which can improve the data security and integrity in data transmission. In [12], Bian et al. uses information dispersal method, divides the data into pieces and scatters in mediums in different geographical location. Capturing a small amount of data slices is meaningless and when the data is not badly destroyed we can use the rest part to recover the complete data. Zhou et al. applies information dispersal method to HPC and cloud computing in [6], they compare the advantages and disadvantages of commonly used data replication and information dispersal method.

In order to reduce the overhead of bandwidth consumption and data re-encryption in authority revocation, we proposed a new encryption algorithm SIDA, which combines the advantages of IDA and the over-encryption algorithm.

3 Security Information Dispersal Algorithm

The algorithm is based on transformation of invertible matrix over finite field. In (m, n) -IDA, a file F is divided into n pieces $F_i (1 \leq i \leq n)$, and the length of each piece is $|F_i| = L/m$. Let $F = (b_1 \cdots b_m), (b_{m+1} \cdots b_{2m}), \cdots$, the characters b_i is an eight-bit byte with range $[0 \cdots 255]$. Take a prime $p > 255$, supposing $p = 257$, the value of $(b_i \% p)$ belongs to the finite field $GB(2^8)$.

Choosing an n -order invertible matrix G , every n -dimensional vector in G is $G_i = (G_{i1}, G_{i2}, \cdots, G_{in}) \in GB(2^8)$, and any m vectors are linearly independent. Choosing any m pieces from the total n pieces can reconstruct the original data while any $(m - 1)$ pieces cannot.

SIDA is based on Information Dispersal Algorithm (introduced in the second layer encryption in Sect. 3.1.4, and adversaries can't obtain the original data without decrypting the data encrypted by the second layer encryption), all byte operations in algorithm are over the finite field $GB(2^8)$.

3.1 Data Encryption Using SIDA

The data encryption process is shown in Fig. 1a. Each time read one data group from the original file, and then executes the first layer encryption with *key1*. After that, use matrix *G* for data transformation and *G* is an *n*-order invertible matrix without zeros in it. Then, execute the second layer encryption with *key2*, that is, encrypt the result of the former step and finally write the cipher text back into its data group. It is worth mentioning that the encryption algorithm used in the first and second layer can be different. For the sake of simplicity, supposing we use the same symmetric encryption algorithm. Figure 1b shows the process of the data transformation, including data extraction, matrix transformation and data reconstruction, which reflect the core idea of information dispersion. The pseudo code algorithm of SIDA encryption is as follows.

```

program SIDA_Encryption(
  F: plaintext file to encrypt; N: segment count in each data group;
  M: number of bytes in segment; G: invertible n-order matrix;
  key1: the secret key for the first layer encryption;
  key2: the secret key for the second layer encryption)
{
  While (f=read (F, M, N)) != NULL){
    //Read one data group from a plaintext file each time
    f=Encrypt1 (f, key1); //First layer encryption with key1
    for (i=0; i<M; i++) {
      Temp=Extract (f); //Extract N bytes from a data group
      //Transform the extracted data with G
      Temp2=Convert (Temp, G);
      //Put the extracted data back to the original location
      Reconstruct (Temp2);
    }//end for
    f=Encrypt2 (f, key2); //Second layer encryption with key2
    Write (f, F); //Write the encrypted data back to file F
  }//end while
  Return the encrypted file F;
} //end program

```

In the above algorithm, *Encrypt1*(\dots) and *Encrypt2*(\dots) are the first layer and second layer encryption respectively. *Extract*(\dots) is the process for extracting data. *Convert*(\dots) is a matrix transformation and *Reconstruct*(\dots) is the data reconstruction operation. For simplicity, in the rest of the paper, *E1*, *E2*, *Ex*, *Cvr* and *Re* are the abbreviations for function *Encrypt1*, *Encrypt2*, *Extract*, *Convert*, and *Reconstruct*.

The first layer encryption is to encrypt the first segment of the data group (1/N of the data group). The transformation operation first does data extracting, and then executes the matrix transformation for the extracted data. The goal of the

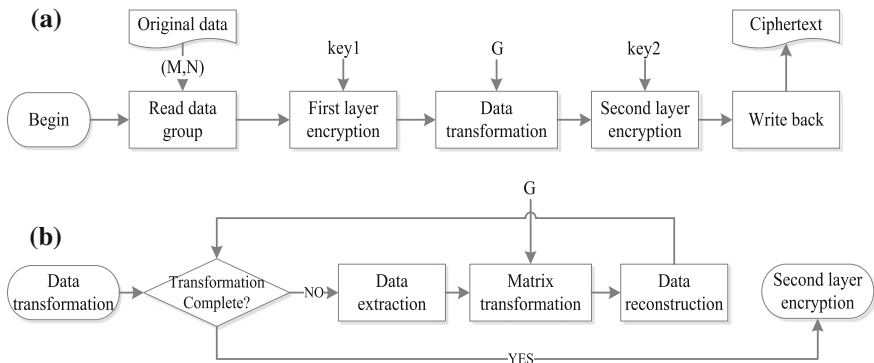
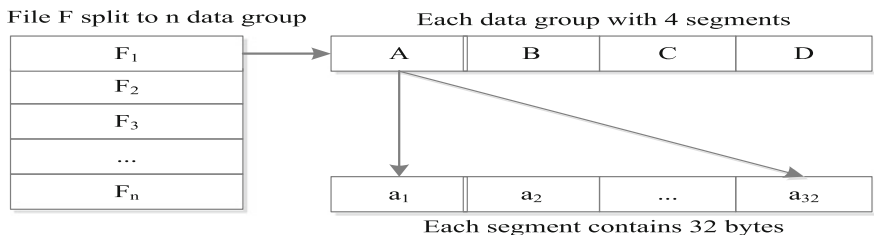


Fig. 1 a Flow chart of data encryption, b data transformation process

transformation operation is letting the $1/N$ (for SIDA4, $N = 4$ and for SIDA8, $N = 8$) encrypted data affects the rest part of data group. In this way, we obtain a cipher-text with good randomness. After this, we do the second encryption, that is, encrypt the first segment of the former result. So, we can see that only users who have both secret keys ($key1, key2$) can decrypt the data. Next, we will describe the main operations in SIDA in detail.

3.1.1 Data Partition

The data partition schemes are as follows: in SIDA4, each data group has four segments and each segment has 32 bytes; in SIDA8, each data group has eight segments and each segment also has 32 bytes. The program read one data group from file each time. If the length of the file is not an integral multiple of 128 (256 in SIDA8), padding some bytes after the tail of the file. The data structure of SIDA4 is shown in Fig. 2. The detailed analysis of the algorithm in rest of the paper will take SIDA4 as an example; and the SIDA8 scheme is similar to SIDA4.



F_i is the i -th data group of file F ; A,B,C and D are segments in each data group; a_i is the i -th byte of segment A.

Fig. 2 The data structure of SIDA4

3.1.2 The First Layer Encryption

In this layer, it will encrypt the first segment of each data group. We can describe it as Eqs. 1 and 2.

$$E1(F_i, key1) = E1(A, key1) + B + C + D \tag{1}$$

$$E1(A, key1) = \{E1(a_1), E1(a_2), \dots, E1(a_{32})\} \tag{2}$$

As mentioned above, $E1$ is the first layer encryption and the “+” operation in Eq. 1 indicates the concatenation operation.

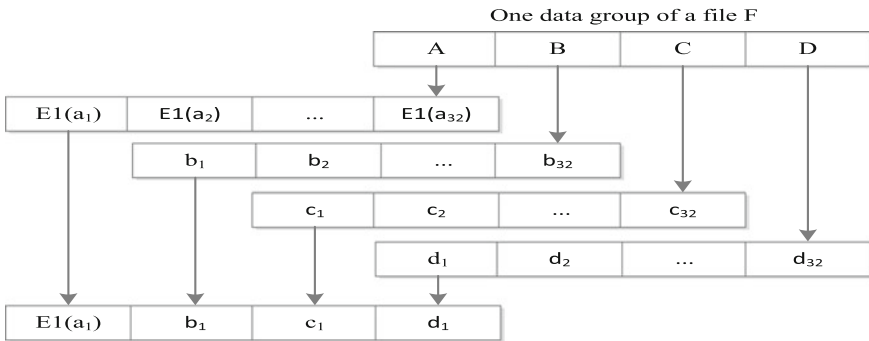
3.1.3 Data Transformation

- (1) **Data Extraction Process** Extract the j -th byte of each segment in data group and construct a four-byte data unit K_i (the size of each data unit is eight bytes in SIDA8). So, there are 32 data units after all data in a data group are extracted. The formal process can be described as Eqs. 3 and 4. Figure 3 shows the detailed process of Data Extraction.

$$Extract(F_i) = \{M_1, M_2, \dots, M_{32}\} \tag{3}$$

$$M_j = \{E1(a_j), b_j, c_j, d_j\} \tag{4}$$

where in Eqs. 3 and 4, M_j is the j -th unit formed in data extraction process which contains every j -th byte of the segment.



Extract i -th byte of each segment, i.e. $i=1$. A, B, C and D are segments of each data group; a_i , b_i , c_i and d_i are corresponding bytes of A, B, C and D. $E1(x)$: encrypt x with the first layer encryption algorithm.

Fig. 3 Data extraction process

- (2) **Matrix Transformation** Does matrix transform on extracted data and G is an n -order invertible matrix without 0. The transformation process can be depicted as Eqs. 5 and 6.

$$Convert(F_i, G) = G \times \{M_1, M_2, \dots, M_{32}\} \quad (5)$$

$$M_j = G \times M_j = G \times \begin{bmatrix} E1(a_j) \\ b_j \\ c_j \\ d_j \end{bmatrix} = \begin{bmatrix} a_j \\ b_j \\ c_j \\ d_j \end{bmatrix} \quad (6)$$

- (3) **Data Reconstruction** Put the 32 data units back to their original position in data group. And now A', B', C', D' are the new four segment of the data group.

$$Reconstruct(F_i) = Reconstruct\{M_1, M_2, \dots, M_{32}\} = \{A', B', C', D'\} \quad (7)$$

3.1.4 The Second Layer Encryption

Encrypt the first segment of the new data group with $key2$, which can be described as Eqs. 8 and 9. And in Eq. 8, “+” is also the concatenation operation.

$$E2(F_i, key2) = E2(A', key2) + B' + C' + D' \quad (8)$$

$$E2(A', key2) = \{E2(a_{1'}), E2(a_{2'}), \dots, E2(a_{32}')\} \quad (9)$$

After the second layer encryption, the first byte of all 32 data units is encrypted. So, concerning encryption time performance, it can be approximately represented as the sum of the time cost during the first layer encryption, the data transformation and the second layer encryption, which can be described as Eq. 10.

$$\begin{aligned} T_{encryption} &= T_{E1} + T_{transformation} + T_{E2} \\ &= \frac{1}{4} T_{original_encryption} + T_{transformation} + \frac{1}{4} T_{original_encryption} \end{aligned} \quad (10)$$

3.2 Data Decryption

Similar to data encryption, decryption also contains 3 steps: the first layer decryption, reverse matrix transformation and the second layer decryption. The first layer decryption is the decryption of the second layer encryption. The reverse matrix transformation is the reverse process of matrix transformation and the matrix

used in this process is the inverse matrix of G . The second layer decryption is the decryption of the first layer encryption. The pseudo-code algorithm of SIDA decryption is as follows.

```

program SIDA_Decryption(
  F: the encrypted file to decrypt; N: segment count in each data
  group; M: number of bytes in each segment; G*: inverse matrix of G;
  key1: the secret key for the second layer decryption;
  key2: the secret key for the first layer decryption)
{
  While ((f = Read(F, M, N)) != NULL) {
    // Read one data group from a decrypted file each time
    f =Decrypt1(f, key2); //First layer decryption with key2
    For (i=0; i<M; i++) {
      Temp=Extract(f); //Extract N bytes from a data group
      Temp2=Convert(Temp,G*); //Transform the extracted data
      //Write the extracted data back to the original location
      Reconstruct(Temp2);
    } //end for
    f=Decrypt2(f, key1); //The second layer decryption with key1
    Write(f, file); //Write the decrypted data back to file F
  } //end while
  Return the decrypted file F;
}

```

First layer decryption Equation 11 presents the first layer decryption, where $D1(\dots)$ indicates the first layer decryption and $E2(\dots)$ is the second layer encryption.

$$D1(F_i, key2) = D1(E2(A', key2), key2) + B' + C' + D' = A' + B' + C' + D' \quad (11)$$

Reverse of matrix transformation Equation 12 presents the reverse of matrix transformation, where G^{-1} is the inverse matrix of G .

$$Convert(F_i, G^{-1}) = G^{-1} \times \{M'_1, M'_2, \dots, M'_{32}\} = \{M_1, M_2, \dots, M_{32}\} \quad (12)$$

And in Eq. 12,

$$G^{-1} \times M'_i = G^{-1} \times \begin{bmatrix} a'_i \\ b'_i \\ c'_i \\ d'_i \end{bmatrix} = \begin{bmatrix} E1(a_i) \\ b_i \\ c_i \\ d_i \end{bmatrix}$$

Second layer decryption The time complexity of decryption can be approximately represented as the sum of the time cost during the first layer decryption, the reverse data transformation and the second layer decryption.

$$\begin{aligned}
T_{\text{decryption}} &= T_{D1} + T_{\text{reverse transformation}} + T_{D2} \\
&= \frac{1}{4}T_{\text{original_decryption}} + T_{\text{reverse transformation}} + \frac{1}{4}T_{\text{original_decryption}} \quad (13)
\end{aligned}$$

4 Performance and Security Analysis of SIDA

4.1 The Encryption and Decryption Complexity of SIDA

As mentioned above, we can get the following SIDA time complexity.

$$\begin{aligned}
T_{\text{encryption}} &= \frac{1}{4}T_{\text{original_encryption}} + T_{\text{transformation}} + \frac{1}{4}T_{\text{original_encryption}} \\
T_{\text{decryption}} &= \frac{1}{4}T_{\text{original_decryption}} + T_{\text{reversetransformation}} + \frac{1}{4}T_{\text{original_decryption}} \\
\text{So, } T_{\text{encryption} + \text{decryption}} &= \frac{1}{2}T_{\text{original_encryption} + \text{original_decryption}} \\
&\quad + T_{\text{transformation} + \text{reversetransformation}} \quad (14)
\end{aligned}$$

$$\begin{aligned}
\text{Then, } \Delta T &= T_{\text{original_encryption} + \text{original_decryption}} - T_{\text{encryption} + \text{decryption}} \\
&= \frac{1}{2}T_{\text{original_encryption} + \text{original_decryption}} - T_{\text{transformation} + \text{reversetransformation}} \quad (15)
\end{aligned}$$

where, ΔT is the gap between original encryption algorithm and SIDA. While the time complexity of matrix transformation is less than the original encryption algorithm. So, ΔT must be larger than 0. It is about half of the complexity of the original encryption algorithm and the performance of SIDA has been improved theoretically.

4.2 The Re-Encryption Complexity of SIDA

The client-side re-encryption model requires the data owner retrieves the whole data from the remote cloud server when the secret key leakage or other situations that need to revoke authority. So, the owner should download the data, re-encrypt and then upload to server. The time complexity is as Eq. 16.

$$T_{\text{normal_authority_revoke}} = T_{\text{download}} + T_{\text{re_encryption}} + T_{\text{upload}} \quad (16)$$

When using SIDA for the client side re-encryption, the second layer encryption just encrypt 1/4 of data. When revoking authority, the owner should only download this part of data and re-encrypt it. So, theoretically, SIDA can significantly reduce the cost of network communication and re-encryption.

$$\begin{aligned}
 T_{SIDA_authority_revocation} &= \frac{1}{4}T_{download} + \frac{1}{4}T_{re-encryption} + \frac{1}{4}T_{upload} \\
 &= \frac{1}{4}T_{normal_authority_revocation}
 \end{aligned}
 \tag{17}$$

4.3 Security Analysis of SIDA

The security of SIDA is guaranteed by the first layer encryption, matrix transformation and the second layer encryption.

In the first layer encryption, it encrypts 1/4 of the original data. And in data transformation process, SIDA chooses an n-order invertible matrix G , and use G to do matrix transform for the data unit $[E(a_1), b_1, c_1, d_1]$ and then get a new data unit $[a'_1, b'_1, c'_1, d'_1]$ with good randomness. Over all, we obtain a cipher text with good randomness by encrypting 1/4 part of original data and do matrix transformation to the whole data. In the second layer encryption, it encrypts the first segment A' of the data group again, and the corresponding cipher text is $[E(a'_1), b'_1, c'_1, d'_1]$. As a result, users who have $key1, A^{-1}$ and $[E(a'_1), b'_1, c'_1, d'_1]$ can't obtain the $[E(a_1), b_1, c_1, d_1]$ by reverse of data transformation. The above three layer model consists of SIDA security structure. Moreover, its encryption and decryption complexity is about half of the original one. So, SIDA significantly improved the encryption and decryption performance.

As of page limit, we omit the SIDA threat model and detailed experiments description. Anyone who interested in please contact the corresponding author.

5 Conclusion and Future Work

In this paper, a new encryption model SIDA is proposed based on the information dispersal and multi-layer encryption. The theoretical analysis shows that: (1) SIDA model is secure; (2) SIDA significantly improves the speed of data encryption and decryption; (3) SIDA reduces the communication and re-encryption costs when revoking authority. Taking SIDA4 algorithm as an example, the overhead of re-encryption when revoking authority, SIDA4 in communication and computation are 1/4 of AES.

However, there are three important parameters in SIDA: number of segments in a data group, number of bytes in segment and the n-order invertible matrix

G. Adjusting these parameters can affect the security and performance of SIDA. For example, if increase the number of segments in each data group, the performance of encryption and decryption will be improved, but the security strength also be affected. Similarly, the number of bytes in each segment determines the number of different values each segment has, which also affects the security strength of SIDA. The principle of selecting the invertible matrix G is choosing an invertible matrix without 0 elements. If the number of segments in each data group increases, the size of matrix G will increase too. But how to generate and optimize large scale reversible matrixes quickly should be studied further.

Acknowledgments This work is partially supported by Shanghai Municipal Natural Science Foundation (13ZR1416100).

References

1. Li H, Sun WH, Li FH, Wan BY (2014) Secure and privacy-preserving data storage service in public cloud. *J Comput Res Dev*
2. Bethencourt J, Sahai A, Waters B (2007) Ciphertext-policy attribute-based encryption. In: *Proceedings of the 2007 IEEE symposium on security and privacy*, pp 321–334, Piscataway, NJ. IEEE press, New York
3. Hong C, Zhang M, Feng DG (2010) AB-ACCS: a cryptographic access control scheme for cloud storage. *J Comput Res Dev (Sup):*259–265
4. Hong C, Zhang M, Feng DG (2011) Achieving efficient dynamic cryptographic access control in cloud storage. *J Commun* 32(7):125–132
5. Tian XX, Wang XL, Zhou AY (2009) DSP re-encryption: a flexible mechanism for access control enforcement management in DaaS. In: *IEEE international conference on cloud computing (CLOUD'09)*, Bangalore, September 21–25, pp 25–32
6. Zhou DH (2013) Studies on proxy re-encryption schemes. Dissertation of Shanghai Jiaotong University
7. Vimercati SDC, Foresti S, Jajodia S, Paraboschi S, Samarati P (2007) Over-encryption: management of access control evolution on outsourced data. In: *VLDB 2007*, pp 123–134
8. Fu K (1999) Group sharing and random access in cryptographic storage file system, Master's thesis, MIT
9. Sun H, Shieh S (1997) Optimal information-dispersal for increasing the reliability of a distributed service. *IEEE Trans Reliab* 46(4):462–472
10. Zhao D, Burlingame K, Debains C et al (2013) Towards high-performance and cost-effective distributed storage systems with information dispersal algorithms. In: *2013 IEEE international conference on cluster computing*, pp 1–5. IEEE, New York
11. Rabin M (1989) Efficient dispersal of information for security, load balancing, and fault tolerance. *J ACM* 36(2):335–348
12. Bian G, Gao S, Shao B (2011) Security structure of cloud storage based on dispersal. *Acad J Xi'an Jiaotong Univ* 45(4):41–45 (In Chinese)

Software Behavior Analysis Method Based on Behavior Template

Lai Yingxu, Zhao Yiwen and Ye Tao

Abstract Software security is not only related to our life, but also close to the security of our society. This paper proposed a method called software behaviors analysis method based on behavior template (SABT). According to the context of source code, we build and form a behavior template as a system to detect malicious behavior of software, including function transfer map and function block transfer map. We utilize some relative algorithms and technology in SABT, which include the method of stubbing interrupts, building behavior template and forming automaton to detect abnormal software behavior. Behavior template consists of function transfer map and minimum function transfer map. Compared with traditional method, such as N-gram, FSA, Var-gram, SABT can get higher cover rate of code and detect abnormal more effectively and efficiently.

Keywords Software behavior · Software interrupt · Behavior template · Minimum function block · Finite state automata

1 Introduction

Software security is not only related to our life, but also close to the security of our society. However, not all of software operates properly with the engineers' instructions, some of them have malfunction under attack. It will lead to serious results, due to the code defects or source code modified maliciously.

In order to respond to the challenges, we propose a method, SABT (Software behavior Analysis method based on Behavior Template), to analyze the software

L. Yingxu (✉) · Z. Yiwen (✉) · Y. Tao (✉)
Beijing University of Technology, Beijing, China
e-mail: laiyngxu@bjut.edu.cn

Z. Yiwen
e-mail: zhaoyiwen@emails.bjut.edu.cn

Y. Tao
e-mail: yt3262@126.com

source code. According to the context of source code, we build a behavior template to detect malicious software behavior.

In this paper, Sect. 2 is the related work on software behavior. The relative algorithms and technology are described in Sect. 3. In Sect. 4, we propose SABT model to build the software behavior model to detect abnormal. In Sect. 5, compare experiments on SABT, including the effectiveness evaluation and efficiency evaluation, are analyzed. And last section makes the conclusion about the method.

2 Related Work

Software behavior modeling methods can be divided into three categories: static analysis, dynamic analysis and hybrid analysis method.

Static analysis method does not run the application dynamically. Static analysis approaches with source code, binary analysis to build software behavior model [1]. Clariso [2] used the symbolic three-valued logic to improve data-flow analysis based on abstract interpretation. ASTREE [3] found the balance between efficiency and accuracy to reduce the computational cost, which enhanced precision by iterative calculation. Model detection [4, 5] expressed system behaviors by the state migration. Reference [6] used logic formula model and temporal method to detect the program. In addition, program slicing and constraint solving [7] were also utilized in static analysis techniques.

Dynamic Analysis method is applied to run the executable files. It carries on the massive executions to record the information of system call, memory, stack, etc. Based on the white-box testing, VEX [8] defined flow pattern to detect source-level vulnerabilities by data-flow analysis. KLEE [9] proposed an optimization algorithm including rewriting expression to improve the VEX. Wang etc. [10] put verification and sensation guidance in testing to improve the pertinence and effectiveness based on the black-box testing. In recent years, techniques including characteristics of pace and time, dynamic taint tracking are rapidly growth [11, 12].

To combine the advantages of static analysis and dynamic analysis, the behavior model is established by static analysis, and then uses the dynamic analysis model [13] for auxiliary correction, including Duck [14] model and Context Sensitive Host-Based IDS [15], etc.

Previous works are difficult in building the relationship between the system calls and software functions. In this paper, we analyze software source code firstly. And then, we obtain the relative system calls dynamically when running software marked by soft interrupts. Based on information, we obtain the corresponding relation between system call sequences and software functions, which is the base to build SABT model.

The Advantage of SABT is shown as follow:

- Our approach increases the model determinacy in software behavior. In present research, many methods used state transfer diagram to build software behavior

model. Our method is based on the corresponding relationship between the functions and system call sequence, which ensures that the malicious behavior detection accuracy.

- Our approach solves the problems that the extraction method of sequence corresponding function is not accurate. Our methods use the soft interrupt to obtain a variable-length system calls sequence corresponding the function. Compared to other methods, our method increase program coverage.
- Our approach solves the problems to detect fixed-length system calls sequences. Based on the block information in FBTM (function block transfer map), we build an automaton which can detect variable-length system sequence.

3 The Relative Algorithms and Technology

In SABT, we utilize some relative concepts and algorithms, including method of inserting soft interrupts, to build behavior automaton template.

We definite some concepts applied in the model before introducing our model.

Definition 1 Function (F): Each program fragment in software source code is a function. Function begins with the function name and ends with finishing running the function' execution. In the function, its arguments represent the file location (FL) and times of circulation if the times is greater than one.

If F_1 is located at FL_1 without loop, we record the function as $F_1(FL_1)$. If the times of F_1 circulation is 3, we definite it as $F_1(FL_1, 3)$.

Definition 2 Minimum function block of system call (Abbreviation function block) B: Each software program is composed of a number of related and integral system call blocks, which is called the minimum function blocks. B_i is one of the minimum function blocks, $B_i = \{A_1, A_2, \dots, A_n\}$, here, A is system calls. Therefore, by monitoring the program operation, the system call sequences are obtained. Each function corresponds to a function block, such as, $F_1(FL_1)$ corresponds to a function block B_1 , $F_1(FL_1) \rightarrow B_1$.

Definition 3 Function Sequence (FS): $FS = \{F_1(FL_1), F_2(FL_1), \dots, F(FL_j)\}$, it lists all the functions of the files.

Definition 4 System Calls Sequence (SCS): A SCS consists of two parts, one is the block of system calls corresponding function, and the other is system calls corresponding soft interrupts.

3.1 Function Transfer Map (FTM)

Function transfer map (FTM) is a part of behavior template. For the source code of the software, we build function sequence table to track the process, including nested functions, jump functions, circulation, etc. Based on the function sequence table, function transfer map can be built.

The algorithm of FTM is shown as follows.

```

Algorithm: function transfer map (FTM)
INPUT {FS, sourcecode}
Procedure of FTM{FS, sourcecode}
Init (FS)
For each even  $S \in \text{Sourcecode}$  do
For each even  $F \subset S$  do
If ( $S==s$ ) then //S for the condition statement
For each even  $S \in \text{true\_condition}$ ; //True branch
    F.next= New Node  $F_{i+1}$ ;
    For each even  $S \in \text{false\_condition}$ ; // false branch
F.next= New Node  $F_{i+1}$ ;
If ( $S==c$ ) then // c is a circulation statement
F.next= New Node  $F_{i+1}$ ;
    Update F.Fw; // F.Fw is restrain of function
    If (each F.Fw==circulation) //When the node in-
formation meets the circulation condition
break;
ReturnFS // Complete transfer function map

```

When some a function runs, the function sequence table will record the function's relative information. One function may contain multiple nested functions. In the process of operation, we set functions and jump functions to the next function operation. Thus the problems about it can get the effective solution.

When the condition statement is analyzed, each branch is used as the next node of the previous node by learning true or false branch gradually.

The analysis of circulation statement is difficult by code, such as F_1 cycle for the third times in function sequence. We use the exited method called multi-path analysis. Each path is set some a restraint value $F.Fw$ which limits the times of circulation function. According the run-time, $F.Fw$ is set to solve the problem that loop the same function over times. This method can reduce the cyclic redundancies to avoid massive repeat data in the paraphrase.

According the function sequence table, the FTM map can be built. Because the FTM is about the transfer of functions, sparse graph is utilized to record the FTM. We choose the adjacent list being the function sequence table.

3.2 Minimum System Calls Function Block Transfer Map

In order to generate minimum system calls function block transfer map (FBTM), soft interrupts are inserted to the front of each function, as Fig. 1 shown.

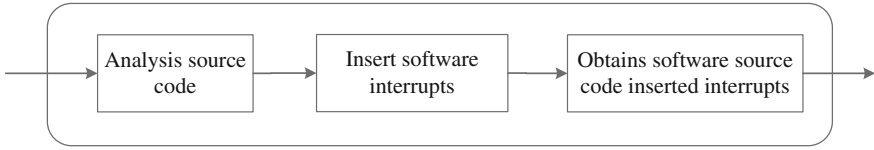


Fig. 1 The process of the pre-process module

When program marked by soft interrupts [16, 17] is executed, the normal operation of program is not affected, just like single step debugging. We can monitor the system calls with the help of soft interrupts. We apply the application to monitor and collect system calls sequences, which corresponds actual movement function path.

The minimum function block transfer map repeats several times to obtain the system call function block of each function. Analyzing for the system call function block of each function, Longest Common Subsequence (LCS) of each B is obtained by LCS extraction algorithm [18]. And then, build the set \sum removing a part of the LCS. LCS can measure the fixed part of the string. And \sum can calculate the similarity of the changeable sequence.

In this process, we get the LCS of each functional block, and finally build the FBTM.

3.3 Automata Based on FTM

Because software function statements are certain and the system call sequence is determined by the corresponding software function, finite-state automata can be utilized to detect the software behavior. In the finite-state automata based on behavior template, we define six elements in automata, $FTM = (B, I, G, Entry, Exit, W)$. The elements are shown as follow.

B is a state set of the automata, corresponding LCS invocation information of the minimum function block. In the process of building the model, according to the function block transfer map to build finite state automata, we add the LCS of function block transfer map in the automaton.

I is input alphabet.

G is for transfer path of set B , $(B \times I) \rightarrow B$. G is transformation rules of finite state automaton.

$Entry$ is an entrance behavior finite state automata, $Entry \in B$. Because the software has the unique entrance, $Entry$ is the only one state in the automaton.

$Exit$ is the export of the finite state automaton marked the end of the operation, $Exit$ is the only one state in automaton. $Exit \in B$.

W is as a restraint on the set B , which is used to store the information corresponding the function, W is constrain information about the automaton jumps and the constraint information including circulation times.

Finite-state automata are generated based on the minimum transfer function block map, which can be applied in the model as a learning content in the process of establishing automata.

4 System of SABT

SABT includes pre-process module, modeling module and examination module. The pre-process module takes the pretreatment on the source code of software. The modelling module learns software information and behavior, then build the software behavior template. The examination module according to the template detects malicious behavior in software running.

4.1 The Pre-process Module

The pre-process module takes the pretreatment on the source code of software. Soft interrupts was inserted between each two functions. Apply for debuggers attach, soft interrupts are set a destination addresses in software debugging. The data on the target address is stored by soft interrupt. Then the first byte of the destination address is replaced by soft interrupts. According the definition of breakpoint, it facilitates further observation on the program in kernel layer of system. When the interrupt is running, the running process of the original software is not affected.

4.2 The Modelling Module

The modelling module learns software information and behavior, and then builds a behavior template. Our method combines advantage of static analysis and dynamic method. We propose a two-layer method to generate software behavior map including function transfer map and minimum function block transfer map, then build the automaton based on minimum function block transfer map (Fig. 2).

- Sept 1 Analyze the source code of software to obtain FS . According to the context of FS , the FTM is built.
- Sept 2 Run the executable program marked by soft interruption and monitor the system calls, in order to obtain the complete SCS of executable program. In SCS, the same system call sequences are obtained by running of soft interrupts. System call function blocks are used to build the system call

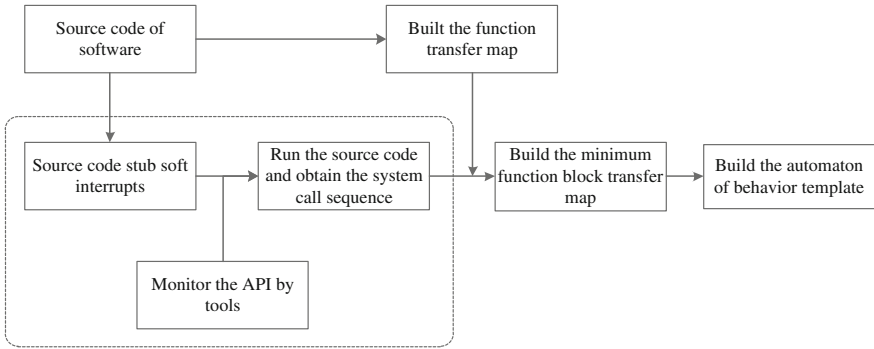


Fig. 2 The process of the modeling module

function blocks map according to the context of whole system call sequences. Each set of system calls functions block is run and learn repeatedly, then, gets the LCS for the common parts and other sets.

Sept 3 Build the automata based on FBTM to detect whether the software is abnormal.

4.3 The Detection Module

The detection module detects malicious behavior in software running in according with automaton (Fig. 3).

Sept 1 System call sequence of the tested software matches the automata entrance. This step compares variable-length sequences to obtain effective system call sequence.

Sept 2 When the system call sequences are loaded into the automaton, it starts detection. The system call sequences match the LCS with other states in automata individually, and cut the matching part that the distance between the tested sequence and automaton meet the specification.

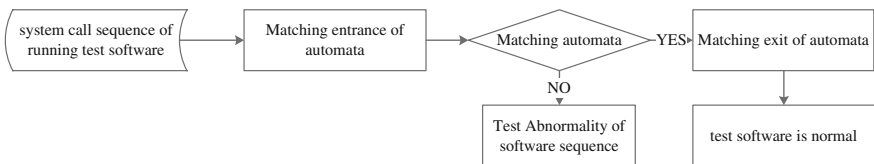


Fig. 3 The process of the detection module

Sept 3 If the tested software system calls sequence matches the automaton, the software is normal without abnormal condition. If there is no deviation in detection, the software behavior is abnormal.

The detection module can detect in time code injection attacks, Denial of Service attacks, etc.

5 Experiments

In this section, we conduct extensive experiments to evaluate the effectiveness and efficiency of the proposed framework SABT.

We make experiments on multi-platforms, including Windows system and Ubuntu system, to verify effect of the SABT. On the Windows system, RSS subscribe as the experimental object. Experiments show that the system calls coverage rate reached 99.85 %, the deviation range of module is $k = 0.15\%$ on Windows. On the system of Ubuntu 14.04, we take contrast experiment of various software behavior methods, including N-gram, Var-gram and FSA. Bash 4.3 and Tcpdump are tested software, and are under token composite attack. The experimental results indicated SABT can detect various abnormal behaviors caused by vulnerability. Analysis is shown in Table 1.

Code injection attacks insert illegal code into a software program to change the operation process and achieve additional functions. When malicious code is executed, the illegal system calls will be invoked, and are inconsistent with the normal. Function transfer model can detect anomalous sequence of function. If it is failed to detect abnormal state, SABT takes the further examination.

Abnormality caused by injecting code attack can be detected, such as vulnerability CVE-2014-6271, CVE-2014-6277 and CVE-2014-6278 in Bash 4.3, which invoked system call sequences are legitimate. N-gram's detections utilize fixed-length short system calls sequences, ignoring the relationship between sequences. Therefore detection for code injection appears omission. Var-gram is affected by the modeling file reported alarm mistakenly.

Table 1 Experiment Result

Vulnerability	Applications	N-gram (K = 7)	FSA	Var-gram (L = 6, W = 8, K = 20)	SABT
CVE-2014-6271	Bash	Normal	Anomaly	Anomaly	Anomaly (K = 0.015)
CVE-2014-6277	Bash	Normal	Anomaly	Anomaly	Anomaly (K = 0.018)
CVE-2014-6278	Bash	Normal	Anomaly	Anomaly	Anomaly (K = 0.022)
CVE-2014-8767	Tcpdump	Anomaly	Anomaly	Anomaly	Anomaly (K = 0.031)

Table 2 The size of the model files (kB)

	Perl	Tcpdump
FSA	483	502
N-gram	426	664
Var-gram	61	82
SABT	148	187

Remote denials of service attacks consume system resources, execution on the function repeatedly, and lead to the server denies new requests. Because SABT contains run times of functions and system calls' information of functions, this kind of attack can be detected.

Vulnerability CVE-2014-8767 belongs to remote denial of service attacks and aimed to cause the target program crashes, which can be detected by each method.

The model size shows the performance of model, we take experiment on the software tool, Perl and Tcpdump (Table 2).

In the comparison of the model size, SABT's modelling performance beyond the performance of N-gram and FSA.

6 Conclusion

According to the context of source code, we build a behavior template, SABT, to detect malicious behavior of software. SABT includes pre-process module, modeling module and examination module. Compared the method of N-gram and Var-gram, SABT beyond others performance. In later work, we will improve the algorithm in modelling module and enhance the performance more effety and efficiently of SABT.

References

1. Hong M, Qianxiang W, Lu Z, Ji W (2009) Software analysis: a road map. *Chin J Comput* 32 (9):1697–1710
2. Clariso R, Cortadella J (2007) The octahedron abstract domain. *Sci Comput Program* 1 (64):115–139
3. Bouissou O, Conquet E, Cousot P (2009) Space software validation using abstract interpretation. In: *The international space system engineering conference on data systems in aerospace*. CiteseerPress, DASIA, pp 1–7
4. Jhala R, Majumdar R (2009) Software model checking. *ACM Comput Surv* 41(4):1729–1739
5. Schlich B, Kowalewski S (2009) Model checking C source code for embedded systems. *Int J Softw Tools Technol Transfer* 3(11):187–202
6. Gulwani S, Srivastava S, Venkatesan R (2008) Program analysis as constraint solving. *ACMSIGPLAN Notices* 6(43):281–292
7. Bixin L (2000) Program slicing techniques and its application in object-oriented software metrics and software test. Nanjing University

8. Bandhakavi S, King ST, Madhusudan P (2010) VEX: vetting browser extensions for security vulnerabilities. In: The 19th USENIX conference on security. CiteseerPress, Washington, pp 1–16
9. Cadar C, Dunbar D, Engler D (2008) KLEE: unassisted and automatic generation of high-coverage tests for complex systems programs. In: The 8th USENIX symposium on operating systems design and implementation. CiteseerPress, San Diego pp 209–224
10. Wang T, Wei T, Gu G (2010) TaintScope: a checksum aware directed fuzzing tool for automatic software vulnerability detection. In: The 2010 IEEE symposium on security and privacy. IEEE Press, Oakland, pp 497–512
11. Tian R, Batten LM, Islam R (2010) Differentiating malware from clean ware using behavioral analysis. In: The 5th IEEE international conference malicious and unwanted software. IEEE Press, Nancy, pp 23–30
12. Sami A, Rahimi H, Yadegari B (2010) Malware detection based on mining API calls. In: The ACM symposium on applied computing. DBLP, Sierra, pp 1020–1025
13. Ruoyu Z, Shiqiu H, Zhengwei Q, Haibin G (2011) Combining static and dynamic analysis to discover software vulnerabilities. In: Fifth international conference on innovative mobile and internet service in ubiquitous computing. IEEE Press, Washington, pp 175–181
14. Giffin JT, Jha S, Miller BP (2003) Efficient context-sensitive intrusion detection. Nds
15. Wen L, Yingxia D, Yifeng L (2009) Context sensitive host-based IDS using hybrid automaton. *J Softw* 20(1):138–151
16. Mcdowell CE, Helmbold DP (1989) Debugging concurrent programs. *ACM Comput Surveys* 21(4):593–622
17. Wahbe R (1992) Efficient data breakpoints. In: The 5th international conference on architectural support for programming, vol 27, issue (9), pp 200–212
18. Kaiyun W, Siqi K, Yunsheng F (2013) Two longest common substring algorithms based on bi-directional comparison. *Comput Res Dev* 50(11):167–170

Formalizing Dynamic Service Interaction Based on Pi-Calculus

Yaya Liu, Jiulei Jiang and Wenwen Liu

Abstract In order to ensure the correct transmission of concurrent data and resolve the uncertainties of communication channels in the across-organizational business process, a new modeling method of dynamic service interaction based on pi-calculus is proposed in this paper. The pi-calculus is selected as the formal modeling language in this method. And three service interaction patterns, which includes request with referral, relayed response and dynamic routing, is studied to build the formal model with the channel mobility and the messaging mechanism of pi-calculus. In case of the bidding activities, a formal model is established and simulated based on pi-calculus in this paper. Furthermore, a tool named MWB is used to automatically validate the model in order to ensure the accuracy and the consistency of process. It proved the applicability and feasibility of the modeling based on the pi-calculus in the dynamic service interaction.

Keywords Dynamic service interaction · Pi-Calculus · Service interaction patterns · MWB

1 Introduction

1.1 Background

With the continuing development of information technology, the traditional object-oriented model faded gradually and the *Service-Oriented Architectures* (SOA) as a component model took its place [1]. SOA is the best way for companies

Y. Liu · J. Jiang (✉) · W. Liu
College of Computer Science and Engineering,
The Beifang University of Nationalities, Yinchuan, China
e-mail: 525128060@qq.com

Y. Liu
e-mail: 503665638@qq.com

W. Liu
e-mail: 645040920@qq.com

to achieve business process management (BPM), and can support not only the business processes in enterprises but also the across-organizational processes. The services communicate with each other through messages in the cross-organizational processes [2]. To some extent, interaction-centric modeling of process is becoming a major concern for business process modelers due to the complexity of interactions among different organizations [3, 4].

Pi-calculus as a model of describing variable interactive systems has the characteristics of mobility and can describe complex interaction with its rigorous algebra semantics in a form of formal [5]. Meanwhile, MWB supports the automatic validation for formal models based on *pi-calculus* so as to ensure the correctness of semantics and interaction. Currently in the formal modeling of *service interaction patterns* based on *pi-calculus*, Gero Decker and Frank Puhmann mainly completed the description of the *single-transmission* [6], and Jiulei Jiang built the formal model for all interaction patterns. Nevertheless, the description is so simple that can not reflect the dynamic characteristics of the interaction [7].

1.2 Content of the Article

This paper focus on how the formal modeling in the dynamic service interaction process should be conducted. First, we respectively build the formal models for three dynamic interaction patterns with the mobility of channel and messaging mechanism of the *pi-calculus*. Moreover, an instance about the bidding activities is presented and formal modeled based on *pi-calculus* in order to enhance the modeling ability for dynamic service interaction. Finally, the interaction is simulated with reaction rules and automatically validated with MWB.

2 Modeling Method

2.1 Dynamic Service Interaction

The cross-organizational service interaction includes not only internal operations in an enterprise but also external operations among enterprises [2, 8, 9]. In order to avoid the simultaneously access to the same channel from multiple enterprises at the same time, the traditional service interactions play with such conundrums by respectively setting different channel for each of the joint venture. However, this scheme increases the complexity of system modeling, and is not conducive to the later expansion of the system.

The dynamic service interaction refers to that it does not to open up different channel for each data type in advance, but adopts the feature that the channel name can be passed along the channel to achieve the dynamic binding of the follow-up

channels in the migration. It is worthy of noting that the follow-up channels is dynamically generated and removed in the dynamic service interaction. Figure 1 shows the specific interaction process.

2.2 Model About Dynamic Interaction

In the *service interaction patterns*, the *request with referral*, *relayed response* and *dynamic routing* all have the obvious dynamic characteristics [10]. According to the way of information exchange [9, 11], the appropriate modeling approaches based on *pi-calculus* are given in the following.

Model 1 (Request with Referral). In the *request with referral*, the middle agent *Y* is to provide a reliable class broadcast mechanism based on some specific criteria. First agent *X* sends a service request to the agent *Y* with the reference set of channels. Agent *Y* then makes judgment which agents (Z_1, Z_2, \dots, Z_n) will meet the request by the received message and decides the next communication channels according to the reference set. Eventually, it will broadcast the message to all of these default agents along these channels. Figure 2a shows the specific interaction and the completed semantic description based on the *pi-calculus* is shown below:

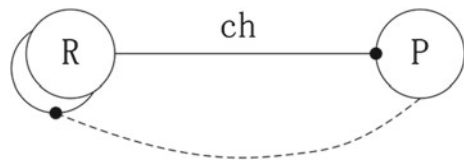
$$X = \overline{ch}1\langle z, msg \rangle \cdot 0 \quad Y = ch1(z, resp) \cdot \prod_{i=1}^n (\overline{z_i}\langle resp \rangle \cdot 0) \quad Z = z(resp) \cdot Z \quad (1)$$

There are three points should be observed in the interaction:

- The default agent eventually broadcasted could be another nominated party (possibly agent *A*);
- The agent *Z* is merged with agent *X* into a single interactive partner not through a shared name *z*. That is, multiple processes can also be integrated together to participate in the interaction by z_i ;
- The channel used to broadcast message from agent *Y* is dynamically generated and removed according to the reference set of channels included in the message from agent *X*.

Model 2 (Relayed Response). *Relayed response* also named the delayed request is to achieve the forwarding of request and monitor the errors in the view of interaction through agent *Y*. Agent *X* sends a request to agent *Y* with a channel for respond, and agent *Y* broadcasts the request to other agents (Z_1, Z_2, \dots, Z_n). These

Fig. 1 The dynamic binding of channel



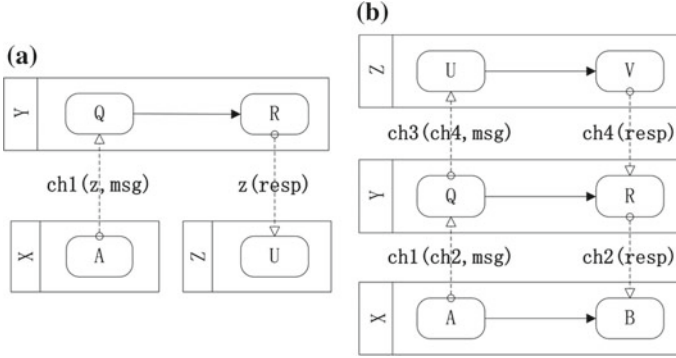


Fig. 2 The request with referral and the relayed response

agents then continue interactions with the agent X. Figure 2b shows the specific interaction. According to the above interaction, the completed description of the model based on the *pi-calculus* is shown in the below:

$$X = \overline{ch1}\langle ch2, msg \rangle . \prod_{i=1}^n (ch2\langle resp \rangle . 0) \quad Z = ch3\langle ch4, msg \rangle . \overline{ch4}\langle resp \rangle . Z \quad (2)$$

$$Y = ch1\langle ch2, msg \rangle . \prod_{i=1}^n (\overline{ch3_i}\langle ch4, resp \rangle . ch4\langle resp \rangle . \overline{ch2}\langle resp \rangle . 0) \quad (3)$$

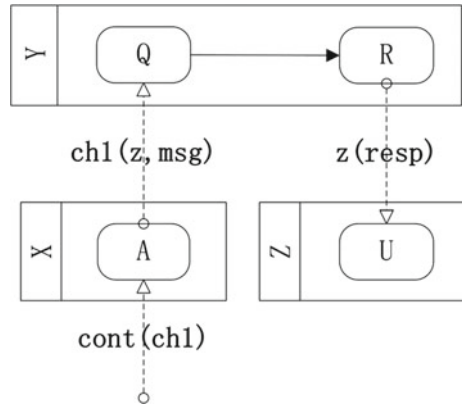
Agent X does not directly interact with agents (Z_1, Z_2, \dots, Z_n) . The interaction between them fully completed by the middle agent Y. And the channel for respond is a private channel of the parent agent and dynamically generated and removed. In addition, from a global perspective, the view monitored by the middle agent Y remains transparent for all the agents.

Model 3 (Dynamic Routing). *The dynamic routing* emphasizes the uncertainty and dynamic of behavior by adding an external control signal also named routed conditions in service interaction. A service request is routed to several agents based on the external routing conditions. The next set of processes respond to the request and then determine which channel is used to send the message based on the request. Figure 3 illustrates the process, and the completed description of the model based on the *pi-calculus* is shown in the below:

$$X = cont\langle ch1 \rangle . \overline{ch1}\langle z, msg \rangle . 0 \quad Z = z\langle resp \rangle . Z \quad (4)$$

$$Y = ch1\langle z, resp \rangle . [resp = msg] \prod_{i=1}^n (\overline{z_i}\langle resp \rangle . 0) \quad C = \overline{cont}\langle ch1 \rangle . 0 \quad (5)$$

Fig. 3 The dynamic routing



The dynamic of the interaction is mainly reflected in the uncertainty choice of routing paths in the dynamic routing. The reason for uncertainty can be attributed to the following two aspects:

- The routing order is flexible, and more than one agent can be activated to receive the request;
- The data included in the initial request or obtained from the interactive process is likely to have impact on the selection of routing paths.

3 System Modeling

A specific formal model about bidding activities is built, as follows.

3.1 Classification of the Agents

The agents of the system can be divided into the following categories:

- The tenderer: generally refers to the legal entities or organizations proposing bidding projects;
- The bidding agency: generally refers to the intermediary organizations that is legally established and commissioned by the tenderer to organize activities and provide related services;
- The center of bidding and tendering (CBT): mainly to provide the qualification of bidders.
- The bidder: generally refers to the legal person or organizations that aim to winning the bidding;
- The e-bank: generally provides the services of fund in the bidding activities.

3.2 Service Interaction

Following description about the interaction of bidding activities is carried out: the basic business process is that the tenderer entrusts a bidding agency for the execution of bidding operations according to self-demand, and the bidders are reviewed through the prescribed procedures by bidding agency. In a word, all the activities about bidding is aimed at the selection of the successful bidder. Figure 4 shows the specific interaction process.

3.3 Formal Description Based on Pi-Calculus

To discuss the approach of the formal modeling [12] based on *pi-calculus*, this paper simplified the business processes on some of the details.

The specifically semantic description is as follows:

$$Tenderer = \bar{x}(p, deleRequest) . p(deleResponse) . 0 \quad (6)$$

$$PA = x(pch, deleRequest) . \overline{pch}(deleResponse) . PAService \quad (7)$$

$$PAService = \bar{y}(k, invitation) . (PAService + w(checkOk)) \\ \cdot \bar{y}(g, payInfo) . n(paySuccess)PAService \quad (8)$$

$$CBT = k(quaCheck) . [quaCheck = checkOk] \bar{w}(checkOk) . 0 \quad (9)$$

$$Bidder = y(kch, invitation) . (Bidder + \bar{k}(quaCheck)) \\ \cdot y(gch, payInfo) . \overline{gch}(payRequest) . Bidder \quad (10)$$

$$EBank = g(payRequest) . \bar{n}(paySuccess) . 0 \quad (11)$$

The realization of the whole interaction may be described as follow:

$$I = Tenderer | PA | Bidder | \prod CBT | \prod EBank \quad (12)$$

After the description above, we can conclude that the *pi-calculus* can vividly and accurately describe the actual interaction of the completed system, and also can clearly depict the dynamic interaction about the delivery of channels in detail [5]. Consequently, the *pi-calculus* has a greater ability to the expression of the mobility and dynamics.

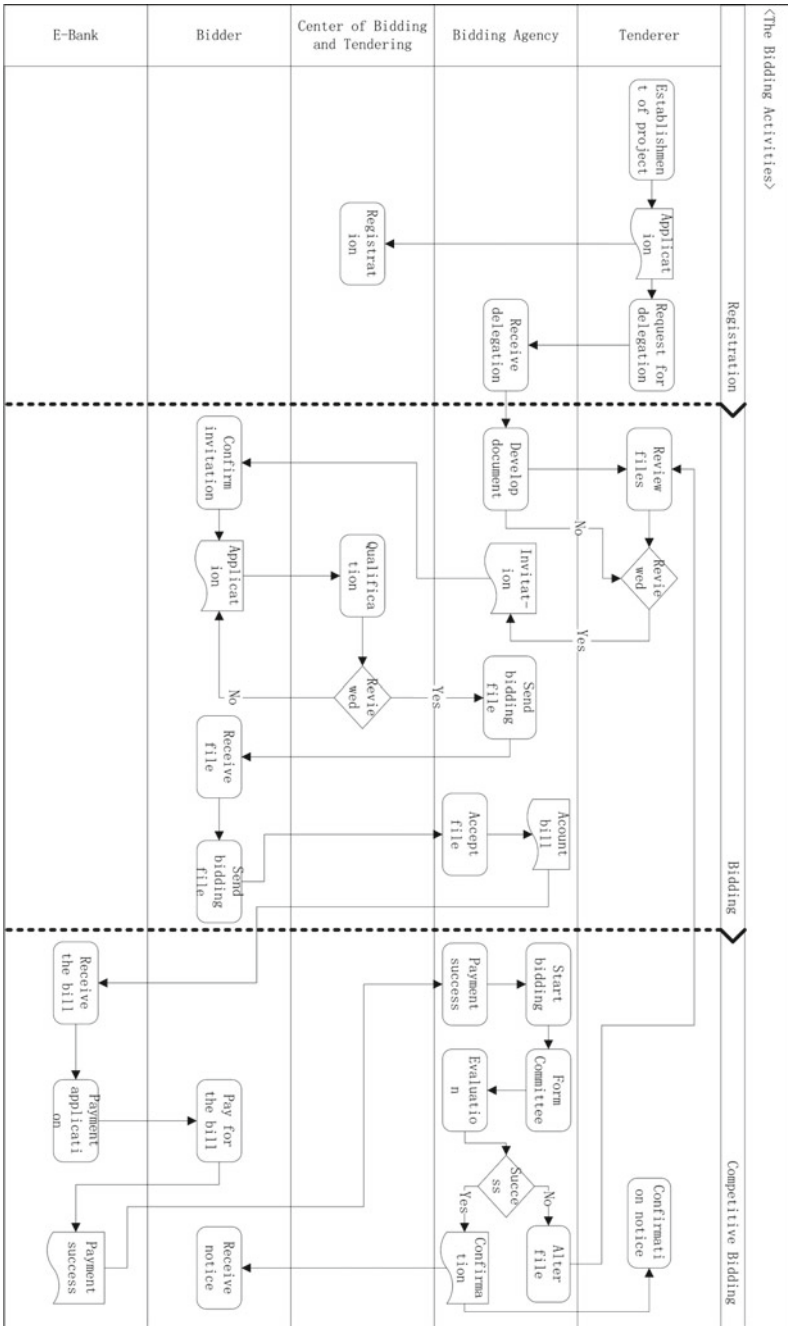


Fig. 4 The flow diagram of system

4 Analysis and Validation

In order to ensure the accuracy, completeness and consistency of the model, interaction is simulated and analyzed based on the reaction rules of the *pi-calculus* [7, 13]. Whether the interaction is in keeping with the expectations is the prerequisite and foundation to achieve that there exists no deadlock between various participants. The specifically process of the validation is listed below, and the first step is the triggered request of mandate:

$$\begin{aligned}
 I &= \text{Tenderer} | PA | \text{Bidder} | \prod CBT | \prod EBank \\
 &\stackrel{x}{=} p(msg) . 0 | \overline{pch} \langle deleResponse \rangle . P\text{AService} | \text{Bidder} | \prod CBT \\
 &\quad | \prod EBank \stackrel{p}{=} P\text{AService} | \text{Bidder} | \prod CBT | \prod EBank = \text{State1}
 \end{aligned} \tag{13}$$

Then the request of invitation is triggered:

$$\begin{aligned}
 \text{State1} &= P\text{AService} | \text{Bidder} | \prod CBT | \prod EBank \\
 &\stackrel{y}{=} (P\text{AService} + w(\text{checkOk}) . \overline{y} \langle g, \text{payInfo} \rangle . n(\text{paySuccess}) \\
 &\quad P\text{AService}) | (\text{Bidder} + \overline{k} \langle \text{quaCheck} \rangle . y(\text{gch}, \text{payInfo}) . \overline{gch} \\
 &\quad \langle \text{payRequest} \rangle . \text{Bidder}) | \prod CBT | \prod EBank = \text{State2}
 \end{aligned} \tag{14}$$

There are two kinds of migrations of the process when the interaction reaches to the state 2. The first one refers to a migration of invalid invitation, namely a time-barred message or a attempt to absence from the bidding activities, and the other is a valid invitation. After a series of operations, the interaction reaches to the state 1 once again:

$$\begin{aligned}
 \dots &\stackrel{g}{=} (n(\text{paySuccess}) . P\text{AService}) | \text{Bidder} | \overline{n} \langle \text{paySuccess} \rangle . 0 | \prod CBT \\
 &\quad | \prod EBank \stackrel{n}{=} P\text{AService} | \text{Bidder} | \prod CBT | \prod EBank = \text{State1}
 \end{aligned} \tag{15}$$

It means that a full interaction between the agents is completed when the system enters into state 1. The above process just acts it out in the continuous circular behaviors. In order to ensure the correctness of interaction, this paper mainly uses a combined formal validation, namely the artificial deduction based on reaction rules and the formal verification based on MWB tool [14]. The codes in MWB and detected result of deadlocks are shown below.

The Formal Description and Validation

MWB>agent Bidder = (\y,k,quaCheck,gch,payReq)y(kch,invitation). (Bidder<y,k,quaCheck,gch,payReq> + 'k<quaCheck> .y(gch,payInfo).'gch<payReq> . Bidder<y,k,quaCheck,gch,payReq>)

MWB>agent CBT = (\k,w,quaCheck,checkOk)k(quaCheck).[checkOk=quaCheck]'w<checkOk> .CBT<k,w,quaCheck,checkOk>

MWB>agent EBank = (\g,n,payS)g(payReq).'n<payS> .EBank<g,n,payS>

```

MWB>agent I = (\x,p,deleReq,pch,y,k,invitation,w,g,payInfo,n,dele Resp,quaCheck,
gch,payReq,checkOk,payS)(Tenderer<x,p,deleReq>|PA<x,pch,y,k,invitation,
w,g,payInfo,n,deleResp>| Bidder<y,k,quaCheck,gch,payReq> | CBT<k,w,quaCheck,
checkOk>| EBank<g,n,payS>)
MWB>agent PA = (\x,pch,y,k,invitation,w,g,payInfo,n,deleResp)x(pch, deleReq).’
pch<deleResp>.PAService<y,k,invitation,w,g,payInfo,n>
MWB>agent PAService=(\y,k,invitation,w,g,payInfo,n)’y <k,invitatio n>.(PASer-
vice<y,k,invitation,w,g,payInfo,n>+w(checkOk).’y<g,payInfo>.n(payS).
PAService<y,k,invitation,w,g,payInfo,n>)
MWB>agent Tenderer = (\x,p,deleReq)’x<p,deleReq>.p(deleResp).0
MWB>deadlocks PA
No deadlocks found.
MWB>deadlocks CBT
No deadlocks found.
MWB>deadlocks EBank
No deadlocks found.
MWB>deadlocks Bidder
No deadlocks found.
MWB>step I

```

The automatic validation occurs with the input of command step I. At each step of the simulation MWB presents us with the different possible actions of the process (numbered from 0 to n). We can choose one of the actions and MWB will then present us for the new choices etc. It means that a full interactive process is completed when the system enters into a predictable cycle. The specific result of detection is shown below. Therefore, the correspondence of result between the automatic validation based on MWB and artificial deduction shows that the *pi-calculus* is applied to the modeling of dynamic interaction.

The Detection of Circular Behavior

Step>4

[Circular behaviour detected]

0: >t. ((PAService<y,k,invitation,w,g,payInfo,n> + w(checkOk18)).

5 Conclusions

The principle of dynamic binding of channels is applied to suggest a new method for the formal modeling of the dynamic service interaction. It can greatly reduce the complexity and increase the flexibility and scalability of the system modeling. This paper provides an extraordinarily detailed description of service interaction patterns based on *pi-calculus*, and establishes a semantic model of the bidding activities on it. In order to assure the validity of interaction, a combined validation between the automatic validation based on MWB and artificial deduction is put out to validate

and analyze the model that whether there exists deadlocks or not in the interaction. The next study is focus on the modeling for the other service interaction patterns and more complex service interaction.

Acknowledgments This study on the dynamic service interaction is partially supported by the Postgraduate Innovation Project of Beifang University of Nationalities, and NSFC pre-breeding program of Beifang University of Nationalities (No. 2012QZP02).

References

1. Massuthe P, Reisig W, Schmidt K (2005) An operating guideline approach to the SOA. In: Proceedings of the 2nd South-East European workshop on formal methods 2005 (SEEFM05). Ohrid, Republic of Macedonia
2. da Silva FSC, Venero MLF (2013) Interaction protocols for cross-organisational workflows. *J Knowl Based Syst* 37(1):121–136
3. Yuxi FU, Lu HAO (2010) On the expressiveness of interaction. *J Theor Comput Sci* 411 (11/13):1387–1451
4. Gero D, Mathias W (2011) Interaction-centric modeling of process choreographies. *J Inf Syst* 36:292–312
5. Milner R (1999) *Communication and mobile systems: the pi-calculus*. Cambridge University Press, Berlin
6. Decker G, Puhlmann F, Weske M (2006) Formalizing service interactions. In: Dustdar S, Faideiro J, Sheth A (eds) *International conference on business process management (BPM 2006)*, LNCS, vol 4102. Springer, Berlin, pp 414–419
7. Jiulei J, Jiajin L (2012) Using-calculus for formalizing service interactions. *Int J Digital Content Technol* 6(7):190–197
8. Workflow Management Coalition the Workflow Reference Model. <http://www.wfmc.org/standards/docs/tc003v11.pdf>
9. Russell N, Aalst W, Hofstede A, Edmond D (2005) Workflow resource patterns: identification, representation and tool support. In: Pastor O, Falcao e Cunha J (eds) *Proceedings of the 17th conference on advanced information systems engineering (CAiSE'05)*. LNCS, vol 3520, pp 216–232. Springer, Berlin
10. Barros A, Dumas M, Hofstede A (2005) Service interaction patterns. In: Aalst W, Benatallah B, Casati F, Curbera F (eds) *International conference on business process management*. LNCS, vol 3649, pp 302–318. Springer, Berlin
11. Mulyar N, Aldred L, Aalst W (2007) The conceptualization of a configurable multi-party multi-message request-reply conversation. In: Felber P, Pu C, Moorsel A (eds) *Proceedings of the OTM conference on distributed objects and applications (DOA2007)*. LNCS, vol 4803, pp 735–753. Springer, Berlin
12. Zaha J, Dumas M, Hofstede A, Barros A, Decker G (2006) Service interaction modeling: bridging global and local views. In: *International enterprise distributed object computing conference (EDOC2006)*. IEEE Computer Society, pp 45–55
13. van der Aalst WMP, Mooij AJ, Stahl C, Wolf K (2009) Service interaction: patterns, formalization, and analysis. In: *9th international school on formal methods for the design of computer, communication, and software systems* (eds) June 1–6, 2009. LNCS, vol 5569, pp 42–48. Springer, Berlin
14. Jonsson B (1994) A verification tool for the polyadic pi-calculus. Licentiate thesis, Department of Computer Systems, Uppsala University, Sweden. Available as report DOcs 94/50

Applications of Video Structured Description Technology for Traffic Violation Monitoring

Qianjin Tang, Zheng Xu, Zhizong Wu, Yixuan Wu and Lin Mei

Abstract Action analysis and semantic interpretation in surveillance video have recently attracted increasing attention in the computer vision community. In this paper, video structural description model is proposed for practical applications for traffic violation monitoring. Conceptual space is defined to bridge the gap between low-level syntax which is quantitative and high-level semantic where information is handled by qualitative means. Based on the conceptual space, conceptual relating model is proposed to simulate and recognize the targets' behaviors in the scene. Applications for traffic violation monitoring experimental results demonstrate the performance of the proposed semantic interpretation model of video structural description.

Keywords Video structural description · Surveillance video · Semantic · Traffic violation

1 Introduction

The surveillance video data has grown tremendously in recent years, but traffic violations still cause a lot of accidents and personal injury every year. There are the problems of “cannot find”, “difficult to understand” for the mass surveillance video retrieval system. The increasing need of video based applications issues the importance of parsing and organizing the content in videos. However, the accurate

Q. Tang · Z. Xu · Z. Wu (✉) · Y. Wu · L. Mei
The Third Research Institute of Ministry of Public Security, Shanghai, China
e-mail: wuzizong@163.com

Q. Tang
e-mail: Tangqj2008@163.com

Z. Xu
e-mail: xuzheng@shu.edu.cn

Y. Wu
e-mail: Wyx876@163.com

understanding and managing video contents at the semantic level is still insufficient. These features are at a higher level description of the video content, but because of the lack of means of the unified representation and modeling of human eye domain knowledge, thus forming a gap between low-level grammar features and high-level semantic features, which is a difficult problem of semantic understanding faced in surveillance video. The semantic gap between low level features and high level semantics cannot be bridged by manual or semi-automatic methods. In this paper, a semantic based model named video structural description (VSD) for representing and organizing the content in videos is proposed. Video structural description aims at parsing video content into the text information, which uses spatiotemporal segmentation, feature selection, object recognition, and semantic web technology. The proposed model uses the predefined ontologies including concepts and their semantic relations to represent the contents in videos. The defined ontologies can be used to retrieve and organize videos unambiguous. In addition, besides the defined ontologies, the semantic relations between the videos are mined. The video resources are linked and organized by their related semantic relations. To illustrate the VSD technology, we choose the traffic violation monitoring scene. Two kinds of the traffic violation action are derived by the semantic relation. The applications show that VSD technology contains real information obtained by analyzing the image sequence, so that the monitoring system has the ability to understand, and use semantic concept of people used to describe video, enabling automatic and efficient intelligent video processing.

2 Related Work

Video structural description (VSD) aims at parsing video content into the text information, which uses spatiotemporal segmentation [1], feature selection [2], object recognition [3], and semantic web technology [4]. The parsed text information preserves the semantics of the video content, which can be understood by human and machine. Generally speaking, the definition of VSD includes two aspects. Firstly, VSD aims at extracting the semantic content from the video. Relying on the standard video content description mechanism, the objects and their features of the video are recognized and expressed in the form of text. Secondly, VSD aims at organizing the video resources with their semantic relations. With the semantic links across multiple cameras, it is possible to use the data mining methods for effective analysis and semantic retrieval of videos. Mei [5] proposed video structured description prototype system and benefit from knowledge modeling, visual information distilled by the proposed method could be accessed by other information systems much easier than before. Jiang [6] introduced that knowledge map in the traffic violations area was constructed to detect traffic violations. Xu [7] proposed a semantic based model for

representing and organizing video big data. The proposed surveillance video representation method defines a number of concepts and their relations, which allows users to use them to annotate related surveillance events. The defined concepts include person, vehicles, and traffic signs, which can be used for annotating and representing video traffic events unambiguously. Li [8] provided an overview of the algorithms and technologies used in extracting static properties of vehicle in the video. Xu [9] introduces a video annotation platform, which enables user to semantically annotate video resources using vocabularies defined by traffic events ontologies and provides the search interface of annotated video resources. Xu [10] proposed a semantic based model for representing and organizing video big data. The proposed method defines a number of concepts and their relations, which allow users to use them to annotate video traffic events. The defined concepts including people, vehicle, and traffic sign, which can be used by users for annotating and representing video traffic events unambiguously.

3 Surveillance Video Semantic Model

3.1 *Sematic Need*

According to industry standards “GA/T 832—2009 Technology specifications of image forensics for road traffic offence”, a motor vehicle violation should meet the following conditions:

1. Number of pictures. A panoramic image of a motorized vehicle with no less than 2 different time or different positions.
2. Displacement interval. The displacement of the motor vehicle in the two pictures is over 1.0 m.
3. Information included in the picture. The picture contains clear identification of motor vehicle driving characteristics, vehicle front features, the tail end of a motor vehicle features panoramic views, number license, motor vehicle driving direction signs, motor vehicles through the stop line at the traffic signal lamp indicating state, guide arrow, parking is prohibited marking instructions.

3.2 *Video Sematic Model*

Concepts, objects, attributes, spatial relations, temporal relations, and events are basic components of the proposed metadata for traffic violation monitoring. In this section, we give the basic definitions of these components.

Definition 1 Traffic Violation Event (TVE): TVE is the combine of the objects and their spatial-temporal relation, which can be denoted as

$$TVE = (\text{Objects}, \text{Spatial Relation}, \text{Temporal Relation}) \quad (1)$$

Definition 2 Objects (O): Object is the extracted component from a video. The extracted object is mapped to a concept. The object can be denoted as

$$\text{Object} = \{O_1, O_2, O_3, O_4\} \quad (2)$$

Where O_1 means motor vehicle, O_2 means traffic light, O_3 means traffic direction arrow, O_4 means traffic marking line.

Definition 3 Temporal Relation (TR): Temporal is the time relation between the different time intervals video. The temporal relation can be denoted as:

$$TR = \{T_1, T_2, T_3, T_i\} \quad (3)$$

Definition 4 Spatial Relation (SR): Spatial relation is the position relation between the different objects of a video. The spatial relation can be denoted as:

$$SR = \{P_1, P_2, P_3, P_i\} \quad (4)$$

Definition 5 Attribution (A): Attributions are the parameters of the objects. In traffic violation monitoring applications, the object attribution is as following:

$$O_1 = \{\text{speed}, \text{license number}, \text{straight}, \text{left turn}, \text{right turn}\};$$

$$O_2 = \{\text{red}, \text{yellow}, \text{green}\};$$

$$O_3 = \{\text{straight}, \text{left turn}, \text{right turn}\};$$

$$O_4 = \{\text{solid line}, \text{dotted line}, \text{stop line}, \text{strait line}, \text{left turn line}, \text{right turn line}\}.$$

On the basis of the concept space, the relationship between concepts can be further explored through the concept correlation model. Concept correlation model is a model based on rule, which purpose is to find the relationship of concept correlation in lens. For example, the semantic set of current set is {vehicle, traffic direction arrow}, and the semantic set of next set is likely to be {vehicle, traffic direction arrow, traffic lights}. The correlation model of concept finds correlation rules of concept directly from semantics. In video semantic modeling process, based on statistical methods can not very good to complete complex concepts and meanings of the concept of task processing. Use concept relative model to semantic description corresponds to the concept of space in the description of the combination. Through the concept of space and concept association model, human visual system accepts video information is converted to semantic video classification, association and description.

4 Traffic Violation Monitoring Application

4.1 Run the Red Traffic Light

According to the red light punishment basis “GA/T 832—2009”, police requires three photos as evidence, when the signal light is red light, the first one is crossing the stop line at the wheel, the second is the vehicle passes through the intersection, the third is the vehicles pass through another stop line, which are shown in Fig. 1 at P1, P2 and P3, respectively.

According to the definitions of video sematic model, the rule of run the red light action (*RRL*) can be described as following:

$$RRL_1 = \{O_1.\text{speed} > 0, O_1.\text{license number} = A, O_2 = \text{red}, O_3 = \text{straight}, TR = T_1, SR = P_1\}$$

$$RRL_2 = \{O_1.\text{speed} > 0, O_1.\text{license number} = A, O_2 = \text{red}, O_3 = \text{straight}, TR = T_2, SR = P_2\}$$

$$RRL_3 = \{O_1.\text{speed} > 0, O_1.\text{license number} = A, O_2 = \text{red}, O_3 = \text{straight}, TR = T_3, SR = P_3\}$$

Through the above three sequence of scenes composed of red light, the rule of run red light event description model as following:

$$RRL_3 = \{RRL_1 \text{ AND } RRL_2 \text{ AND } RRL_3 \text{ AND } T_1 < T_2 < T_3 \text{ AND } P_2 - P_1 > 1m \text{ AND } P_3 - P_2 > 1m\}$$

Where A is a fix number, which means the same vehicle; m is the length unit; “*O*_{1,+}” means the attribution of *O*₁.

4.2 Combined Lane Judge Violation

As shown in the Fig. 2, straight and turn left traffic lights are red, and the right turn signal light is green, the lane traffic marking of the vehicle is straight and right turn. according to these information, the lawful driving actions can be derived as following:

Fig. 1 The illustration of a vehicle run the red light

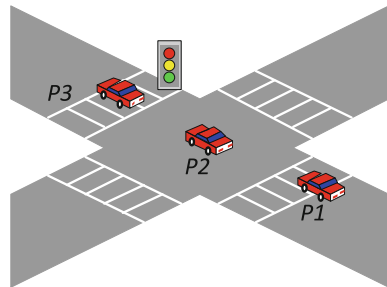
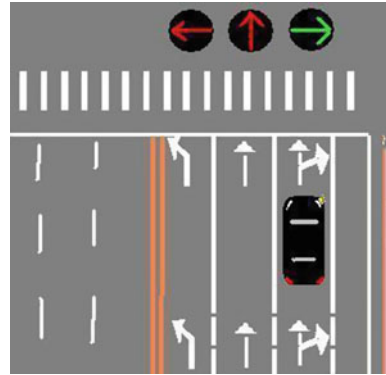


Fig. 2 Schematic diagram of vehicle driving on combined drive



$$\text{Lawful Driving}_1 = \{O_1.\text{speed} > 0 \text{ AND } O_1.\text{license number} = \text{A} \text{ AND } O_1.\text{turn right}\}$$

$$\text{Lawful Driving}_2 = \{O_1.\text{speed} > 0 \text{ AND } O_1.\text{license number} = \text{A} \text{ AND } O_1.\text{turn right}\}$$

According to the above information, the following acts are illegal:

$$\text{Illegal Driving}_1 = \{O_1.\text{speed} > 0 \text{ AND } O_1.\text{license number} = \text{A} \text{ AND } O_1.\text{turn left}\};$$

$$\text{Illegal Driving}_2 = \{O_1.\text{speed} > 0 \text{ AND } O_1.\text{license number} = \text{A} \text{ AND } O_1.\text{turn straight} \text{ AND } PRL_3\};$$

$$\text{Illegal Driving}_3 = \{O_1.\text{license number} = \text{A} \text{ AND } O_4.\text{solid line}\}.$$

5 Conclusions

This paper describes the semantic video understanding technology trends and research contents and methods are proposed based on VSD technology. Semantic information is related to the meaning of these elements and layouts, such as those information of the critical objects. These features are at a higher level description of the video content, but because of the lack of means of the unified representation and modeling of human eye domain knowledge, thus forming a gap between low-level grammar features and high-level semantic features, which is a difficult problem of semantic understanding faced in surveillance video. The application experiment proves the validity of semantic understanding model.

Acknowledgment This work was supported in part by National High-tech R&D Program of China (863 Program) under Grant 2013AA014604, and in part by the project of Shanghai Municipal Commission of Economy and Information under Grant 12GA-19.

References

1. Chen H, Ahuja N (2012) Exploiting nonlocal spatiotemporal structure for video segmentation. In: IEEE conference on computer vision and pattern recognition, pp 741–748
2. Javed K, Babri H, Saeed M (2012) Feature selection based on class-dependent densities for high-dimensional binary data. *IEEE Trans Knowl Data Eng* 24(3):465–477
3. Choi M, Torralba A, Willsky A (2012) A Tree-based context model for object recognition. *IEEE Trans Pattern Anal Mach Intell* 34(2):240–252
4. Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
5. Mei L, Cai X, Zhang H et al (2012) Video Structured description—vitalization techniques for the surveillance. *Video Data IFTC, CCIS* 331:219–227
6. Jiang Y, Xu Z, Chen H (2011) Semantic analysis on the knowledge map in the area of traffic violations. *Int J Distrib Sens Netw* 1–15
7. Xu Z, Liu Y, Mei L et al (2014) Semantic based representing and organizing surveillance big data using video structural description technology. *J Syst Software*. dx.doi.org/10.1016/jss.2014.07.024
8. Li J, Xu Z, Jiang Y et al (2014) An overview of extracting static properties of vehicles from the surveillance video. In: Proceedings of 2014 IEEE 13th international conference on cognitive informatics and cognitive computing, pp 317–322
9. Xu Z, Jiang Y, Li Z (2014) Construction and application of ontology in traffic surveillance video systems. *J Shanghai Univ (Nat Sci)* 20(5):658–669 (in Chinese)
10. Xu Z, Mei L, Liu Y et al (2013) Video structural description: a semantic based model for representing and organizing video surveillance big data. In: IEEE 16th international conference on computational science and engineering, pp 802–809

Research of Mining Multi-level Association Rule Models

Wen-Hsing Kao, Chin-Wen Lo, Kuo-Pin Li, Hsien-Wei Yang
and Jeng-Chi Yan

Abstract With the rapid development of Internet and the popularization of information technology and computers nowadays, data mining technology is becoming increasingly sophisticated. The main purpose of data mining is to obtain potentially relevant information from large databases correctly and efficiently. As to the association rule model, the main purpose of it is to find out possibly related product items. For example, with each transaction records in stores, we can dig out association rules like “80 % of customers that purchase PCs may also purchase screens.” The aim of this essay is to construct mining multilevel association rules and to analyze and discuss its integrity. The original multilevel association rules only explore associations at single concept level, so this essay will examine the integrity of multilevel association rules and use the original rules to find association relationships at multiple concept levels, coupling with the operation of filtering information from databases. The analysis of this essay can help companies in making marketing strategies and providing customized services to raise their overall sales.

1 Background and Motivation

A single transaction record tells us individual consumer behavior but we can understand overall consumption habits if gathering and analyzing a large number of transaction data. The purpose of association rules is to find out association relationships among items in each transaction. For example, we can find 80 % of customers that purchase PCs may also purchase screens. With respect to the multilevel association rules, it express association relationships at a lower concept level and provide more detailed information than the association rules do. The aim of this

W.-H. Kao (✉) · C.-W. Lo · H.-W. Yang · J.-C. Yan
Department of Information Technology, Overseas Chinese University, Taichung, Taiwan
e-mail: star@ocu.edu.tw

K.-P. Li
Department of Business Administration, Asia University, Taichung, Taiwan

essay is to construct the integrity of mining multilevel association rules and to analyze and discuss relevant limits and theories.

2 Literature Review

Apriori algorithm and the concept hierarchy are included in the literature review of this essay as follows.

2.1 Apriori Algorithm

Apriori algorithm is one of most classical algorithms which study association rules. Apriori algorithm finds associations among items in a given database step by step in order to form a rule, which is shown in Fig. 1.

2.2 The Concept Hierarchy of Merchandise

- (a) Although association rules at lower concept level present more detailed information than at high level, it may lead to the state of relative low support of itemsets [1].
- (b) If we want to find association rules at lower concept hierarchy, the minimum support must be lowered relatively.
- (c) The association rule at a higher level has higher support, but its outcome may be obviously predicted by common experiences.
- (d) We use concept hierarchies to perform multilevel mining and set different minimum support flexibly to different levels.

Fig. 1 The process of the Apriori algorithm

Apriori algorithm	
Step1 :	$L_1 \leftarrow \{\text{large } 1 - \text{itemsets}\}$;
Step2 :	for ($k = 2$; $L_{k-1} \neq \emptyset$; $k++$) do begin
Step3 :	$C_k = \text{Candidate_gen}(L_{k-1})$;
Step4 :	for each transactions t
Step5 :	$C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$ for candidates $c \in C_t$
	$\text{count}[c] \leftarrow \text{count}[c] + 1$;
Step6 :	$L_k \leftarrow \{c \mid c \in C_k \wedge \text{count}[c] \geq \epsilon\}$;
Step7 :	end
Step8 :	return $L = \bigcup_k L_k$

3 Method

The research process and steps in this paper are as follows:

1. Encode items:
If there are 4 itemsets, their concept hierarchies of items are represented in Figs. 2, 3, 4, 5 and the item ‘PCs laptops ASUS’ is encoded as ‘111’:
2. Recode items of transaction database T[1] according to their concept hierarchies, as shown in Table 1.
3. Use mining multilevel rules to compute itemsets and multilevel association rules, the basic spirit of which is introduced below.
 - (a) Use top-down progressive deepening method. First we calculate itemsets at level 1 and then itemsets at level 2 and so on until no itemsets are found.

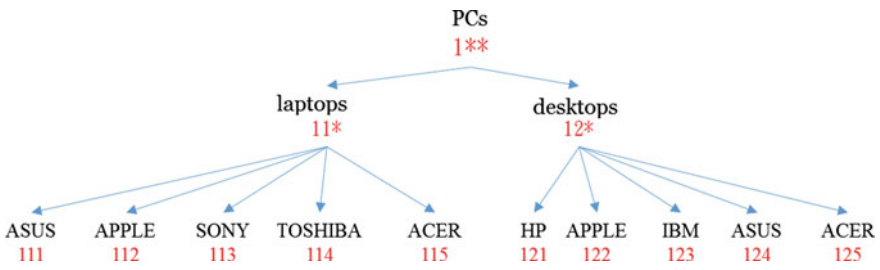


Fig. 2 Concept hierarchy of PCs

Fig. 3 Concept hierarchy of monitors

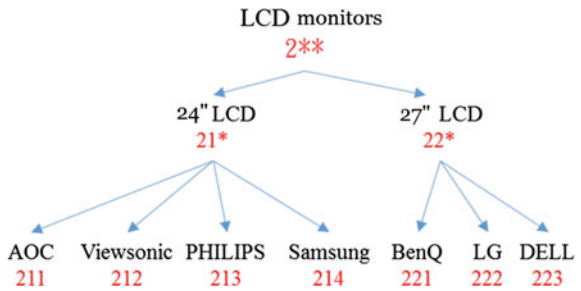


Fig. 4 Concept hierarchy of output devices

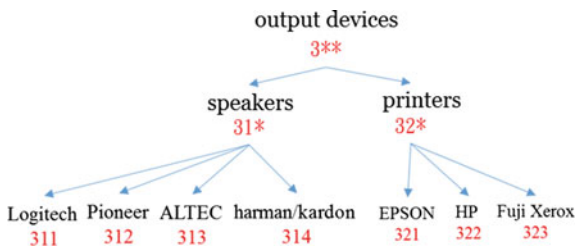


Fig. 5 Concept hierarchy of input devices

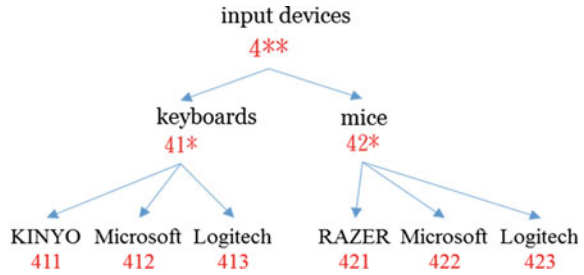


Table 1 Transaction database T[1]

Tid	Items
1	{111,211}
2	{113,313}
3	{112,423}
4	{122,212,321}
5	{125,214,311}
6	{113,313}
7	{115,321}
8	{123,221,322}
9	{412}
10	{411}

- (b) We can use Apriori algorithm to generate itemsets at any levels.
- (c) In the process of mining multilevel association rules, we set minimum support on current level and add requirements to limit it: item X at level I is considered when the parent node of X at level I-1 represents large items.
- (d) The multilevel mining considers the support of parent node: item X at level I is considered when its parent node at level I-1 is large 1-itemsets if not ignore it.

4 Analysis and Discussion of Mining Model

First, we calculate itemsets at a single level; next, calculate itemsets at multiple levels; finally, multilevel association rules will be generated completely. The process will be elaborated below:

4.1 Itemsets at the Same Level

At first, we read database and use Candidate_gen to generate large 1-itemset C [1,1] = {{1**},{2**},{3**},{4**}}. If minisup[1] = 4, we remove itemset{4**}. L [1,1] can be generated as shown in Table 2.

Table 2 Level-1 large 1-itemset L[1,1]

Itemsets	Support
{1**}	8
{2**}	5
{3**}	6

Table 3 Level-1 large 2-item set L[1,2]

Itemset	Support
{1**,2**}	5
{1**,3**}	6

Because {4**} does not comply with minimum support 4, we reduce transaction database T[1].

From L[1,1] we combine itemsets and obtain candidates 2-candidate itemset, C [1,2] = {{1**,2**},{1**,3**},{2**,3**}. However, support of {2**,3**} is 3, which does not comply with user-defined minisup[1] = 4, so we delete {2**,3**}. L[1,2] can be derived as shown in Table 3.

From L[1,2], candidate itemset C[1,3] = {1**,2**,3**} is generated and its minimum support = 3, which does not comply with user-defined support = 4, so L [1,3] = ∅ and so on at level-2 and level-3.

4.2 Large Itemsets of Merchandise Between Different Levels

Next, we discuss multilevel association rules completely, the purpose of which is to calculate the large itemsets of merchandise at multiple levels. In the beginning, we filter itemsets on every levels and delete any items which are not large. Back to the large 1-itemset in L[1,1], level-1 candidate item input devices {4**} ≤ minisup[1] so we delete it. Therefore, PCs {1**}, LCD monitors {2**} and input devices {3**} are only considered.

We, furthermore, consider the filtering of level-2. Level-2 large 1-itemset L[2,1] is {{11*},{12*},{21*},{22*},{31*},{32*}} so itemset {22*} is not considered. Due to deleting the parent nod{22*}, its child nod is not considered too. Then perform the same steps at level-3, the concept hierarchy of filtered merchandise is shown in Figs. 6, 7, 8.

Fig. 6 Filtered concept hierarchy of PCs

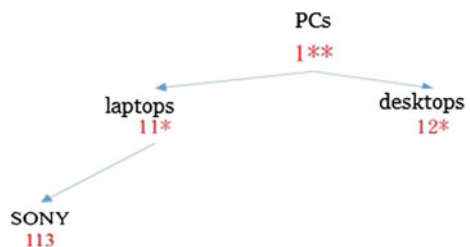


Fig. 7 Filtered concept hierarchy of LCD monitors

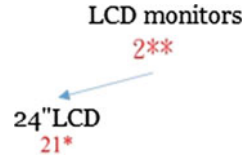
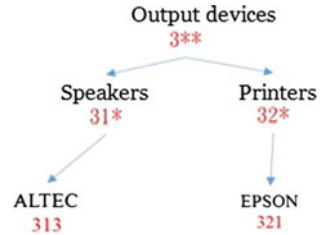


Fig. 8 Filtered concept hierarchy of output devices



We consider $L[1,2] = \{\{1^{**},2^{**}\}, \{1^{**},3^{**}\}\}$, so $\{1^{**}\}$ is related to $\{2^{**}\}$ at level-1 and $\{1^{**}\}$ is correlated with $\{3^{**}\}$. $\{2^{**},3^{**}\}$ is not large itemset and represents no association in additional requirements, so we only need to compute the association between $\{1^{**}\}$ and $\{2^{**}\}$ and the association between $\{1^{**}\}$ and $\{3^{**}\}$ without additional requirements.

Then consider associations in $L[2,2]$. $L[2,2] = \{\{11^*,31^*\},\{12^*,21^*\}, \{12^*,32^*\}\}$, 3 sets only, so compute the association between $\{11^*\}$ and $\{31^*\}$, $\{12^*\}$ and $\{21^*\}$, and $\{12^*\}$ and $\{32^*\}$, as shown in Fig. 9.

If $\text{minisup}[1] = 3$, we compute merchandise at level-1 and level-2 candidate 2-itemset. $C[1,2,2] = \{\{1^{**},21^*\},\{1^{**},31^*\},\{1^{**},32^*\},\{2^{**},11^*\},\{2^{**},12^*\}, \{3^{**},11^*\},\{3^{**},12^*\}\}$ support count of candidate itemset $\{2^{**},11^*\} < \text{minisup}[1]$, so we delete $\{2^{**},11^*\}$ and keep the rest to get large 1-itemset $L[1,2,2] = \{\{1^{**},21^*\},\{1^{**},31^*\},\{1^{**},32^*\},\{2^{**},12^*\},\{3^{**},11^*\},\{3^{**},12^*\}\}$, as shown in Table 4.

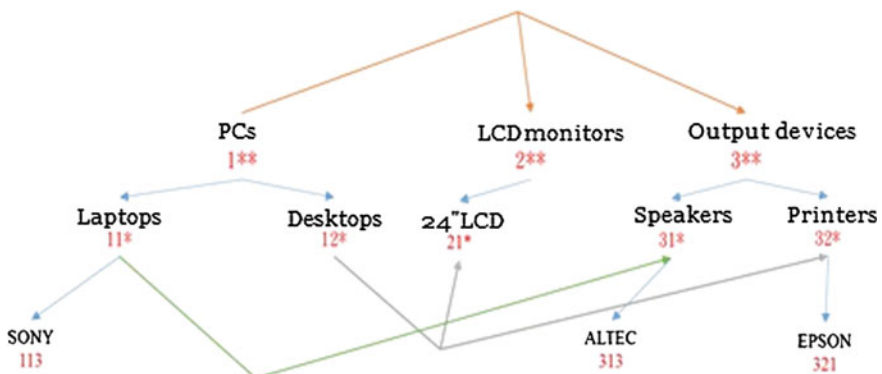


Fig. 9 Association of paired comparison

Table 4 Level-1 and level-2 large 2-itemset $L[1,2,2]$

Itemsets	Support
{1**,21*}	3
{1**,31*}	3
{1**,32*}	3
{2**,12*}	3
{3**,11*}	3
{3**,12*}	3

There are itemsets $\{ \{1**,21*\}, \{1**,31*\}, \{1**,32*\} \}$ and $\{ \{3**,11*\}, \{3**,12*\} \}$ in $L[1,2,2]$. If paired itemsets share the same first item, in accord with the condition of combination, they are combined into a candidate itemset $C[1,2,3] = \{ \{1**,21*,31*\}, \{1**,21*,32*\}, \{1**,31*,32*\}, \{3**,11*,12*\} \}$. Support count of $L[1,2,3] < \text{minisup}[1]$ so $L[1,2,3] = \emptyset$.

The association between level-1 and level-3 will be discussed below. Suppose $\text{minisup}[2] = 2$, according to the matching condition we can get candidate itemset $C[1,3,2] = \{ \{1**,313\}, \{1**,321\}, \{2**,113\}, \{3**,113\} \}$. We delete $\{2**,113\}$ of them because of its support count = 0, so we get large itemset $L[1,3,2] = \{ \{1**,313\}, \{1**,321\}, \{3**,113\} \}$, as follows in Table 5.

In $L[1,3,2]$, $\{1**,313\}$ and $\{1**,321\}$ share the same first item, corresponding to the condition of combination, so $C[1,3,3] = \{ \{1**,313,321\} \}$ is derived. Then compute support count of $\{1**,313,321\}$ and it is $0 < \text{minisup}[2]$; therefore, delete $\{1**,313,321\}$ and get $L[1,3,3] = \emptyset$.

Finally, compute the association between level-2 and level-3. Suppose $\text{minisup}[3] = 2$, compute combinations in level-2 and level-3 according to Fig. 9, and then we can get candidate itemset $C[2,3,2] = \{ \{11*,313\}, \{12*,321\}, \{31*,113\} \}$.

If $\text{minisup}[3] = 2$ is known and support count of candidate $\{12*,321\}$ in $C[2,3,2]$ is 1, we delete $\{12*,321\}$ and get $L[2,3,2] = \{ \{11*,313\}, \{31*,113\} \}$, as shown in Table 6.

Two itemsets in level-2 and level-3 large itemset have different first item so the process of combining them is unnecessary.

Table 5 Level-1 and level-3 large 2-itemset $L[1,3,2]$

Itemsets	Support
{1**,313}	2
{1**,321}	2
{3**,113}	2

Table 6 Level-2 and level-3 large 2-itemset $L[2,3,2]$

Itemsets	Support
{11*,313}	2
{31*,113}	2

4.3 Complete Multi-Level Association Rules

Suppose minimum confidence of level-1 = 0.8, association rules can be generated as follows:

$$\begin{array}{ll} \{1^{**}\} \rightarrow \{2^{**}\} & \text{confidence } 5/8 = 0.625 \\ \{2^{**}\} \rightarrow \{1^{**}\} & \text{confidence } 5/5 = 1 \\ \{1^{**}\} \rightarrow \{3^{**}\} & \text{confidence } 6/8 = 0.75 \\ \{3^{**}\} \rightarrow \{1^{**}\} & \text{confidence } 6/6 = 1 \end{array}$$

The second rule and fourth one comply with minimum confidence.

Suppose minimum confidence of level-2 = 0.7, association rules can be generated as follows:

$$\begin{array}{ll} \{11^*\} \rightarrow \{31^*\} & \text{confidence } 2/5 = 0.4 \\ \{31^*\} \rightarrow \{11^*\} & \text{confidence } 2/3 = 0.66 \\ \{12^*\} \rightarrow \{21^*\} & \text{confidence } 2/3 = 0.66 \\ \{21^*\} \rightarrow \{12^*\} & \text{confidence } 2/3 = 0.66 \\ \{12^*\} \rightarrow \{32^*\} & \text{confidence } 2/3 = 0.66 \\ \{32^*\} \rightarrow \{12^*\} & \text{confidence } 2/3 = 0.66 \end{array}$$

No association rule can comply with minimum confidence.

Suppose minimum confidence of level-3 = 0.8, association rules can be generated as follows:

$$\begin{array}{ll} \{113\} \rightarrow \{313\} & \text{confidence } 2/2 = 1 \\ \{313\} \rightarrow \{113\} & \text{confidence } 2/2 = 1 \end{array}$$

The first rule and second one comply with minimum confidence.

Suppose minimum confidence of level-1 and level-2 = 0.8, association rules can be generated as follows:

$$\begin{array}{ll} \{1^{**}\} \rightarrow \{21^*\} & \text{confidence } 3/8 = 0.375 \\ \{21^*\} \rightarrow \{1^{**}\} & \text{confidence } 3/3 = 1 \\ \{1^{**}\} \rightarrow \{31^*\} & \text{confidence } 3/8 = 0.375 \\ \{31^*\} \rightarrow \{1^{**}\} & \text{confidence } 3/3 = 1 \\ \{1^{**}\} \rightarrow \{32^*\} & \text{confidence } 3/8 = 0.375 \\ \{32^*\} \rightarrow \{1^{**}\} & \text{confidence } 3/3 = 1 \\ \{2^{**}\} \rightarrow \{12^*\} & \text{confidence } 3/5 = 0.6 \\ \{12^*\} \rightarrow \{2^{**}\} & \text{confidence } 3/3 = 1 \\ \{3^{**}\} \rightarrow \{11^*\} & \text{confidence } 3/6 = 0.5 \\ \{11^*\} \rightarrow \{3^{**}\} & \text{confidence } 3/5 = 0.6 \\ \{3^{**}\} \rightarrow \{12^*\} & \text{confidence } 3/6 = 0.5 \\ \{12^*\} \rightarrow \{3^{**}\} & \text{confidence } 3/3 = 1 \end{array}$$

The 2nd, 4th, 6th, 8th and 12th association rules comply with minimum confidence.

Suppose minimum confidence of level-1 and level-3 = 0.8, association rules can be generated as follows:

$$\begin{aligned}
 \{1^{**}\} &\rightarrow \{313\} && \text{confidence level } 2/8 = 0.25 \\
 \{313\} &\rightarrow \{1^{**}\} && \text{confidence level } 2/2 = 1 \\
 \{1^{**}\} &\rightarrow \{321\} && \text{confidence level } 2/8 = 0.25 \\
 \{321\} &\rightarrow \{1^{**}\} && \text{confidence level } 2/2 = 1 \\
 \{3^{**}\} &\rightarrow \{113\} && \text{confidence level } 2/6 = 0.33 \\
 \{113\} &\rightarrow \{3^{**}\} && \text{confidence level } 2/2 = 1
 \end{aligned}$$

The 2nd, 4th and 6th association rules comply with minimum confidence.

Suppose minimum confidence of level-2 and level-3 = 0.8, association rules can be generated as follows:

$$\begin{aligned}
 \{11^*\} &\rightarrow \{313\} && \text{confidence level } 2/5 = 0.4 \\
 \{313\} &\rightarrow \{11^*\} && \text{confidence level } 2/2 = 1 \\
 \{31^*\} &\rightarrow \{113\} && \text{confidence level } 2/3 = 0.66 \\
 \{113\} &\rightarrow \{31^*\} && \text{confidence level } 2/2 = 1
 \end{aligned}$$

The 2nd rule and 4th one comply with minimum confidence.

At last, we replace merchandise code with their real names and obtain the association rules, as shown in Tables 7, 8.

Table 7 Association rule of merchandise at a single level

Association rule	Confidence
LCD monitors → PCs	1
Output devices → PCs	1
SONY laptops → ALTEC speakers	1
ALTEC speakers → SONY laptops	1

Table 8 Association rule of merchandise at multiple levels

Association rule	Confidence
24" LCD → PCs	1
Speakers → PCs	1
Printers → PCs	1
Desktops → LCD monitors	1
Desktops → Output devices	1
ALTEC speakers → PCs	1
EPSON printers → PCs	1
SONY laptops → Output devices	1
ALTEC speakers → Laptops	1
SONY laptops → Speakers	1

5 Future Research

A single transaction record tells us individual consumer behavior but we can analyze overall consumption habits after collecting huge amounts of transaction data. The purpose of mining association rules is to find out possibly related items in every transactions, such as digging out “80 % of customers that purchase PCs may also purchase screens.” Multilevel association rules express relationships among items at lower concept hierarchies and provide more detailed information than association rules do.

This essay is mainly about constructing the integrity of mining multilevel association rule models. We use original multi-level association rules to study and correct the deduced results. The original association rules are about relationships among items at a single level from different product groups and provide more detailed information than general association rules do. The essay changed items at a single level from different product groups into at multiple levels and in different product groups. We obtain associations at multiple levels after the process of filtering. The information, which comes from analysis in more complete mining multi-level association rule models, can help companies make more effective marketing strategies and offer customized services to raise overall sales.

Also, the integrity of mining quantitative multi-level association rule models is interesting topic for future study.

Reference

1. Natek S, Zwilling M (2014) Student data mining solution—knowledge management system related to higher education institutions. *Expert Syst Appl* 41(14):6400–6407

Research of the Dimension Combination Strategy Model

Bo-Shen Liou, Ruei-Yang Lin, Kuo-Pin Li, Wen-Hsing Kao
and Jeng-Chi Yang

Abstract This study quantifies the usage and evaluation of Data Reduction. Within Data Reduction, there are three different measurement of methods: association measurement, discrimination measurement, and information measurement. Through analysis of the importance of each measurement stage, we generated sequences of forward generation to select the best combination of Data Reduction. The purpose of the sequences of forward generation is to increase efficiency and accuracy from the selected combination of Data Reduction. Based on the method of generating our model, we want only a single field to appear, in order to measure the amount of information based on the most suitable model law for the three measurement methods. The purpose of this model is to allow users of data mining to explore the selected field, in addition to the single characteristic attribute field as a reference, but also according to different dimensions of the resulting combination of all the chaos of the target attributes and how they affect the relationship, so that users can analyze and use the field to solve the most troublesome mining field dimension selections.

Keywords Data mining · Data reduction · Information measurement

1 Research Motivation and Purpose

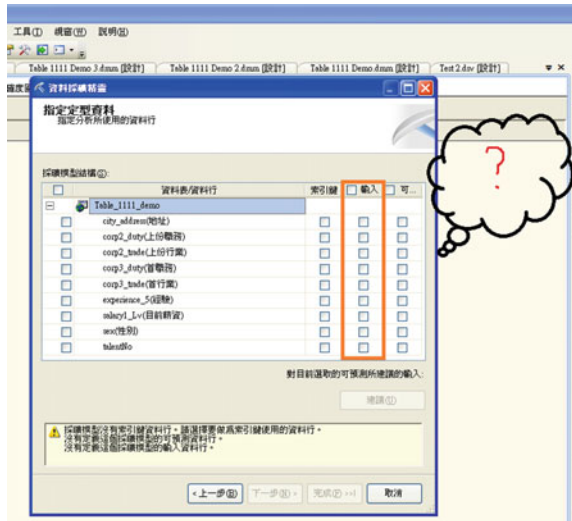
When the data analyst runs the selection of Data Reduction from data mining, he or she will tend to get exhausted during the selection process.

From Fig. 1, it will be time-consuming to figure out which fields and suitable column dimension to select. In order to allow data mining users to reduce the time for this segment, the target of this study will be on how to choose the most correct column dimensions and combinations.

B.-S. Liou · R.-Y. Lin · W.-H. Kao (✉) · J.-C. Yang
Department of Information Technology, Overseas Chinese University, Taichung, Taiwan
e-mail: star@ocu.edu.tw

K.-P. Li
Department of Business Administration, Asia University, Taichung, Taiwan

Fig. 1 Data reduction from data mining to get exhausted during the selection process



In this study, the use of Data mining related to 1111 Job Resource Bank’s database as an information source, the actual dimensions of the combined analysis of data to analyze the true of the various dimensions of dimension in achieving the following objectives:

1. The use of pre-processing data mining is to related to technologies, materials dimensionality reduction in the connection resistance, the ability to identify, measure the amount of information law as a model based on a combination of data dimensions, and with rules of thumb including policies and progressive selection methods to ensure the user does not waste time. It will be treated according to the data repository dimensional combination model to build our operations.
2. Choose from a single field characteristic attributes through a single result of our model.
3. Select all fields or a combination pick and choose the desired results through our complete model.
4. From those three models, a single field results and complete dimensions combined results generated by the model, to select the most suitable model.

2 Discussion Document

Documents of this paper to explore the theory of portfolio strategies including data dimensions, are described as follows.

2.1 Streamlining Data

With the evolution of time and an increasing amount of data, a user spends relatively longer time in exploration to mine the quality and usefulness of the application, but ends up with results that are not as good as in the past. Thus, streamlining data (Data Reduction) has a part to play in the process of data mining, which is to set data selection, filter out the required information, and reduce the time and cost required for data mining. Pre-positioning in the data processing stage, exploration stage and the post-processing phase three stages can be used to streamline information technology, and this paper will primarily use the information to streamline the data pre-processing stage. This data can be found from your library or storage link, where you may select the necessary information and create a user set, and from the combined set of information, filter out irrelevant deviation, null, or repeat information. Of course, streamlining data can be regarded as one of the important pre-processing applications, which contains data dimensionality reduction, data recording and data value streamlining, and this study uses data dimensionality reduction technology.

2.2 Connected Resistance, the Ability to Identify and Measure the Amount of Information Law

1. Correlation measurement concept

Connects the dimensions of the measurement data, between two data dimensions A and B. The higher the degree of correlation, the data value of A can determine the possibility of B in the value of the higher data transmission connection of measurements. Thus through this analysis, you can gain the correlation between subject's data dimension data sheet and its relations' dimension with the other of the data even in degrees.

2. The ability to identify the concept of measurement method

Through the measurement data table and other information dimensions, such as dimension of information, we have the ability to identify the value of information and the ability to identify higher dimension in data information for the underlying dimensions of more importance and influence.

3. Measuring the amount of information law concept

Through the measurement of data collection in other dimensions of information to gain the underlying profit, the limited information provided is mainly used to determine the value of the data record bid. The more data from the amount of information provided by Dimension Data, the higher the value of its importance. Without considering other dimensions of data, the random value formula underlying data dimensions are as follows:

$$E(D) = - \sum_{i=1}^d P_D(c_i) \log_2 P_D(c_i) \tag{1}$$

$P_D(c_i)$ represents the probability values c_i dimension data, and d represents the size of the target range dimension information.

In considering a combination of X dimension data, the value of the underlying data chaos dimension.

$$E(D_j^x) = - \sum_{i=1}^d P_{D_j}(c_i) \log_2 P_{D_j}(c_i) \tag{2}$$

$P_{D_j}(c_i)$ represents that dimension X indicates when the data is data value j , the underlying data dimensions, c_i probability value data and d represent the size of the underlying data dimensions range.

Calculating the combination of dimensions— X dimension of the underlying data information to profit.

$$IG(X) = E(D) - \sum_{j=1}^p \frac{|D_j|}{|D|} E(D_j^x) \tag{3}$$

p represents the size range of a combination of X dimension data.

2.3 Progressive Selection Method

When looking for the best information in line with the termination conditions or combination of dimensions, a combination of shorter dimension data is assessed by a combination of the longer dimension of the information, and the information we have at this time will determine the dimension combinations resulting in the sequential methods listed below. In order to produce the information dimension combination aspect, the present paper uses the progressive selection method. {} Lattice from below to above the lattice, each time requiring more consideration to calculate the dimensions of a single information data, using first layer lattice dimensions based on the selection of the measurement method {"Candidates' sex"}, {"work experience"} and {"living area "}, and pick the best of the information dimension, as shown in Fig. 2.

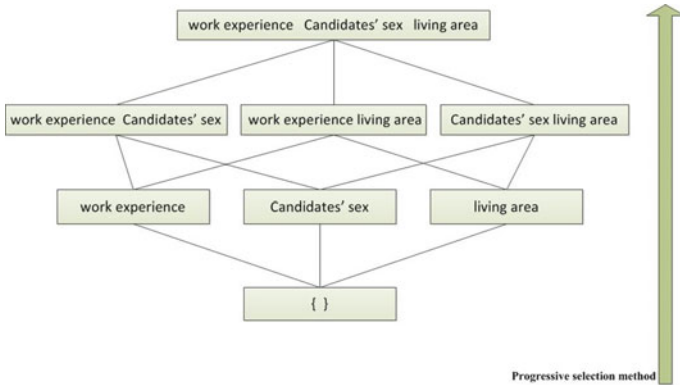


Fig. 2 Progressive selection method

3 Research Methods

In this paper, using the 1111 Job Resource Bank’s 2010 database as an example, we hope to find a combination of dimensions affecting salary-related fields. The first is pre-processing of data mining dimensionality reduction of measurement through the data, including the connection resistance, the ability to identify and measurement method to calculate the amount of information in each field of random values, then determining the importance of each field, and then step strategy and progressive selection method which is based on the rule of thumb list, to find the best combination of data dimensions. The results were analyzed and discussed, and we wrote the program and produced the relevant rules, with the model architecture as shown in Fig. 3.

(a) 1111 Job Bank Database

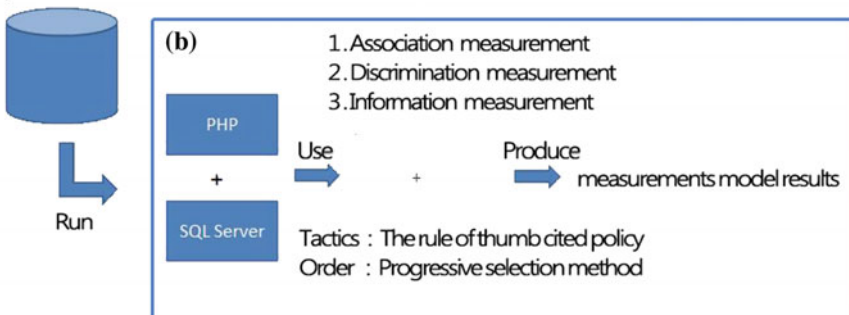


Fig. 3 Model architecture

4 Implementing Results and Discussion

After a long period of testing and adjustment, model accuracy and speed have reached a certain standard, and have been repeatedly tested by the process to improve the quality of the model. The results will be analyzed to verify the following references.

4.1 Single-Column Model and Compare the Results of Analysis

- 1. Connected single field of measurement results

Shown in Fig. 4, where we can find it in the first three individual fields of 1.sex (gender) 2.corp3_duty (first position) 3.corp2_trade (the parts industry).

- 2. The ability to identify a single field of measurement results

Figure 5 below, where we can find it in the first three individual fields of 1.sex (gender) 2.corp3_duty (first position) 3.corp2_trade (upper parts industry).

- 3. The amount of information results in a single field measurement

Figure 6 below, where we can find it in the first three individual fields of 1. corp2_trade (the parts industry) 2.corp3_duty (first position) 3.corp2_duty (duties on parts).

Fig. 4 Connected single field of measurement results



Fig. 5 The ability to identify a single field of measurement results



Fig. 6 The amount of information results in a single field measurement

資料維度之組合策略分析如下

1	{sex(性別)}	0.079
2	{corp3_duty(首職務)}	0.024
3	{corp2_trade(上份行業)}	0.024
4	{corp3_trade(首行業)}	0.007
5	{corp2_duty(上份職務)}	0.007
6	{experience_5(經驗)}	0.005
7	{city_address(地址)}	0.001

返回重算 顯示折線圖 下一層 全展開

Fig. 7 SQL Server features the single field attribute results

建議相關資料行

資料行名稱	分數	輸入
experience_5(經驗)	0.058	x
corp2_duty(上份職務)	0.042	
corp3_duty(首職務)	0.036	
corp2_trade(上份行業)	0.027	
corp3_trade(首行業)	0.027	
sex(性別)	0.018	
city_address(地址)	0.016	

4. SQL Server features the single field attribute results

We use SQL Server to perform real results generation in Fig. 7, where we found that the top three individual fields of 1. Experience_5 (experience) 2. Corp2_duty (sake duties) 3. corp3_duty (first position). Its model of compliance: the amount of information > Connected sex discrimination capability =

4.2 Complete Field Analysis and Comparison of the Results of the Model

1. Connected to full field measurement method analysis

From Fig. 8 below, from the bar chart view, we are unable to identify the cause, but we can find a line graph showing the head in this figure, the high end of the case, this case may be due to the data pre-processing being not clean enough.

2. The ability to identify the complete measurement field analysis

From Fig. 9 below, the bar chart view, you will find there are more and more based on the increased dimension of lower trend line chart from the point of view that we can find. We later presented a convergence of state, that is, we ignore multiple dimensions with a degree, in order to save time.

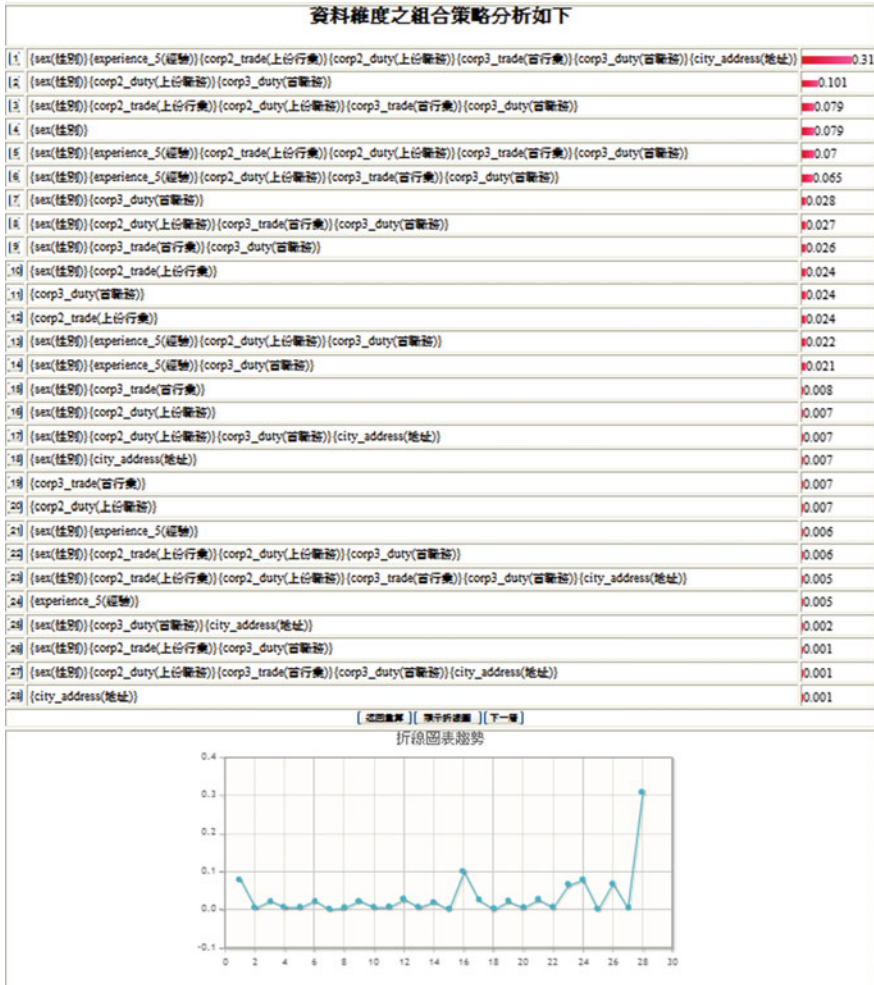


Fig. 8 Connected to full field measurement method analysis result

3. Information quantity measurement complete field analysis

From below Fig. 10, with the long bar chart view, you will find that there is an increasing trend in the increased dimension, while the line chart presents a convergence of state that is to return dimension to a multi-level, that we will be not be considering, in order to save time.

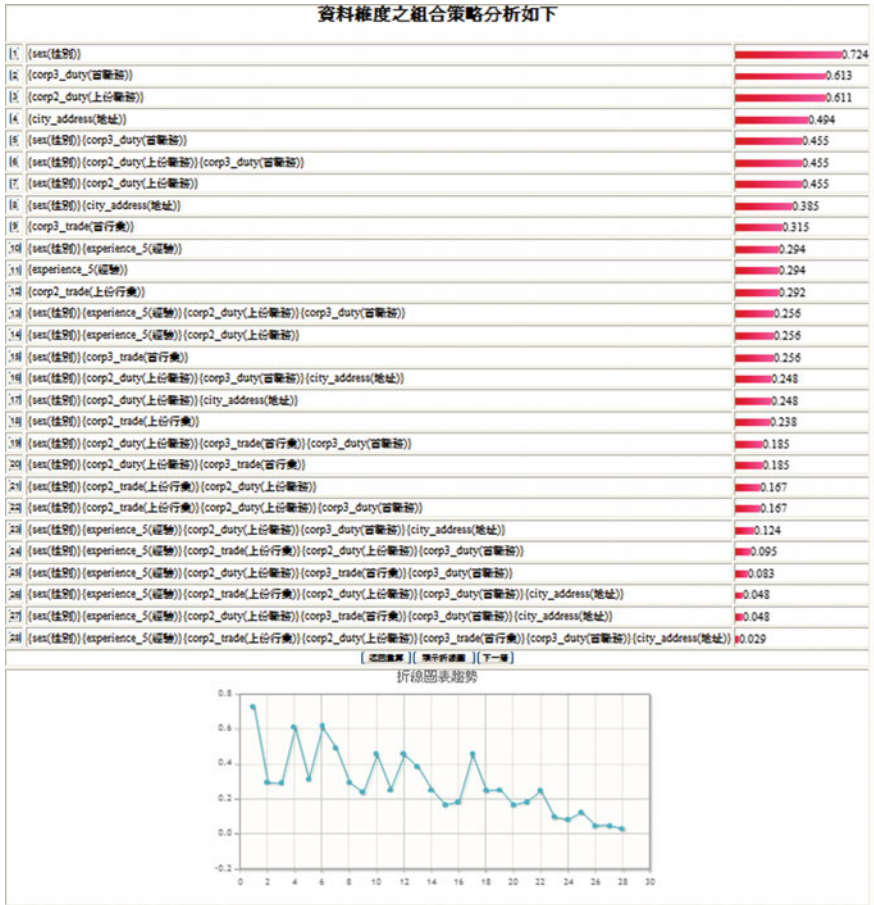


Fig. 9 The ability to identify the complete measurement field analysis result

According to our comparison of the complete field dimension bar charts and line charts whether a line graph converges to obtain individual model applicability, our results are found in the following Table 1.

Finally, we found that the accuracy of this model for the measurement of the amount of information > discriminating > Connected based on the comparison of the aforementioned.

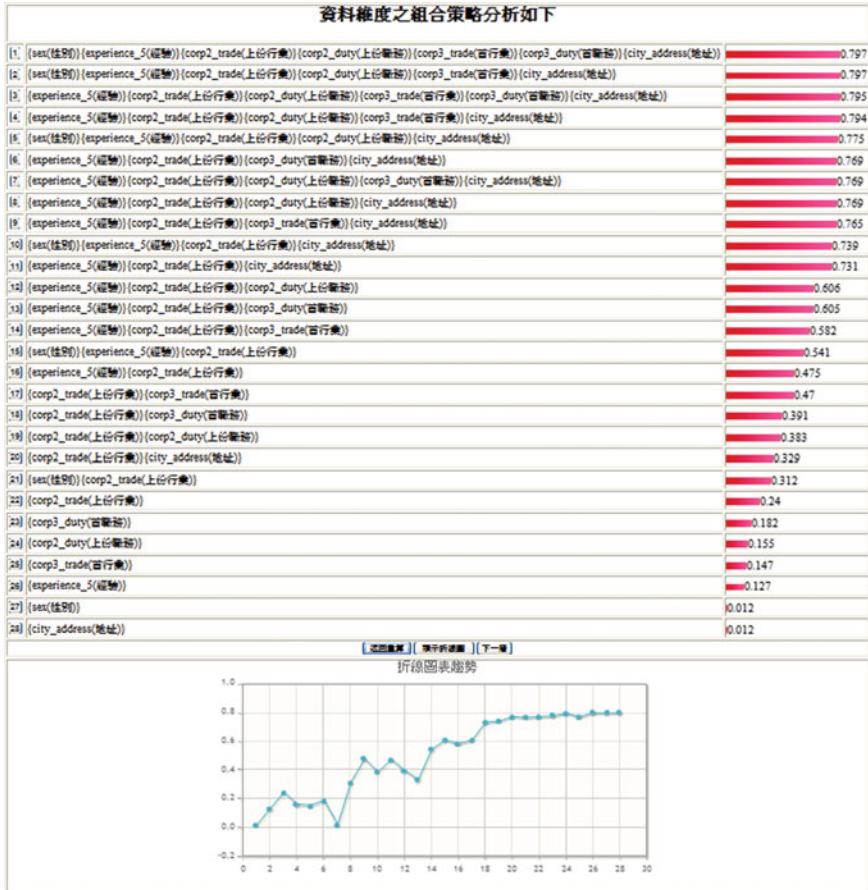


Fig. 10 Information quantity measurement complete field analysis result

Table 1 Comparison of the complete field dimension

	Association	Discrimination	Information
Bar charts law	X	O	O
Line chart converge to whether	X	O	O

5 Conclusions and Future Research Directions

In this paper, the 1111 Job Resource Bank cooperation has made its database of job seekers and job seekers with information relevant as a model. As the information is not complete as part of the cleanup, we used multiple methods including the correlation measurement method, the ability to identify the measurement method, the rule of thumb with measurements including policies and progressive selection

method, and the establishment of three measurement-based models, after the model, in order to identify the most influential model of our line graph data in three dimensions combined model. The length of the bar chart of the data for analysis and comparison of the results of a single field generated by the model, the complete dimensions of combined results, and the use of the decision tree model scoring results comments comparison, allowed us to select the most suitable model for 1111 Job Bank database, maximizing the amount of information based on the measurement data of the dimensional model portfolio strategy. In order to allow data mining users to speed up all kinds of information to understand the relationship between the dimension combinations, we applied data dimensionality reduction techniques to the 1111 Job Bank database as an information source, and successfully achieved the following objectives:

1. Using pre-processing data mining related technologies, materials dimensionality reduction in the connection resistance, the ability to identify, measuring the amount of information law as a model based on a combination of data dimensions, and with rules of thumb including policies and progressive selection method to reduce time wastage to treat accordingly to the data repository dimensional combination model to build our operations.
2. Choose from a single field characteristic attributes through a single result of our model.
3. Select All fields or a combination pick and choose the desired results through our complete model.
4. From our three models, a single field results and complete dimensions combined results are generated by the model, for us to select the most suitable model.

This study uses associated resistance, ability to identify, and measures the amount of information law to build dimensional data model portfolio analysis. We hope that the future strategy in the data combination of dimensions can be added to different measurement methods of analysis and discussion, in order to increase the dimensions of the data Portfolio Analysis rich model. In addition, to try to use the hybrid selection method, to further reduce user waiting time.

References

1. Su JH, Lin WY (2004) CBW: an efficient algorithm for frequent itemset mining. In: Proceedings of the 37th Hawaii international conference on system sciences, pp 9
2. Berry MJA, Linoff GS (2003) Data mining techniques: for marketing, sales, and customer support. Wiley, New York
3. Tong Y, Chen L, Cheng Y, Yu PS (2012) Mining frequent itemsets over uncertain databases. Proc VLDB Endowment 5(11):1650–1661
4. Liu B, Zhang L (2012) A survey of opinion mining and sentiment analysis. Springer, New York
5. Bramer M (2013) Principles of data mining. Springer, London

Short Latency Bias in Latency Matrix Completion

Cong Wang, Min LI and Yan Yang

Abstract For latency-sensitive applications, a key issue is how to estimate the latencies between any couple of nodes. Latency Matrix Completion method provides a simple but efficient way to estimate the latencies instead of measure them directly. In this paper, we make comparative studies on several Internet latency data sets, and report an easy overlooked shortcoming exists in Latency Matrix Completion. For short latencies, their relative estimation errors are much higher than those of long latencies. In this paper, we propose a brief model to analyze why this bias exists. We believe that the loss function which used in the optimizing process is a possible reason for this phenomenon. How to remove the short latency bias should cause our consideration in the future.

Keywords Matrix completion · Optimization · Network coordinate system

1 Introduction

LATENCY-SENSITIVE applications play an important part in today's Internet. Many kinds of applications, such as P2P Networks [1], Online games [2, 3], and even The Onion Router (Tor) [4–6] are all latency sensitive. To improve the application performance, a key issue is how to estimate the latency between each couple of nodes. Ping or traceroute operations get all the latencies of the system but

C. Wang (✉)

Digital Media College, Sichuan Normal University, Chengdu, Sichuan, China
e-mail: wongcong@gmail.com

M. LI

School of Computer, Sichuan Normal University, Chengdu, Sichuan, China
e-mail: lm_turnip@126.com

Y. Yang

College of Mechanical and Electrical Engineering, Chongqing University
of Arts and Sciences, Chongqing, China
e-mail: yangzicb@163.com

the complexities of these measure methods are up to $O(n)^2$. Thus these methods cannot update the latency information real-timely while the system scale is too large [7]. It has been widely concerned about how to estimate latencies from limited measurement data accurately. Early studies of the latency estimating problem embeds all the nodes of the applications into a specific metric space, then the latency between any couple of nodes can be replaced by the spatial distance between this couple of node. This kind of studies is well known as Network Coordinate System, NCS. NCS have been deployed in many Internet applications [8] because of its acceptable accuracy. A potential but not appropriate assumption of NCS is that nodes can be embedded into the metric space ideally. Recent studies show that the Internet latency matrix is approximately sparse [9]. The research result of [10] indicates that NCS is a special kind of completion method of Internet latency matrix.

In full decentralized environment, each node i can be assigned a n -dimensional row vector \mathbf{u}_i and a n -dimensional column vector \mathbf{v}_i , then the empty entry $d_{i,j}$ in the latency matrix, e.g. the real latency between node i and j , can be calculated approximately by $\mathbf{u}_i \cdot \mathbf{v}_j$. Comparing with NCS, this idea has three advantages: At first, it can be translated into a couple of convex problems to solve; Secondly, the triangle inequality violations problem, TIVs can be easily handled since the restricted condition $d_{a,b} + d_{b,c} \geq d_{a,c}$ is no longer needed; At last, the completed matrix is not necessarily to be symmetric, thus the asymmetry of the Internet latencies can be expressed well. IDES [11] is the first algorithm derived from this idea. But an ill-posed matrixes inversion computing process is inevitable because of the multi-manifold property of the Internet latency space [12], thus the accuracy of the latency estimation is dropped significantly. In DMF [13] and DMFSGD [14] algorithm a regularization factor is imported to get a biased estimation of the latency, which improves the robustness significantly. On the basis of DMF, Phoenix algorithm proposes a non-negative ensuring scheme to eliminate the negative latency estimation [15].

This paper reports a potential shortcoming, e.g. the short latency bias of latency matrix completion algorithms. This shortcoming exists in all the methods mentioned above even though acceptable estimation accuracy can be provided by them. Firstly we establish a model to illustrate how this shortcoming is produced theoretically, and find out a possible reason of this shortcoming. And then, we prove that this shortcoming is widely existed in latency matrix completion algorithms through our experiments thus it cannot be ignored, and need to be studied deeply.

2 An Overview of Latency Matrix Completion

In this section let's review the fundamental idea of the latency matrix completion problem firstly: For an application which contains P nodes, there exists a latency matrix \mathbf{D} with $P \times P$ entries. Each element $d_{i,j}$ of \mathbf{D} represents the latency between

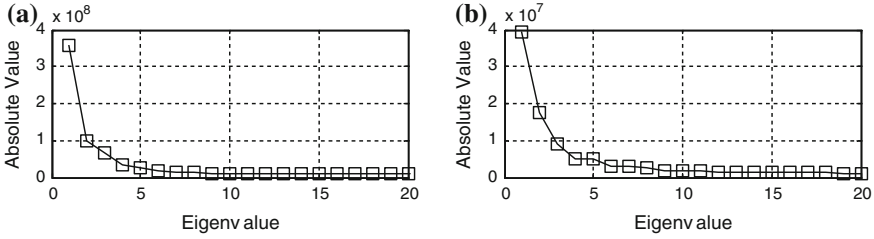


Fig. 1 The 20 largest eigenvalues of the latency matrixes of King and Planetlab datasets. **a** King dataset **b** Planetlab dataset

node i and j . In general, node can hardly get the latencies between all other nodes and themselves in real time. As a result we can only get an incomplete matrix \mathbf{D}' since some entries in \mathbf{D} cannot be obtained. But the approximately sparsity of the latency matrix makes it possible to complete this matrix: Most of the eigenvalues of the matrix are close to zero. Figure 1 shows the biggest 20 eigenvalues of the latency matrixes of Planetlab [16] and King [17] datasets respectively.

We can complete \mathbf{D}' to get a completed matrix $\hat{\mathbf{D}}$ as a good fitting of \mathbf{D} . Each entry $d_{i,j}$ in \mathbf{D} can be estimated by the corresponding entries $\hat{d}_{i,j}$ in $\hat{\mathbf{D}}$.

In full decentralized environment, it is difficult to extract a global consistent latency matrix. But in consideration of the approximate sparsity of the latency matrix \mathbf{D} , in general it is much easier to get a consensus about the prior estimation N of the l_0 norm of $\hat{\mathbf{D}}$ to guarantee the rigid sparsity of $\hat{\mathbf{D}}$. Then the matrix can be completed by solving the following optimization problem:

$$\min_{\hat{\mathbf{D}}} (L(P_{\Omega}(\hat{\mathbf{D}} - \mathbf{D}'))), s.t. \|\hat{\mathbf{D}}\|_0 = N \tag{1}$$

where $P_{\Omega}(\cdot)$ extracts those entries which exist in \mathbf{D}' and the corresponding entries in $\hat{\mathbf{D}}$ to compute; $L(\cdot)$ is the pre-defined loss function; $\|\cdot\|_0$ represents the l_0 norm of a matrix, e.g. the number of non-zero eigenvalues. $\hat{\mathbf{D}}$ can be expressed as a multiplier of two $P \times N$ matrixes \mathbf{U} and \mathbf{V} :

$$\hat{\mathbf{D}} = \mathbf{U} \cdot \mathbf{V}^T \tag{2}$$

Hence $d_{i,j}$ in \mathbf{D} can be estimated by the corresponding entry $\hat{d}_{i,j}$ in $\hat{\mathbf{D}}$:

$$d_{i,j} \approx \hat{d}_{i,j} = \mathbf{u}_i \cdot \mathbf{v}_j^T = u_{i,1} \times v_{j,1} + \dots + u_{i,N} \times v_{j,N} \tag{3}$$

where $u_{i,n}, v_{j,n}$ is the n -th corresponding entry of \mathbf{u}_i and \mathbf{v}_j . \mathbf{U} and \mathbf{V} can be maintained by all nodes: We can assign each node i the corresponding vectors \mathbf{u}_i and \mathbf{v}_i , then node i can estimate the latency between any other node j and itself by getting

the vectors which maintained by j . Correspondingly, Eq. (1) can be factorized to the following couple of sub-problems:

$$\left. \begin{aligned} \min_{\mathbf{u}_i} & \left(\sum_{\langle i,j \rangle \in \Omega} l(d_{i,j} - \mathbf{u}_i \cdot \mathbf{v}_j^T) \right) \\ \min_{\mathbf{v}_i} & \left(\sum_{\langle j,i \rangle \in \Omega} l(d_{j,i} - \mathbf{u}_j \cdot \mathbf{v}_i^T) \right) \end{aligned} \right\} \quad (4)$$

The global optimum of this couple of problems can be obtained numerically if $l(\cdot)$ is a convex function. To enhance the robustness of the numerical solutions, a regularized factor can be introduced:

$$\left. \begin{aligned} \min_{\mathbf{u}_i} & \left(\lambda \mathbf{u}_i \cdot \mathbf{u}_i^T + \sum_{\langle i,j \rangle \in \Omega} l(d_{i,j} - \mathbf{u}_i \cdot \mathbf{v}_j^T) \right) \\ \min_{\mathbf{v}_i} & \left(\lambda \mathbf{u}_j \cdot \mathbf{u}_j^T + \sum_{\langle j,i \rangle \in \Omega} l(d_{j,i} - \mathbf{u}_j \cdot \mathbf{v}_i^T) \right) \end{aligned} \right\} \quad (5)$$

Obviously that couple of optimization problems is also a convex one though the regularized factors $\lambda \mathbf{u}_i \mathbf{u}_i^T$ and $\lambda \mathbf{u}_j \mathbf{u}_j^T$ are introduced. It can be easily solved by the alternating direction method of multipliers (ADMM). DMFSGD algorithm proposes the alternating direction sub-gradient descending method to update \mathbf{u}_i and \mathbf{v}_i alternatively:

$$\left. \begin{aligned} \mathbf{u}_i^{(t+1)} &= (1 - \eta_u \lambda) \mathbf{u}_i^{(t)} + \eta_u \sum_{\langle j,i \rangle \in \Omega} \frac{\partial l(\mathbf{u}_i)}{\partial \mathbf{u}_i} \\ \mathbf{v}_i^{(t+1)} &= (1 - \eta_v \lambda) \mathbf{v}_i^{(t)} + \eta_v \sum_{\langle j,i \rangle \in \Omega} \frac{\partial l(\mathbf{v}_i)}{\partial \mathbf{v}_i} \end{aligned} \right\} \quad (6)$$

where η_u and η_v are the sub-gradient descending lengths, t is the timestamp.

If the non-negative property needs to be ensured, the optimizing problem turns to be:

$$\left. \begin{aligned} \min_{\mathbf{u}_i \in R^{+N}} & \left(\lambda \mathbf{u}_i \cdot \mathbf{u}_i^T + \sum_{\langle i,j \rangle \in \Omega} l_2(d_{i,j} - \mathbf{u}_i \cdot \mathbf{v}_j^T) \right) \\ \min_{\mathbf{v}_i \in R^{+N}} & \left(\lambda \mathbf{u}_j \cdot \mathbf{u}_j^T + \sum_{\langle j,i \rangle \in \Omega} l_2(d_{j,i} - \mathbf{u}_j \cdot \mathbf{v}_i^T) \right) \end{aligned} \right\} \quad (7)$$

The original and the completed latency matrixes of Planetlab and King Datasets are shown in Fig. 2. It can be seen easily the completed matrixes are similar with the originals.

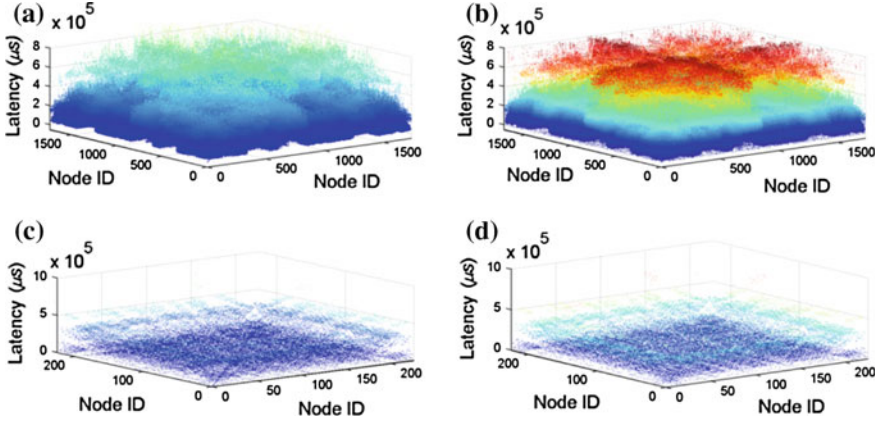


Fig. 2 The comparisons of original and completed latency matrices. **a** The original latency matrix of King **b** The completed latency matrix of King **c** The original latency matrix of Planetlab **d** The completed latency matrix of Planetlab

3 Theoretical Analyze of Short Latency Bias

Latency matrix completion improves the performance of latency-sensitive applications significantly, and be widely deployed in many applications. For this kind of applications, the estimation accuracy of short latency generally much important than that of long latency since nodes always communicate with their neighbors. But during the computing process of completion algorithm, a potential problem is that short and long latencies are treated equally by algorithms, this mechanism may leads to a bias, e.g. the relative estimation accuracy of short latency always much higher than that of long latency. For an instance, for a link with 500 ms latency, 10 ms estimation error can be seemed as very small, but for a link with 10 ms latency, 10 ms estimation error is very high. We believe that l_1 or l_2 loss functions which are widely used in almost all the algorithms maybe partly lead to this phenomenon. Take l_2 loss function for example, our deduce is shown as following:

Under this situation, Eq. (1) can be rewritten as:

$$\min_{\mathbf{D}} \left(\sum_{\langle i,j \rangle \in \Omega} (e_{i,j})^2 \right), e_{i,j} = \text{abs}(d_{i,j} - \mathbf{u}_i \cdot \mathbf{v}_j^T) \quad (8)$$

For discussion purposes, an equality constraint can be added:

$$E = \sum_{\langle i,j \rangle \in \Omega} (e_{i,j}) \quad (9)$$

To solve this constraint optimization problem, the Lagrange Multiplier method can be used, thus we can get:

$$\min_{\mathbf{D}}(F) = \min_{\mathbf{D}} \left(\sum_{\langle i,j \rangle \in \Omega} (e_{ij})^2 + \lambda \left(E - \sum_{\langle i,j \rangle \in \Omega} (e_{ij}) \right) \right) \quad (10)$$

While the optimum is gotten, for each e_{ij} , we have:

$$\frac{\partial F}{\partial e_{ij}} = 2e_{ij} - \lambda = 0 \quad (11)$$

Then we can get:

$$\forall e_{ij} \rightarrow e_{ij} = \frac{\lambda}{2} \quad (12)$$

This is to say, for l_2 loss function, the estimation error for all nodes tend to be equal no matter how the real latency is short or long. While l_1 loss function is adopted, we can get a similar result.

When non-negative ensuring scheme, it always be NMF algorithm, is introduced, this situation tend to be even worse. Let's assume the n -th entry of \mathbf{u}_i , $u_{i,n} < 0$, then we can get:

$$\begin{aligned} e_{ij} &= \text{abs}(d_{ij} - \mathbf{u}_i \cdot \mathbf{v}_j^T) \\ &= \text{abs}(d_{ij} - (u_{i,1} \times v_{j,1} + \cdots + u_{i,n} \times v_{j,n} \cdots + u_{i,N} \times v_{j,N})) \\ &\leq \text{abs}(d_{ij} - (u_{i,1} \times v_{j,1} + \cdots + 0 \times v_{j,n} \cdots + u_{i,N} \times v_{j,N})) \end{aligned} \quad (13)$$

This is to say, non-negative ensuring scheme will enlarge the estimation error.

4 Experiments

At first we introduce the relative error, RE of the latency estimation as an evaluation criterion. The relative error value $RE_{i,j}$ between node i and j is defined as:

$$RE_{i,j} = \text{abs} \left(\frac{\hat{d}_{ij} - d_{ij}}{d_{ij}} \right) \times 100 \% \quad (14)$$

And more, we define those latencies which are shorter than 50 ms as short latencies, while the others as long latencies. The experiments are performed on Planetlab and King datasets, which includes 226 and 1740 nodes respectively. All the experiments of this article set the dimension of vector \mathbf{u}_i and \mathbf{v}_i to be 5.

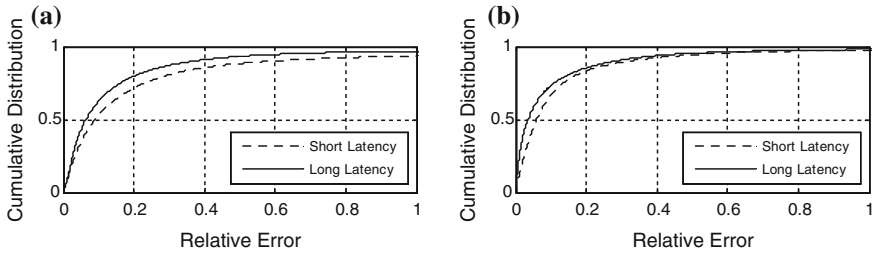


Fig. 3 The cumulative distributions of the relative error of short and long latency for different datasets. **a** King dataset **b** Planetlab dataset

Figure 3 shows the cumulative distributions of short and long latencies respectively. From this figure we can see that the estimation of short latency is clearly lower than that of long latency is both datasets. For King dataset, there are about 84.3 % estimations of long latency have no more than 20 % relative error value, while this criterion of short latency is only 69.9 %. The same tendency are appeared in the Planetlab: there about 86.1 % estimations' relative error of long latency are less than 20 %, while for short latency, only 77.2 % entries have no more than 20 % relative estimation error value.

Now we study the influence caused by NMF algorithm. To simplify the discussion, we only extract those entries which have negative estimation value while NMF algorithm is not adopted since those entries are always very short. Figure 4 and Table 1 show the estimation error of King and Planetlab datasets respectively. For Planetlab, there are only 3 entries which have negative estimation value. It is clearly that NMF algorithm can enlarge the estimation error of those entries significantly.

Fig. 4 The comparison of negative entries of King dataset

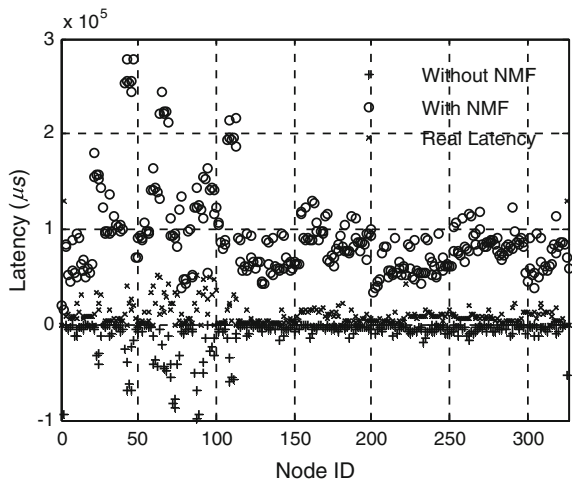


Table 1 The comparison of negative entries of Planetlab dataset

Entries	[20,21]	[21,20]	[21,52]
Real latency (μ s)	1999	2042	51,192
DMFSGD With NMF	-4443	-4604	-2135
DMFSGD Without NMF	9619	11,013	4154

5 Conclusion

This paper reports an easy overlooked shortcoming of latency matrix completion for the first time. Theoretical analysis shows that the inappropriate loss function is a possible reason for this shortcoming. And more, non-negative ensuring scheme make it worse. Our experiments show that this shortcoming influence the applications which performance relies a lot on latency matrix completion algorithms, and should not be neglected. How to overcome this shortcoming and improve the estimation accuracy of short latency is an open issue and worth deeply study in the future.

References

1. Nazanin M, Reza R, Ivica R et al (2014) ISP-friendly live P2P streaming. *IEEE/ACM Trans Netw* 22(1):244–256
2. Armitage G, Heyde A (2012) REED: optimizing first person shooter game server discovery using network coordinates. *ACM Trans Multimedia Comput Commun Appl (TOMCCAP)* 8(2):20
3. Jiang JR, Wu JW, Fan JY et al (2014) Immersive voice communication for massively multiplayer online games. *Peer-to-Peer Netw Appl*, (Already online but not assigned to an issue yet)
4. Klein A, Ishikawa F, Honiden S (2013) Towards network-aware service composition in the cloud. In: *Proceedings of the 21st international conference on world wide web*, Lyon, France, pp 959–968
5. Sherr M, Mao A, Marczak WR et al (2010) A3: an extensible platform for application-aware anonymity. In: *Proceedings of the network and distributed security symposium*, San Diego, USA
6. Wacek C, Tan H, Bauer K et al (2013) An empirical evaluation of relay selection in Tor. In: *Proceedings of the network and distributed security symposium*, San Diego, USA
7. All Pair Ping Project. http://pdos.csail.mit.edu/~strib/pl_app/
8. Frank D, Russ C, Frans K, Robert M (2004) Vivaldi: a decentralized network coordinate system. In: *Proceedings of 2004 SIGCOMM*, Portland, OR, USA, pp 15–26
9. Lee S, Zhang ZL, Sahu S et al (2010) On suitability of euclidean embedding for host-based network coordinate systems. *IEEE/ACM Trans Netw* 18(1):27–40
10. Candes EJ, Plan Y (2011) Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans Inf Theory* 57(4):2342–2359
11. Mao Y, Saul LK, Smith JM (2006) IDES: an internet distance estimation service for large networks. *IEEE J Sel Areas Commun* 24(12):2273–2284
12. Wang Z, Chen M, Xing C et al (2013) Multi-Manifold model of the internet delay space. *J Netw Comput Appl* 36(1):211–218
13. Liao Y, Geurts P, Leduc G (2010) Network distance prediction based on decentralized matrix factorization. In: *Networking*. Springer, Heidelberg, pp 15–26

14. Liao Y, Du W, Geurts P et al (2013) DMFSGD: a decentralized matrix factorization algorithm for network distance prediction. *IEEE Trans Netw* 21(5):1511–1524
15. Chen Y, Wang X, Shi C et al (2011) Phoenix: a weight-based network coordinate system using matrix factorization. *IEEE Trans Netw Serv Manage* 8(4):334–347
16. Planetlab Dataset. <http://www.eecs.harvard.edu/~syrah/nc/sim/pings.4hr.stamp.gz>
17. King Dataset. <http://pdos.csail.mit.edu/p2psim/kingdata/>

Facial Feature Extraction Based on Weighted ALW and Pulse-Coupled Neural Network

Junhua Liang, Zhisheng Zhao, Xiao Zhang, Yang Liu
and Xuan Wang

Abstract In order to improve the robustness of face identification with the changes of illumination, expression and facial alteration, a new facial feature extraction algorithm based on weighted adaptive lifting wavelet (ALW) scheme and pulse-coupled neural network (PCNN) is involved in this paper. The face images are decomposed into several subbands by weighted adaptive lifting scheme. Then the PCNN is utilized to decompose each weighted subbands into a series of binary images, the entropies of which are calculated and regarded as facial features. Experimental results show that the method yields a good robustness against the illumination, expression and facial variability and reduces the computer burden.

Keywords PCNN · Adaptive lifting wavelet · Face recognition

1 Introduction

Face recognition possesses important theoretical research value as a typical research topic in image processing, pattern recognition, and artificial intelligence. Simultaneously, it has broad application prospects in financial security, criminal detection, public security and other fields [1].

The Funding Project of Science & Technology Research and Development in Hebei North University (Grant No. ZD201301).

Major Scientific Research Projects in Higher School in Hebei Province (Grant No. ZD20131085).

J. Liang · Z. Zhao (✉) · X. Zhang · Y. Liu
School of Information Science and Engineering, Hebei North University,
Zhangjiakou 075000, Hebei, China
e-mail: zhaoshisheng_cn@sina.com

X. Wang
School of Physics and Information Technology, Shaanxi Normal University,
Xi'an 710062, Shaanxi, China

Feature extraction algorithm is the main step of face recognition which attracts great attention of scholars [2]. With the change of illumination, expression, posture and facial variation, robustness and rapidity of feature extraction are the key problems in face recognition application [2]. Existing methods for facial feature extraction are classified into four methods including: geometric-based, model-based, subspace-based and spacial-frequency domain-based [3]. Geometric-based method mainly extracts the geometric structure of major facial organs, however, if certain expression or posture changes, the identification accuracy is poor. In order to overcome the above problems, the method based on transform domain or model is proposed. Face image will be decomposed into several subbands, the high frequency of which is considered as feature in Gabor, wavelet and Contourlet transform [4–6]. The Markov or Hidden Markov model is utilized to simulate the statistical characteristics of input signal while the face image is considered as a random variable [7]. These methods partly improve the robustness of expression, posture and illumination changes. But it can't meet the real-time requirement because of its high computation complexity. Face image will be projected into subspace by subspace methods [8], such as principal component analysis (PCA), linear decision analysis [9, 10] (LDA) and independent component analysis (ICA). The projection coefficient is considered as feature. The method has higher computation efficiency, but the recognition rates and robustness will decline with the change of expression, posture and position, because it can hardly distinguish the differences caused by the external environment and face image itself.

Recently, several frameworks of lifting wavelet transform were applied into face feature extraction [11], which inherited the advantages of multiresolution representation over the traditional wavelet transform such as flexible design, low computational complexity and real-time application. According to the discontinuities playing an important role in face image, we put forward a modified adaptive lifting scheme to protect the edge information, and adapt the weighted subbands considering the different contribution each subband takes in the whole image. In this paper, weighted adaptive lifting scheme is utilized to decompose the face image into several subbands, then the weight of each subband is computed, and will be sent to Pulse-coupled Neural Network as input, through which we can obtain a series of binary images, the entropies of these binary images are considered as facial feature. The experimental results in ORL and YALE face database show that the algorithm is efficient contrast to subspace method, while it obtains strong robustness against the change of light, posture and facial expression.

2 Lifting Wavelet Scheme and Pulsed-Coupled Neural Network

2.1 Lifting Scheme

The lifting wavelet scheme called the second generation wavelet transform provides us with a simple explanation of the basic theory of wavelet method [12]. In lifting wavelet scheme, transformation process is divided into three steps: split, predict and update. The decomposition and reconstruction are just shown in Fig. 1.

Spilt: The original signal S is divided into low frequency x and high frequency y by a particular transform such as the simple polynomial decomposition.

Update: An update map U will act on the high frequency y to modify x yielding a new low frequency x' , i.e.

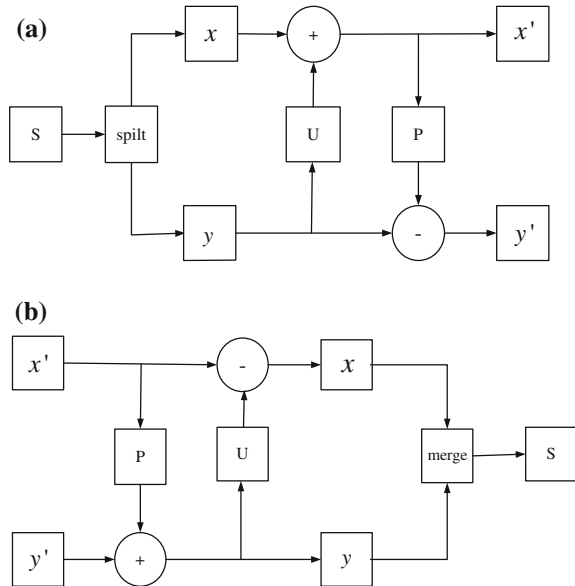
$$x' = x + U(y) \tag{1}$$

Predict: A prediction map P operating on x' is used to modify y resulting in a new high frequency y' , i.e.

$$y' = y - P(x') \tag{2}$$

Actually, the reconstruction process is the reversed lifting scheme steps as shown in Fig. 1b. From the above operations we can see that the lifting wavelet achieved situ operation, in other words, the old data can be replaced with new data

Fig. 1 Lifting scheme (a) and reconstruction (b)



flow stream at each data point, thereby a lot of storage space is saved by lifting scheme. In these way it can avoid complex computation resulted from the repeated convolution, meanwhile real-time performance is necessary for FPGA realization.

2.2 Pulse-Coupled Neural Network

Pulse-coupled Neural Network is proposed by Johnson, according to neurons sync pulse phenomenon in cat visual cortex. PCNN is invariant to rotation, scaling and distortion changes [13, 14], which is composed of receiving unit, connecting unit and pulse generator. In image processing, each pixel and neuron are corresponding to each other. For better application, this paper adopts the following representation of neural network model [15]:

$$F_{ij}(n) = S_{ij} \quad (3)$$

$$L_{ij}(n) = \sum_{kl} W_{ijkl} Y_{kl}(n-1) \quad (4)$$

$$U_{ij}(n) = F_{ij}(n)(1 + \beta L_{ij}(n)) \quad (5)$$

$$X_{ij}(n) = \frac{1}{1 + \exp(E_{ij}(n-1) - U_{ij}(n))} \quad (6)$$

$$Y_{ij}(n) = \begin{cases} 1, & X_{ij}(n) > 0.5 \\ 0, & \text{else} \end{cases} \quad (7)$$

$$E_{ij}(n) = \begin{cases} V_E, & Y_{ij}(n) = 1 \\ \alpha_E E_{ij}(n-1), & Y_{ij}(n) = 0 \end{cases} \quad (8)$$

In which, n is the current iteration, (i, j) represents the neuron location, S is input, M and W are called weight matrix, V_L , V_F and V_T indicate the inherent voltage respectively, β is the connecting coefficient, α_F , α_L and α_T represent attenuation time constant. The model is a simplification over standard Pulse Coupled Neural Network, which should only set V_E , α_E and β avoiding a lot of computation burden caused by preset parameters. These undetermined parameters will be different in application fields, so there is no general theoretical method except conducting experiment.

3 The Proposed Algorithm

3.1 Weighted Lifting Wavelet Scheme

Classical lifting scheme, where the update map U and the prediction map P are fixed, obviously adopts the uniform smooth operation for the whole image, which hardly precisely locates the region of interest or discontinuities. In this section we build a weighted lifting wavelet scheme taking full account of the alternating details.

Spilt: The original image $f(i, j)$ with size $N \times M$ is split into four images $x(n, m)$ $y_h(n, m)$ $y_v(n, m)$ $y_d(n, m)$ with size $N/2 \times M/2$ by the simple polyphase decomposition, shown as Fig. 2.

Prediction: the output approximation signal $x'(n, m)$ equals

$$x'(n, m) = x(n, m) \oplus_{p_{nm}} U_{p_{nm}}(y_h(n, m), y_v(n, m), y_d(n, m)) \tag{9}$$

where p_{nm} is the output of the binary decision map D , that is

$$p_{nm} = D(n \cdot m) = \begin{cases} 0 & \sigma(n, m) > T \\ 1 & \sigma(n, m) \leq T \end{cases} \tag{10}$$

where T is a given threshold, and $\sigma(n, m)$ is defined as

$$\begin{aligned} \sigma(n, m) = & |y_h(n-1, m) - \mu| + |y_h(n, m) - \mu| + |y_v(n, m-1) - \mu| + |y_v(n, m) - \mu| \\ & + |y_d(n-1, m) - \mu| + |y_d(n, m) - \mu| + |y_d(n, m-1) - \mu| + |y_d(n-1, m-1) - \mu| \end{aligned} \tag{11}$$

$f(2n-1, 2m-1)$	$f(2n-1, 2m)$	$f(2n-1, 2m+1)$	$y_d(n-1, m-1)$	$y_v(n-1, m)$	$y_d(n-1, m)$
$f(2n, 2m-1)$	$f(2n, 2m)$	$f(2n, 2m+1)$	$y_h(n, m-1)$	$x(n, m)$	$y_h(n, m)$
$f(2n+1, 2m-1)$	$f(2n+1, 2m)$	$f(2n+1, 2m+1)$	$y_d(n, m-1)$	$y_v(n, m)$	$y_d(n, m)$

Fig. 2 Left coordinates for two-dimensional signal. Right location of the input $x(n, m)$ $y_h(n, m)$ $y_v(n, m)$ $y_d(n, m)$ after polyphase decomposition

where μ denotes the mean of the eight horizontal, vertical and diagonal neighbors of $x(n, m)$, i.e.

$$\mu = \frac{1}{8} [y_h(n-1, m) + y_h(n, m) + y_v(n, m-1) + y_v(n, m) + y_d(n-1, m) + y_d(n, m) + y_d(n, m-1) + y_d(n-1, m-1)] \quad (12)$$

In our adaptive lifting scheme, p_{nm} governs the choice of the update step. For every possible decision p_{nm} , we have a different addition operator $\oplus_{p_{nm}}$ and update operator $U_{p_{nm}}$. In other words, they are location-dependent and adaptive. Update operator $U_{p_{nm}}$ is given by

$$U_{p_{nm}} = \begin{cases} 0 & p_{nm} = 0 \\ \frac{1}{4} [y_h(n-1, m) + y_h(n, m) + y_v(n, m-1) + y_v(n, m)] & p_{nm} = 1 \end{cases} \quad (13)$$

and the addition operator $\oplus_{p_{nm}}$ is of the form

$$x \oplus_{p_{nm}} u = \begin{cases} x + u & p_{nm} = 0 \\ \frac{1}{2}(x + u) & p_{nm} = 1 \end{cases} \quad (14)$$

As shown in Fig. 3, in our adaptive lifting scheme, the update map U is adaptive, while the prediction step is fixed. The prediction operators $P_h(x'(n, m))$, $P_v(x'(n, m))$ and $P_d(x'(n, m))$ are defined as follows:

$$P_h(x'(n, m)) = \frac{1}{2} [x'(n-1, m) + x'(n, m)] \quad (15)$$

$$P_v(x'(n, m)) = \frac{1}{2} [x'(n, m-1) + x'(n, m)] \quad (16)$$

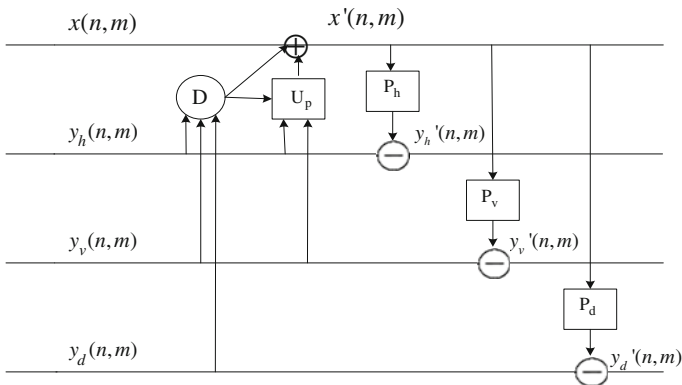


Fig. 3 Our 2-D adaptive lifting scheme

$$P_d(x'(n, m)) = \frac{1}{2}[x'(n-1, m-1) + x'(n, m)] \quad (17)$$

The detail signals $y'_h(n, m)$, $y'_v(n, m)$ and $y'_d(n, m)$ are calculated by

$$y'_h(n, m) = y_h(n, m) - P_h(x'(n, m)) \quad (18)$$

$$y'_v(n, m) = y_v(n, m) - P_v(x'(n, m)) \quad (19)$$

$$y'_d(n, m) = y_d(n, m) - P_d(x'(n, m)) \quad (20)$$

The contribution of each detail signal to the entire image is different, which can be computed as follows:

$$\alpha_i = \frac{\sum x_i(n, m)}{\sum_{m=1}^M \sum_{n=1}^N f(n, m)} \quad (21)$$

The basic idea underlying our adaptive scheme is stated as follows: for smooth regions where $p_{nm} = 1$, we compute $x'(n, m)$ as the weighted average of $x(n, m)$ and its eight horizontal, vertical and diagonal neighbors, whereas for less homogeneous regions where $p_{nm} = 0$, we do not perform any filtering, i.e. $x'(n, m) = x(n, m)$. The main reason to do this is that the discontinuities in face images play an significant role in face identification, but there are many isolated discontinuous points resulted from noise or other distortions, which may disturb face identification, and should be filtered. By adaptive update and prediction, we can extract the face information with filtering the distortions. In addition, our adaptive scheme only involves the addition and subtraction operations in spatial domain, and holds high parallel property, in situ operation, so it presents very low computational cost and easy hardware implementation. Figure 4 shows the decomposition results (at level 1) via our weighted adaptive scheme and the wavelet transform used in [11]. It is noted from the detailed images in Fig. 4 that the weighted adaptive scheme preserves the edges well with less distortions in comparison with the wavelet transform.

3.2 The Proposed Method

Face image is always influenced by some factors such as facial expression, illumination and facial modification, which will reduce the recognition and limit its application. Meanwhile, real-time is a key requirement in practice, so it is necessary to look for a strongly robust and real-time algorithm.

The characteristics obtained by weighted adaptive lifting wavelet is not sensitive to expression change, and fully take into account of contour edge plays a key role in face recognition. Adaptive lifting wavelet outperforms because of flexible design and low computational complexity. Pulse coupled neural network itself has the



Fig. 4 The decomposition results (at level 1) using our weighted adaptive scheme and the wavelet transform used in [11]. *Top* (from left to right): the approximate, horizontal, vertical and diagonal detail images of the wavelet transform. *Bottom* (from left to right): the approximate, horizontal, vertical and diagonal detail images of our weighted adaptive scheme

invariant characteristics against the rotation, scale and translation meanwhile it can better explain the global features of the image. It will get stronger robustness combining the local information by weighted adaptive lifting wavelet and global information by pulse coupled neural network.

The low frequency component in face image represents most information, describing the invariant features of face. It is proved that approximation component is best for recognition at the first layer decomposition in literature [16]. Therefore, this article choose the first layer of low frequency component as the feature, greatly reducing the complexity of calculation. The high frequency component in face image represent details corresponding to illumination, facial expression and location change. Statistical results show that the diagonal component contains too many local informations caused by illumination, rotation, and facial expression change, while the excessive internal pattern changes will affect the recognition performance. Considering the recognition performance and computational complexity, we select the horizontal component and vertical component at one layer as the feature.

The implementation steps in feature extraction process are as follows:

- Step 1: Get approximate, horizontal and vertical subbands by weighted adaptive lifting wavelet at the first layer for a given face image.
- Step 2: The decomposed subbands will be considered as the input of pulse coupled neural network, simulating the perception process of biological visual cortex, and converted into a series of cognition sequence of binary image.

Table 1 The parameters of PCNN

V_E	α_E	β
127.5	0.8	0.03

Step 3: The entropy of each binary sequence is calculated and considered as the final classification recognition characteristics.

After several test experiments, the three input parameters with PCNN model in our feature extraction process are shown in Table 1. From the table, the inherent potential of dynamic threshold V_E is set as half of the biggest grey values of each image. The decay time constant α_E and the strength connection coefficient β is selected the optimal value on the basis of many experiments, which get better effect in ORL and YALE face database.

Connected weighted coefficient matrix M and W set like this:

$$M = W = \begin{bmatrix} 0.5 & 1 & 0.5 \\ 1 & 0 & 1 \\ 0.5 & 1 & 0.5 \end{bmatrix}$$

The subbands decomposed by weighted adaptive lifting wavelet are converted to a series of binary cognition sequences through PCNN model which maybe different because of different grey value distribution. So we calculate the entropy of the binary sequence as final feature as that. Shannon entropy has the scaling, translation and rotation invariant features. The entropy is defined as:

$$H(p) = -p_0 \log(p_0) - p_1 \log(p_1) \tag{22}$$

where p_0 and p_1 respectively represent the probability of pixels 0 and 1 in binary sequence. The introduction of entropy and the combination of local special processing using weighted adaptive lifting wavelet with rotation, scaling, distortion invariant nature of pulse coupled neural network [14], up to the design requirements of face recognition robustness.

4 Simulation and Analysis

4.1 Experiment in ORL

Our experimental simulation environment is Intel(R) Celeron(R) CPU 2.1 GHz, 2 GB memory, Microsoft Windows XP professional system, and Matlab7.0 software. ORL face database contains 256 grey values corresponding to 400 different facial images, the resolution of each image is 92×112 . The 10 face images of each person include different facial expression change, tiny attitude change, and less than 20 % scale change. Figure 5 shows the facial expressions of two person random



Fig. 5 Two person's facial expression images random selected from ORL face database

select in ORL face database. There is no any preprocessing for the test face image in our experimental process.

We respectively random select 3, 4, and 5 samples in ORL face database as training set, the remaining as test set. Our algorithm is compared with the principal component analysis (PCA) [8], linear decision analysis (LDA) [9], independent component analysis (ICA) [10], M-PCNN [15] and the algorithm in literature [17]. In order to eliminate the influence of different classifier for recognition accuracy, each method is classified by support vector machine with the optimum parameters. All kinds of algorithms take the average recognition accuracy of 20 times. The contrast results of robustness are shown in Table 2 in terms of variance which represent the deviation of each training.

From Table 2 the robustness of our algorithm is significantly higher than the classic subspace method and only PCNN method. In addition, the method of literature [17] has been carried on the comparison which combines the visual perception with edge protection. It performs slight difference between the two methods.

Table 2 The recognition performance in terms of identify precision (%) and variance in ORL face database

Method	Recognition rate			Variance
	3 Train	4 Train	5 Train	
PCA	78.52	82.08	87.5	6.3951
LDA	83.57	85.41	92.00	6.2684
ICA	83.21	83.75	91.00	6.1519
M-PCNN	85.35	87.91	94.50	6.6759
Literature [17]	91.02	92.50	99.30	6.0415
Our method	91.07	95.83	99.49	5.9708

4.2 Experiment in YALE

It includes 165 frontage face images from 15 samples with the resolution of 100×100 and 256 grayscale values in YALE face database. These sample images are obtained under more obvious light condition, gesture and facial expression [18], which perform as high as 200° of depth rotation and plane rotation even 10 % of face scale change. Figure 6 shows two person’s face image randomly selected in YALE face database. It can further verify the robustness of these algorithms because of more rich facial expression, uneven illumination change, and sunglasses wearing.

This experiment randomly select 3, 4, 5 sample images of each person as training, the rest for test. Classifier uses support vector machine (SVM) with the optimization parameters. The identification accuracy and variance of all kinds of algorithms are shown in Table 3, as you can see from the table, all recognition rates of these algorithms are decreased, mainly due to the obvious illumination, posture, and accessories change in YALE face database. But the robustness of our method is significantly higher than that of the classical PCA and LDA algorithm, and more stability than the single M-PCNN method. The decline degree of robustness and variance is close to the literature [17] method.



Fig. 6 Two person’s facial expression images randomly selected in YALE face database

Table 3 The recognition performance in terms of identify precision (%) and variance in YALE face database

Method	Recognition rate			Variance
	3 Train	4 Train	5 Train	
PCA	63.08	70.38	73.11	9.2573
LDA	71.08	81.43	83.89	9.6137
ICA	61.50	68.70	71.40	9.1558
M-PCNN	79.30	84.80	92.70	9.5257
Literature [17]	82.16	88.69	93.28	7.9028
Our method	82.23	87.78	93.33	7.8489

Table 4 Feature extraction time (ms) of these compared methods

Methods	PCA	LDA	ICA	M-PCNN	Literature [17]	Our method
Extraction time	3.41	3.75	3.48	3.19	4.34	3.05

Due to the demand of real-time performance and hardware realization in practice, the compute cost of feature extraction time is compared just as shown in Table 4. From the perspective of real-time analysis, our features only contain three subbands with only one layer. Compared with the tens of thousands of dimension in original image, our method reduces nearly two orders of magnitude, and greatly reduce the time complexity, significantly lower than the literature [17] and traditional subspace algorithm. Meanwhile adaptive lifting wavelet can meet the requirements of hardware implementation because of situ operation which greatly save the storage space.

5 Conclusion

In order to improve the robustness of illumination, posture, facial expression change in face recognition, and to meet the requirement of real-time performance and calculation cost in practical application, a weighted adaptive lifting scheme combined with pulse coupled neural networks is proposed in this paper. This method combines the local detail processing operation in weighted adaptive lifting scheme with the scaling, translation and rotation invariant advantages in pulse coupled neural networks, which yields strong robustness and low calculation cost in the contrast experiment of ORL and YALE face database. Considering the requirements of robustness and calculate burden, our algorithm perform high effectiveness.

References

1. Kwaka N (2008) Feature extraction for classification problems and its application to face recognition. *Pattern Recogn* 41(5):1701–1717
2. Qiao L, Chen S, Tan X (2010) Sparsity preserving projections with applications to face recognition. *Pattern Recogn* 43(1):331–341
3. Jafri R, Arabnia HR (2009) A survey of face recognition techniques. *Inf Process Syst* 5(2): 41–67
4. Yang M, Zhang L (2013) Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary. *Pattern Recogn* 46(7):1865–1878
5. Xu B, Li HY (2012) Wavelet face recognition using Bayesian classifier. *Adv Mater Res* 467:561–564
6. He YC, Liu WB, Zhang G (2011) Rotation-invariant texture image retrieval algorithm based on nonsubsampling contourlet transform. *J Image Graph* 16(1):79–83

7. Wang ZHCH, Liu HY (2013) A face recognition based on hidden markov model. *Comput Appl Softw* 30(2):304–307
8. Shchegoleva NL, Kukharev GA (2010) Application of two-dimensional principal component analysis for recognition of face images. *Pattern Recogn Image Anal* 20(4):512–527
9. Yu H, Yang J (2000) A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recogn* 34(10):2067–2070
10. Bartlett MS, Movellan JR, Sejnowskit J (2002) Face recognition by independent component analysis. *IEEE Trans Neural Networks* 13(6):1450–1464
11. Wang X, Liang JH et al (2013) On-line fast palmprint identification based on adaptive lifting wavelet scheme. *Knowl-Based Syst* 42:68–73
12. Sweldens W (1998) The lifting scheme: a construction of second generation wavelets. *SIAM J Math Anal* 29(2):511–546
13. Johnson JL, Ritter D (1993) Observation of periodic waves in a pulse-coupled neural network. *Opt Lett* 18(15):1253–1255
14. Johnson JL (1994) Pulse-coupled neural nets: translation, rotation, scale, distortion and intensity signal invariance for images. *Appl Optics* 33(26):6239–6253
15. Wang X, Yang G (2013) Facial feature extraction based on NSCT and M-PCNN. *Comput Eng* 49(1):213–216
16. Chen CK, Gao XM et al (2005) Wavelet feature-based nonlinear feature extraction technique. *J Electron Inf Technol* 27(2):290–293
17. Gu XH (2013) Visual perception and edge preserving illumination invariant face recognition. *Acta Electronica Sin* 41(8):1500–1505
18. Olivetti & Oracle Research Laboratory. The Olivetti & Oracle Research Face Database of Faces[EB/OL]. 21 Nov 2010. <http://www.camorl.co.uk/facedatabase.html>

Event Representation and Reasoning Based on SROIQ and Event Elements Projection

Wei Liu, Ning Ding, Yue Tan, Yujia Zhang and Zongtian Liu

Abstract Events have become central elements in the representation of information from various semantic web applications. It is necessary to develop a formal language for describing and reasoning event knowledge. Description logic is a well-defined knowledge representation language, but it is difficult to represent the event and elements with different characteristics. This paper proposes an event element projection method which is combined with SROIQ to build a new formalization method for event-centered knowledge. Event element projection unifies representation framework of event and event status, establishes the semantic relations between event and its elements. Through element projection and SROIQ, event classes, event instances and event elements can be effectively described in a unified style. An example of formalization on water pollution emergencies ontology is provided based on SROIQ and element projection method. The semantics of event relations based on element projection and reasoning on event relations are also discussed at last.

Keywords Event model · SROIQ · Event element projection · Event-based reasoning

1 Introduction

Events have become a key concept for representing knowledge and organizing and structuring media on the web and different application domains, such as emergency response, public opinion monitoring, history and cultural heritage, etc. Event ontology is a new paradigm for representation and reasoning on events. It highlights the representation of event classes and relations. However, the complexity of event

W. Liu (✉) · N. Ding · Y. Tan · Y. Zhang · Z. Liu
School of Computer Engineering and Science, Shanghai University, Shanghai, China
e-mail: liuw@shu.edu.cn

N. Ding
e-mail: ces13721024@shu.edu.cn

structure results in the shortage of representation and reasoning language for event-based knowledge. Description logic is a well-defined knowledge representation language which has become the logic foundation of standard ontology language. In recently year, different extension work on description logics emerges massively, and some of them can be used to represent action, time and place concepts respectively. In another word, each element of event can be described isolatedly. There is a lack of a unified formalization framework for describing a whole events, elements and relations. Aims at a unified formal framework for event modeling, we propose an Event Element Projection (EEP) method, which is combined with SROIQ [1] (a standard description logic with rules) to describe and reason event elements. In EEP, SROIQ logic operators are used to construct a complex conceptual axiom that contains different event elements predicates, which can be projected onto an objective element for the simplification of inference. EEP provides inference for event elements and enhances the inference based on event semantic relations.

The remainder of this paper is structured as follows. Section 2 reviews the related work. Section 3 introduces several concepts about event model. Section 4 introduces the syntax and semantics of extended SROIQ and *Event Element Projection* method. Also, several inferences based on EEP are discussed in Sect. 4. An example of formalization on water pollution emergencies ontology is provided in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Related Work

In recent years, research on the use of events as a key concept for representing knowledge is surging, especially in semantic web community [2, 3]. A good deal of relevant research on the modelling of events has been done in the semantic web community. Among these works, SEM is proposed [4, 5] to model events in various domains, without making assumptions about the domain-specific vocabularies used. SEM is designed with a minimum of semantic commitment to guarantee maximal interoperability. Reference [6] proposes a music ontology based on event calculus and OWL-Time. Reference [7] presents a formal model of events, called Event Model-F, which is based on the foundational ontology DOLCE+DnS Ultralite (DUL) and provides comprehensive support to represent time and space, objects and persons, as well as causal, and correlative relationships between events. LODÉ [8] defined event as an action or event happening during a period, but is not capable of describing event relations. In [9], we propose a 6-tuple event model and event ontology, and implement event-related applications of text information [10]. Whereas, most event model are concept-centered, there exist deficiencies while modeling event knowledge, such as separateness of concepts (the person, objects,

places and action of event are not organized as a whole knowledge unit), insufficient capacities of capturing dynamic aspects of event.

Formalization of events, event classes and elements plays a key role of research on event ontology. Existing formal methods for the event knowledge focus on action and the process of event, such as situation and event calculus [11, 12]. Description logic is a series of knowledge-based formal logics which has become the logic basis of ontology language. Due to its clear model and theoretical mechanism [1, 13], description logic can effectively represent domain knowledge in applications via a static concept taxonomy to formalize specific model and reason on it [14]. Some existing work on extending description logic and specific ontologies are designed to represent and reason different event elements, including people, actions, time and so on. FOAF ontology [15] specifies a vocabulary that can be used to define, exchange and search for social information, describing people, their attributes and their relationships. Time ontology [16] is developed for describing the temporal content of web pages and the temporal properties of web services. The formalization of geographical ontology patterns are discussed in [17, 18]. Reference [19] presents a dynamic description logic for representation of actions, with an approach that embrace actions into the description logic. Most of these work ignore semantic relationship between event elements and events, resulting in the elements of the event are isolated, unconnected and static. It is necessary to develop a unified formal language for describing and reasoning the event elements (inner structure) and event relations (outer structure).

3 Concepts of Event Model

Events provide a natural way to express complicated relations between people, places, actions and objects. Event relationships provide more sophisticated description and reasoning of event-centered concepts. In this section, we will introduce an event model structure for representation of generic event information on Web.

Definition 1. (*Event*) We define an event as a thing happens in a certain time period and place, which some actors participate in and show some action features, along with the changing of internal status. Event e can be defined as a 6-tuple formally:

$$Event ::= (A, O, T, P, S, L)$$

- A: an action or a set of actions usually regarded as a trigger word to identify an event.
- O: objects involved in the event, including participants and entities.

- T*: the period of time that event lasting, including absolute time and relative time.
P: the location of an event happens.
S: status of object during an event happens, including *pre-condition* set and *post-condition* set.
L: language expressions of text-based event, it includes a *Core Words Expressions* (CWE) set and a *Core Words Collocations* set. *Core Words Collocations* (CWC) describe the fixed collocations between core words and other word.

Definition 2 Event class is an abstract event that represents a set of events with some common characteristics, denoted as *EC*:

$$EC = (E, C_A, C_O, C_T, C_P, C_S, C_L)$$

$$C_i = \{c_{i1}, c_{i2}, \dots, c_{im}, \dots\} \quad (i \in \{A, O, T, P, S, L\}, m \geq 0)$$

where *E* means an event set. C_i is the set of event elements. It denotes the common characteristics set of certain event element (element *i*). C_{im} denotes one of the common characteristics of event factor *i*. C_{im} is also called event elements class.

The relationships between the events are divided into two categories: taxonomic relation and non-taxonomic relations. The taxonomic relation describes the hierarchical structure of event classes. Non-taxonomic relations describe the internal semantic relationships between events or event classes, including composition relation, follow relation, causality relation and concurrency relation.

Subsumption relation (*is_a*): An event class can subsume or be subsumed by other event classes. It can be formalized as $EC1 \sqsubseteq EC2$.

Causality relation: If an event *e1* (instance of *EC1*) happened, then another event *e2* (instance of class *EC2*) happens at above a specified probability threshold, there is a causality relationship between *e1* and *e2* (or *EC1* and *EC2*). *EC1* is cause and *EC2* is effect, causality relation formalized as $EC1 \rightarrow EC2$.

Follow relation: Follow means events coming after in time order, as a consequence or result, or by the operation of logic. It can be formalized as $EC1 \triangleright EC2$.

Concurrency relation: If there are event *e1* (instance of class *EC1*) and event *e2* (instance of class *EC2*) occur simultaneously or successively in a certain period of time (the two event are coincident events), there is a concurrency relationship between *e1* and *e2* (or *EC1* and *EC2*), formalized as $EC1 || EC2$.

Composition relation: If an event instance *e1* of class *EC1* can be decomposed to several sub-events e_i ($i > 0$, instance of class *ECi*) with smaller granularity, and while all the smaller events e_i happened means *e* happened, there exists composition relation between *e1* and e_i (or *EC1* and *ECi*), which can be formalized as $EC1 \angle EC2$.

4 Event Projection Method Based on SROIQ

Compared with some previous sub-language of description logic [13], SROIQ extends much more features, which describe different event element conceptions, especially dynamic elements. However, event elements described with SROIQ is static, it is difficult to reason about the status of event. In order to reason on the status of event elements, it is necessary to build link among the elements of event. *Event Elements Projection* is proposed in this section to solve this problem.

4.1 SROIQ Syntax and Semantics for Event Model

Definition 3. The language of SROIQ is interpreted in models over \mathcal{I} , which is triples of the form $\mathcal{I} = (\Delta^{\mathcal{I}}, \bullet^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is nonempty set of concepts (the domain of I) and $\bullet^{\mathcal{I}}$ is an interpretation function. $\Delta^{\mathcal{I}}$ is the domain of \mathcal{I} , including such general concepts: as to events, it represents event classes; as to elements, it represents elements, like object, time, place elements. $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ describes the set of roles, that is, of property relation between concepts. In scope of events, it represents event relations, including taxonomic and non-taxonomic relations. In scope of elements, it represents actions in events, describing the property relation between two other elements, such as participants and entities. In this way, both the syntax solutions of event and element are the same but semantic different. Table 1 defines SROIQ syntax and semantics of event class or element.

Table 1 SROIQ syntax and semantics of event (element) class

Constructors	Syntax	Semantic
Concept (event class or element)	C	$C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
Nonempty set of concepts	$\top^{\mathcal{I}}$	$\Delta^{\mathcal{I}}$
Empty set of concepts	$\perp^{\mathcal{I}(r)}$	\emptyset
Negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
Disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
Conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
Concepts inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
Role (event relation or element property relation)	R	$R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
Exist restrict	$\exists R.C$	$\{x \mid \exists y \langle x, y \rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
Value restrict	$\forall R.C$	$\{x \mid \forall y \langle x, y \rangle \in R^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}}\}$
Inverse of role	R^-	$\{\langle x, y \rangle \mid \langle y, x \rangle \in R^{\mathcal{I}}\}$
Role inclusion	$R_x \sqsubseteq R_y$	$R_x^{\mathcal{I}} \subseteq R_y^{\mathcal{I}}$
General role inclusion	$R_1 \circ \dots \circ R_n \sqsubseteq R_m$	$\{\langle x_1, x_n \rangle \mid \langle x_1, x_n \rangle \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}, x_1, \dots, x_n \in \Delta^{\mathcal{I}}, \langle x_i, x_{i+1} \rangle \in V_i^{\mathcal{I}}\}$

Table 2 SROIQ syntax and semantics of event relations

Event relations	Syntax	Semantic
Cause	$EC1 \rightarrow EC2$	$\{\langle e_1, e_2 \rangle \arg \max(P(e_2 e_1))\}$
Concurrency	$EC1 EC2$	$\{\langle e_1, e_2 \rangle e_1^{[t_{11}, t_{12}]}, e_2^{[t_{21}, t_{22}]}, [t_{11}, t_{12}] \cap [t_{21}, t_{22}] \neq \emptyset\}$
Follow	$EC1 \triangleright EC2$	$\{\langle e_1, e_2 \rangle e_1^{[t_{11}, t_{12}]}, e_2^{[t_{21}, t_{22}]}, t_{12} \leq t_{21}\}$
Composition	$EC1 \angle EC2$	$\{\langle e_1, e_2 \rangle e_1^{[t_{11}, t_{12}]}, e_2^{[t_{21}, t_{22}]}, [t_{11}, t_{12}] \subseteq [t_{21}, t_{22}], C_{O1} \subseteq C_{O2}\}$

According to the definitions of event non-taxonomic relations, four extended symbols are introduced to represent four non-taxonomic relations, of which the syntax and semantics are defined in Table 2.

4.2 Event Element Projection

Definition 4. (Event Element Projection) The elements of events are abstracted as concepts (classes) or roles. Based on the constructors of description logic, a complex concept of element α in event e is built, which includes classes or roles of other elements. This complex concept is defined as the projection on α of event e , α represented as $e|_{\alpha}$.

Inside an event, the key element can characterize the type of event is the action element (or trigger words in some events); at the same time, participants, objects, time and place elements are linked with each other by means of action elements or trigger words in different events as the bridge. In description logic, an action element can be represented as the role associated with any other two elements, while the object, time, location and other elements are generally regarded as a class or concept. Figure 1 depicts various elements of the event are projected onto one of object elements, thus to construct a model for a complex concept of object element.

Event element projection plays an important role in the formalization and reasoning of events. In order to define concepts and roles in events based on

Fig. 1 Event element projection in event model

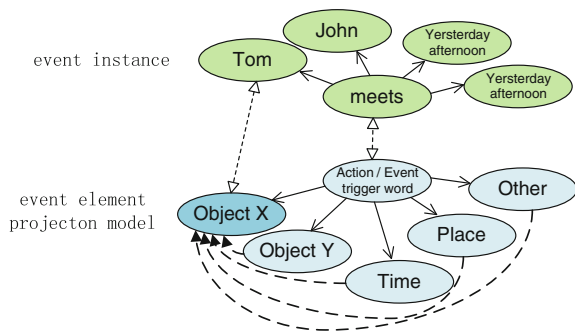


Table 3 Event element projection axioms examples

Event (class)	Element projected onto	Element projection axioms
E_1	Object O_s (active object)	$E_1 _O := O_s \sqcap \exists V_r.O_o$
E_2	Object O_s (active object)	$E_2 _O := O_s \sqcap \exists V \text{ when}.T \sqcap \exists Vat.P \sqcap \exists V_r.O_o$
E_3	Time T	$E_3 _T := T_1 \sqcap \exists V \text{ when}^-.O_s$
E_4	Place P	$E_4 _P := P_1 \sqcap \exists Vat^-.O_s$

description logic, action elements need to be created as several roles linked with different couples of event elements. Specifically, the action elements can be extended to three dimensions, building three roles as follow: in the dimensions of the object element, action element is represented as a role V_r , linking active object O_s with passive object O_o ; in the dimensions of the time element, action element is represented as the role V_{when} , linking with objects O_s with time element T ; in the dimensions of the place element, action element is represented as the role Vat , linking object O_s with place P . At the same time, according to description logic constructors of event elements in 2.1, complex concepts of specific elements under event element projection can be constructed.

As is depicted in Table 3, while an event contains more elements, it is difficult to construct the element concept that can be projected onto. The key issue is how to select the element that should be projected onto. This tends to depend on the features of the event type and element links the event relations. For instance, we can project the couple of events in concurrence relations onto time element. Time consistency can be described as the rule that both the time element projections has conceptual consistency. By means of event element projection, concepts of different elements projected are constructed, which increases the flexibility of knowledge representation and facilitates element reasoning in different events.

4.3 Element Projection of Event Status

Events projection method can not only formalize the event, but also be used to indicate the event status, including preconditions or post-condition. An event seems to be a dynamic process of static event status. It is necessary to integrate status into formalization and inference on events for the improvement of expression and reasoning ability in the formalization framework.

The structure of elements in status is similar to 6-element model in event. As a result, event element projection can also be applied to the formalization and inference about event status. For instance, status in an event (class) is formalized as a 2-tuple, (S_{pre}, S_{post}) , where S_{pr} is the precondition and S_{post} is the post-condition. Their element projection can be described as follows:

$$\begin{aligned}
S_{pre}|_O &:= Os_1 \sqcap \exists V_r. Oo_1, & S_{pre}|_T &:= T_1 \sqcap \exists V \text{ when}^- . Os_1 \\
S_{post}|_O &:= Os_2 \sqcap \exists V_r. Oo_2, & S_{post}|_T &:= T_2 \sqcap \exists V \text{ when}^- . Os_2
\end{aligned}$$

where, V_r is the role of a verb or trigger word in event status. In natural language, the action in event status often exists in attributive clauses of noun, such as “*carrying*” in “*a trunk carrying chemical*” and “*equipped*” in “*firefighters equipped with fire-fighting tools*”. If the event and status are projected onto the same type of element concepts, projected event and projected status can be built as a concept conjunction, like $S_{pre}|_O \sqcap E|_O$. In this way, event element projection unifies formal representation of event and status.

Following is an example for formalization of event status.

Example e₁: Police officers managed to rescue a kidnapped child from kidnappers

This event can be divided into pre-status, post-status and event itself.

pre-status: *The kidnapper abducted a child.*

event: *Police officers rescued a child.*

post-status: *The child has been saved.*

The event instance e_2 includes a police officer instance a , a kidnapper instance b and a child instance c . E_2 is the class of e_2 , $S_{2\ pre}$ is the pre-status and $S_{2\ post}$ is the post status. Roles *Rescue* and *Hijack* are actions in the event.

$$E_2|_{O'} := Child \sqcap \exists Rescue^- . PoliceOfficer$$

$$S_{2\ pre}|_{O'} := Child \sqcap Hijack^- . Kidnapper$$

$$S_{2\ post}|_{O'} := Child \sqcap Saved$$

$$S_{2\ pre}|_{O'} \sqcap E_2|_{O'} \sqsubseteq S_{2\ post}|_{O'}$$

$$E_2(e_2); PoliceOfficer(a), Kidnapper(b), Child(c), Rescue(a, c), Hijack(b, c), Saved(c)$$

In the above axioms, $E_2|_O$ is the E_2 ' projection onto *Child* element. Similarly, $S_{2\ pre}|_O$ and $S_{2\ post}|_O$ are status projections.

There always exists equivalence and inclusion between element concept of events and status. This makes it possible to use elements' concept conjunction to reason in the scope of elements in a unified way.

4.4 Event Relation Reasoning Based on Element Projection

In SROIQ, we formalize the non-taxonomic event relations as four roles, (R_{Cause} , $R_{CompositeOf}$, $R_{Concurr}$, R_{Follow}), and try to build links between elements of different events. These links reflect event relations in the scope of elements, which can help us to abstract more semantic information. To distinguish links in both scopes of event and elements, **explicit link** and **implicit link** are introduced in this section.

Table 4 Event relations semantic definition based on event element projection

Event relation	Semantic based on event element projection
$EC1 \sqsubseteq EC2$	$R_{is_a} = \{(EC1, EC2) EC1 _O \sqsubseteq EC2 _O\}$
$EC1 \rightarrow EC2$	$R_{cause} = \{(EC1, EC2) EC1 _O \sqsubseteq EC2 _O\} \cup \{(EC1, EC2) \exists EC3, then EC1 _O \sqsubseteq EC3 _O, EC3 _O \sqsubseteq EC2 _O\}$
$EC1 EC2$	$R_{concur} = \{(EC1, EC2) EC1 _T \sqsubseteq EC2 _T\}$
$EC1 \triangleright EC2$	$R_{follow} = \{(EC1, EC2) EC1 _T \sqcap EC2 _T \sqsubseteq \perp, \exists a, b, EC1(a), EC2(b), then After(a, b) or After(b, a)\}$ (After is a role that describe time sequence)

In this paper, elemental links of event relations is defined as **implicit link** now. Implicit links can be named as *event relation projection*, represented as follows:

$$Relation|_{element}, Relation \in \{cause, follow, concur, compositeOf\}, element \in \{A, O, T, P\}$$

As is shown in Table 4, Event relations can also be formalized and reasoned based on the element projection method.

5 Formalization Example

In this section, a formalization program for *vehicle chemical leakage water pollution emergency* is provided. As is depicted in Fig. 2, event classes are linked with each other through relations while elements and their links are extended on it as well. These implicit links associate elements of different events. For instance, the

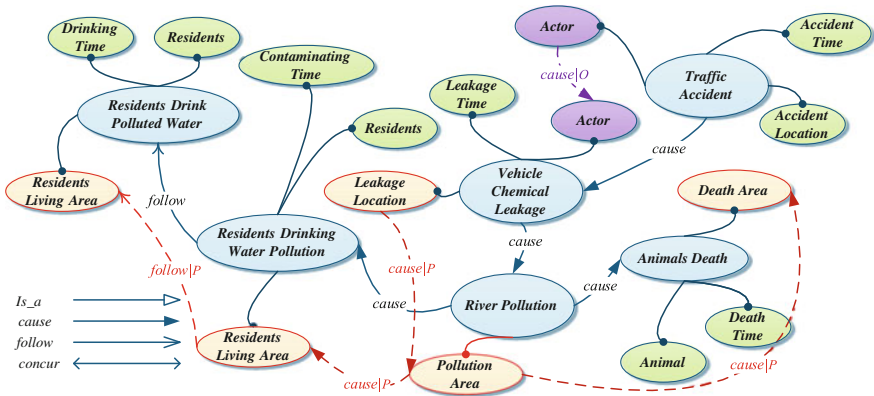


Fig. 2 Event classes model of *vehicle chemical leakage water pollution emergency*

implicit links between place element projection, like $cause|_P$ and $follow|_P$, contribute to represent the spatial associations and changes in scope of event elements.

In this solution, the formalization process of the upper part of *Vehicle chemical leakage water pollution emergency* can be offered as follows. Some event classes can be abstracted like E_0, E_1, E_2, E_3 , and their concepts and roles description are given based on SROIQ, where E_0 denotes *Traffic Accident*, E_1 denotes *Vehicle Chemical Leakage*, E_2 denotes *River Pollution*, E_3 denotes *Animals Death*. Concepts *Vehicle, Chemical, Pollutants, River, Animals* are elements of the above event classes. Particularly, *EncounterTrafficAccident* is regarded as a concept, for the action element of E_0 is intransitive. Roles *Carry, Leak, Leakat, PolluteAt, DieAt*, represent actions element. Among them, *Leak* associates active and passive objects while *Leakat* denotes object and place elements. *Carry* represents action in the pre-status of E_0 . According to formalization method based on element projection, logic program can be established as follow.

$$\begin{aligned}
E_0|_O &= Vehicle \sqcap EncounterTrafficAccident, & S_{0pre}|_O &= Vehicle \sqcap \exists Carry.Chemical \\
E_1|_O &= Vehicle \sqcap \exists Leak.Chemical \sqcap \exists Leakat.River, & E_0|_O \sqcap S_{0pre}|_O &\sqsubseteq E_1|_O \\
E_1|_P &= River \sqcap \exists (Leak^- \circ Leakat)^-.Chemical \\
E_2|_P &= River \sqcap \exists PolluteAt^-.Pollutants \\
Chemical &\sqsubseteq Pollutants \\
E_1|_P &\sqsubseteq E_2|_P, & E_3|_P &= Area \sqcap \exists DieAt^-.Animal, & E_2|_P &\sqsubseteq E_3|_P
\end{aligned}$$

Then definition axioms of event element projection should substituted the element projection in other axioms. In this way, the extended symbols of element projection can be eliminated. The above logic program is simplified and converted into the following program.

$$\begin{aligned}
&Vehicle \sqcap EncounterTrafficAccident \sqcap \exists Carry.Chemical \\
&\sqsubseteq Vehicle \sqcap \exists Leak.Chemical \sqcap \exists Leakat.River \\
&River \sqcap \exists (Leak^- \circ Leakat)^-.Chemical \sqsubseteq River \sqcap \exists PolluteAt^-.Pollutants \\
&Chemical \sqsubseteq Pollutants \\
&River \sqcap \exists PolluteAt^-.Pollutants \sqsubseteq Area \sqcap \exists DieAt^-.Animal
\end{aligned}$$

This logic program contains axioms of elements in events. The description of semantic links strengthens the expression ability of formalization framework. At the same time, event element projection symbols in logic program can be eliminated, so the ultimate logic program conforms to the standard syntax of SROIQ. It is significant that formalization method based on event element projection does not affect the original reasoning ability of standard SROIQ. Therefore, event element projection is an effective method of formalization of event relations.

6 Conclusion

The event element projection proposed in this paper attempts to unify event, elements, status of formalization framework. Based on extended SROIQ, a formalization method for event ontology is proposed. A formalization example of *vehicle chemical leakage water pollution emergency* is provided to demonstrate the process of event element projection. It is noted that the extended symbols of event element projection can be eliminated in the final logic program at last. Therefore, the extended language will not affect the original formalization inference ability. Meanwhile, event relation reasoning can be converted into element inference based on element projection method, which provides a new point for the inference on event relations. Because of the ambiguity and complexity of non-taxonomic relations, event reasoning on these relations need further study.

Acknowledgements This paper is supported by the Natural Science Foundation of China, No. 61305053 and No. 61273328.

References

1. Mosurovic M, Krdzavac N, Graves H et al (2013) A decidable extension of SROIQ with complex role chains and unions. *J Artif Intell Res* 47(1):809–851
2. Bethard S, Martin JH (2006) Identification of event mentions and their semantic class. In: *Proceedings of the empirical methods in natural language processing (EMNLP)*. ACM, New York, pp 146–154
3. Llorens H, Saquete E, Navarro-Colorado B (2010) TimeML events recognition and classification: learning CRF models with semantic roles. In: *International conference on computational linguistics*. ACM, New York, pp 725–733
4. Van Hage WR, Malaisé V, Segers R et al (2011) Design and use of the simple event model (SEM). *Web Semant Sci Serv Agents World Wide Web* 9(2):128–136
5. Van Hage WR, Malaisé V, de Vries GKD et al (2012) Abstracting and reasoning over ship trajectories and web data with the simple event model (SEM). *Multimed Tools Appl* 57(1):175–197
6. Raimond Y, Abdallah S, Sandler M et al (2007) The music ontology. In: *Proceedings of the 8th international conference on music information retrieval*, vol S. 1. ACM, New York, pp 417–422
7. Scherp A, Franz T, Saathoff C et al (2009) F—a model of events based on the foundational ontology dolce+DnS ultralight. In: *Proceedings of the fifth international conference on knowledge capture*. ACM, New York, pp 137–144
8. Shaw R, Troncy R, Hardman L (2009) LODÉ: linking open descriptions of events. *Lecture Notes in Computer Science*, pp 153–167
9. Liu Z et al (2009) Research on event-oriented ontology model. *Comput Sci* 36(11):189–192 (Chinese)
10. Zhong Z, Li C, Liu Z et al (2013) Web news oriented event multi-elements retrieval. *J Softw* 24(10):2366–2378 (Chinese)
11. McCarthy J (1963) Situations, actions, and causal laws. *Seman Inf Process*
12. Shanahan M (1999) The event calculus explained. In: *Artificial intelligence today*. Springer, Heidelberg, pp 409–430

13. Pittet P, Cruz C, Nicolle C (2013) A structural mathcal {SHOIN(D)} Ontology model for change modelling. *Lecture notes in computer science*, vol 8186, pp 442–446
14. Šimánčík F, Motik B, Horrocks I (2014) Consequence-based and fixed-parameter tractable reasoning in description logics. *Artif Intell* 209(2):29–77
15. Finin T, Ding L, Zhou L et al (2005) Social networking on the semantic web. *Learn Organ* 12 (5):418–435
16. Hobbs JR, Pan F (2004) An ontology of time for the semantic web. In: *ACM transactions on Asian language information processing*, vol 3. ACM, New York, pp 66–85
17. Carral D, Scheider S, Janowicz K et al (2013) An ontology design pattern for cartographic map scaling. In: *Semantic web semantics and big data*, vol 7882. Springer, Berlin, pp 76–93
18. Hu Y, Janowicz K, Carral D et al (2013) A geo-ontology design pattern for semantic trajectories. In: *Spatial information theory*, vol 8116. Springer, Berlin, pp 438–456
19. Chang L, Lin F, Shi Z (2007) A dynamic description logic for representation and reasoning about actions. *Knowledge science engineering and management*, vol 4798. Springer, Berlin, pp 115–127

The Research and Application of Data Warehouse's Model Design

The Data Warehouse's Model Design for the Decision Support System of Hospital Drugs

Zhangzhi Zhao, Jing Li, Yongfei Ye, Yang Liu and Yaxu Liu

Abstract Hospital drug business is complex, the data integration and the analysis are imperfect. Lacks of data warehouse system in which information is comprehensive and data is integrated, research the hospital drugs data, and carried on the design, the development and the deployment of the data warehouse project using “business dimension lifecycle method”. Created the data warehouse bus structure; established topic model; used dimension modelling carries on the logical modelling; studied in detail of Indexing strategy, granularity conversion algorithm and form design strategy in the design process, the overall logical organization pattern layout was clear. The construction method was practical. Gives a good hospital drug analysis model of data warehouse.

Keywords Data warehouse · Model design · Hospital drugs · Dimensional modeling

1 Introduction

With the rapid expansion of the hospital management and clinical data, Data Warehouse (DW), Data Mining (DM) and On-line analytical processing (OLAP) technology in Hospital information system (HIS) has been widely used in research

-
1. The Funding Project of Science and Technology Research and Development in Hebei North University (Grant No. ZD201301).
 2. Major Scientific Research Projects in Higher School in Hebei Province (Grant No. ZD20131085).

Z. Zhao (✉) · J. Li · Y. Ye · Y. Liu · Y. Liu
School of Information Science and Engineering, Hebei North Univeristy,
Zhangjiakou, China
e-mail: zhaozhisheng_cn@sina.com

and application. But so far, there are many problems to be solved in building the data warehouse has a good organizational structure, and can effectively real-time online analytical processing and data mining.

As an more mature applied branch of HIS, there are a lot of information about patients diagnosis and medical information, hospital and drug consumption information, hospital treatment drug monitoring data, etc. However, most data handling in hospitals database operations only limited data input, modify, query, statistics, delete and so on. How to use the existing drugs usage information in HIS to calculate simulation data for drug forecast analysis; so as to get medicinal product consumption model for automatically generate the rational use information of drugs; How to make use of analysis of regular change of drug information, guide and assist the pharmacy work; How to seek the optimal management of hospital drug supply chain, all these become the key points and difficulties in the hospital drugs management [1]. At present data is relatively complete in drug management system. But the lack of information comprehensive, integrated data warehouse system, the lack of data integration and analysis, let alone the automatic acquisition of decisions and knowledge. And build a good organization structure of data warehouse system is the only way which set up medical information platform of Decision Support System (DSS) and to improve the level of hospital drug management and information utilization.

2 The Implementation Methods

The realization of the data warehouse mainly based on the database technology, because data storage and management technology is relatively mature of database, so it is cost and complexity lower. But database system can not satisfy the need of data storage and data analysis in data warehouse. Database is designed to capture data, which oriented transaction, generally it is stored on-line trading data, and try to avoid data redundancy and design in the rules normalized. A data warehouse is designed to analyze the data, and aim at the theme design. Generally it is stored historical data. Its two basic element is dimension table and fact sheets. In the design of data warehouse is interested in bringing the redundancy, using the anti-paradigm approach to design [2].

The first step to building a data warehouse is clear its theme, the theme is a higher level data classification standards, each topic corresponds to a macro analysis of the field. Contraposing specific decision, table can be divided into multiple themes, specifically is to determine the scope of decisions and to solve the problem. But the determination of theme must be based on the existing online transaction processing (OLTP) system, otherwise the storage structure of Data Warehouse is designed according to the this topic will become an empty shell. In addition, when determining the theme, also need to find a “balance” between theme and the OLTP

data, according to the need of the subject to collect data, so that to build a data warehouse to meet the needs of the decision and analysis [3]. The designed and implemented environment of hospital drug data warehouse is based on SYABSE Adaptive Server Enterprise 12.0, Database Modeling tools used Power Designer11.0.

2.1 Construction Methods of Data Warehouse and Realization Process

In the process of project implementation, first carefully study the implementation of Data Warehouse theory, combining with service characteristics of hospital drug management, second applying “dimension of business life cycle method” to design, develop and deploy Data Warehouse project [4], the development process is shown in Fig. 1.

“Business dimensions life cycle method” illustrates a series of advanced tasks to design, develop and deploy data warehouse efficiently [3]. We mainly focus on its theme model, dimension modeling, physical modeling, data dump and development. The method most suitable for application in the scope of the project is easy to control and administer, in the scope of the project is difficult to control, demand is relatively difficult to define, it is necessary to across the direct demand of user data to establish the model.

Data Warehouse has used the unified data warehouse and affiliate data mart as a system structure: It adopt the way of data mart to further organize data according to the theme for business analysis of certain subject, affiliate of data marts’ data directly comes from the data warehouse. The main design is divided into the following 5 steps:

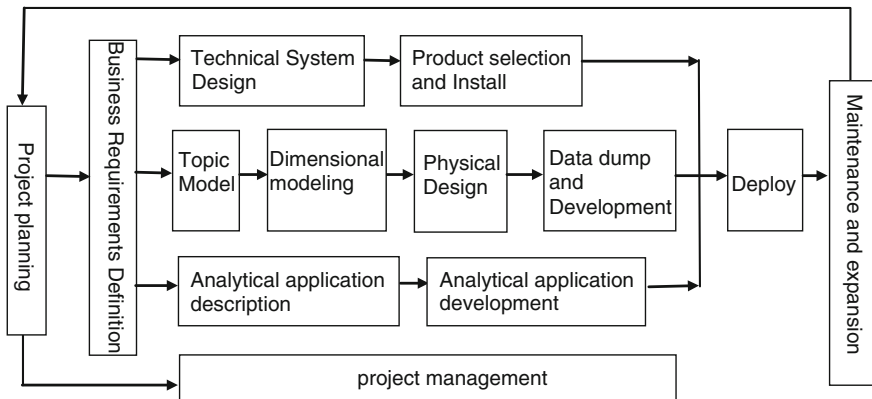


Fig. 1 Business dimension lifecycle chart

- (1) Define the details of each data source required in this theme, including the computer platforms, owner, data structure, using the data processing, drug warehouse updated plan etc.
- (2) Define the principle of data extraction so as to extract the required data from each data source, and define the data transform and load data table topics [5].
- (3) A theme is refined to multiple business themes to form the theme table. According to topic table select multiple data subset from the data warehouse, that is, data mart (DataMart). Data mart is usually targeted at a departmental level or a specific business need, it has a short development cycle and low cost, and can meet the needs of user decision-making in a short time. Therefore, in the actual development process will intend to use the strategy, which is to build a data warehouse after establishing a data mart.
- (4) The data definition directly are inputted into the system, used for data management module and analysis. Metadata is stored in the meta database. It is not only the documentation of the data warehouse, used for management, maintenance staff, and used for the user query, that make users to better understand the structure of data warehouse, improve the level of the use of decision makers.
- (5) Using theme model to carry on dimensional model and physical design, and completes the data extraction, transformation and integration module design; the data management maintenance module design. Figure 2 is based on the original HIS relational database using Power Designer11.0 reverse engineering to extraction data.

2.2 Data Organization Model and Granularity Choose

The data in Warehouse of Hospital Drugs is very huge, because the purpose of service goal is different, the traditional data modeling method such as the Third Paradigm Model (3NF, the third normal form schema) has already can't adapt to the needs of the data model [5]. Here we use the star model and snowflake model to design the data organization.

Granularity is refinement and comprehensive level of the data in the data warehouse unit. It affects in the size of the amount of data in the data warehouse, storage space and the query type that user can answer for related in the data warehouse [2]. In the hospital drugs management, if the granularity of data warehouse to detailed records, so this level data warehouse can answer all the questions (such as a specific number of drugs to the department in one day). However, if the size is just arrived a week bill particle size, then the data warehouse can only do a few quantity statistics, and unable to answer queries about user behavior patterns.

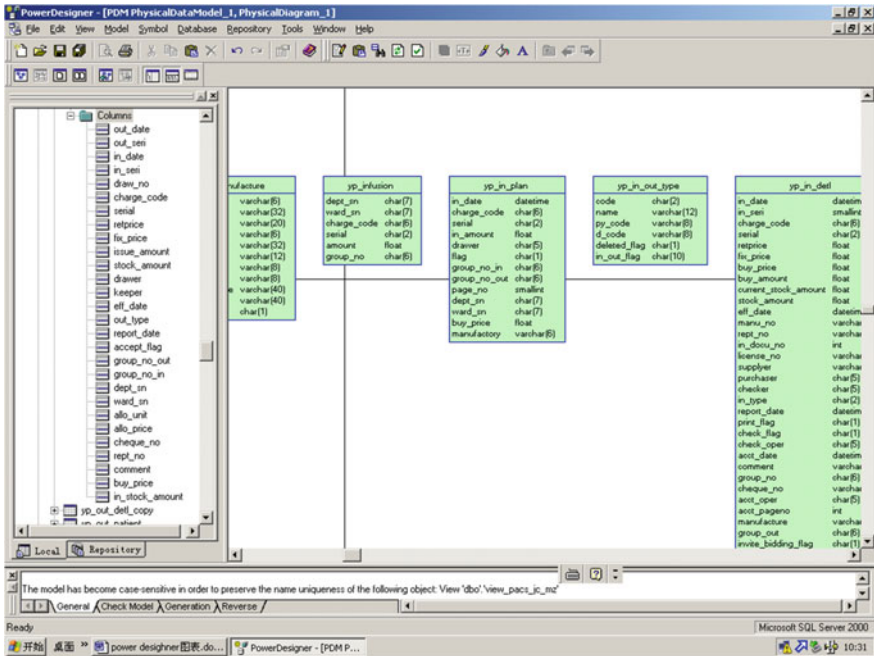


Fig. 2 Power Designer extract data using reverse engineering

Table 1 Hospital drug data warehouse and time-related granularity

Contents	Granularity	Storage capacity/way
Total monthly data	Highly integrated	Capacity is small, stored in a separate model, it may be physically in the data warehouse, it may in the multidimensional database
Daily bill	Mild integrated	Smaller capacity, stored in a data warehouse
Current medicines (reservation 6–12 months)	Least	Large capacity, stored in a data warehouse at the bottom
History drugs	Least	Great capacity, typically stored in tape

As one of the most important problems in data warehouse design, the design of the granularity size is directly related to the quality of the data warehouse, double (or multiple) level of granularity size been consider in the hospital drugs data warehouse. Table 1 shows several levels of granularity related time in the hospital drugs data warehouse.

3 Hospital Drug Data Warehouse Model Design

3.1 Basic Design Process

Hospital drug administration section mainly includes two parts, namely drugs management and clinical use. Drugs in the hospital from the inbound to the out-bound, until the use of the patient is a complex process, which runs through the entire patient’s clinical activities.

The design of the hospital drugs data warehouse of process can be divided into data warehouse model design and data load interface design, the design of the basic process as shown in Fig. 3:

The main task of drug management subsystem is to manage information which used pharmacy, preparation, outpatient pharmacy, hospital pharmacy, drug prices and drug and assist clinical rational drug use of clinical drug, including prescription or orders review for the rational use of drugs, drug information consultation, medication consultation, etc.

Based on detailed analysis and summary of the demand related the core business system of hospital drug business system such as all kinds of statistical analysis data sources. Summarize the hospital drug core business into the following four topics, drug supply subject, outpatient pharmacy subject, dispensary for inpatients subject, organization subject. Data is organized by subject, and the different themes have

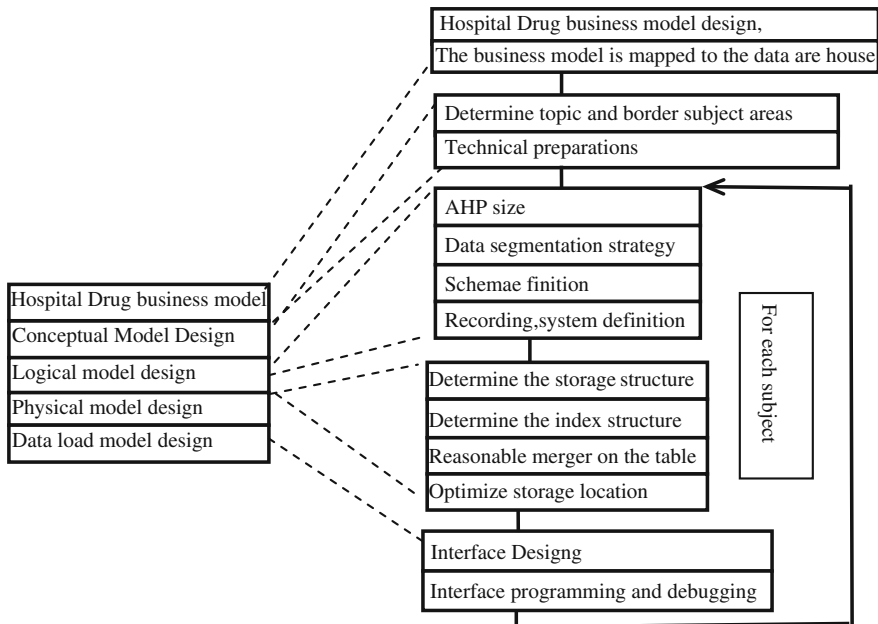


Fig. 3 Hospital drug data warehouse design process

relatively subordinate relations. Such as drug supply subject, its child theme for pharmacy, traditional Chinese medicine (TCM) preparations. Such as drug supply subject, its child theme for pharmacy, traditional Chinese medicine (TCM) preparations. Dimensions include TCM store, medicine Store, Chinese herbal preparation, Herbal medicine library [6].

3.2 *Logical Modeling and Physical Implementation*

Logical modeling is based on the theme of the model and is the basis for the physical design. Here we use dimensional modeling to achieve the logical view of the data warehouse hospital drugs. According to the results of dimensional modeling, design table in Sybase environment. These database tables as the core of data warehouse tables, mainly are divided into the core fact table and dimension table, gathered the fact table.

(1) Hospital Drug Data Warehouse Bus Architecture Matrix

Overall construction idea is to first construct a data warehouse bus, which provides a framework of data warehouse. Data must be dimensions, atomic and attached to the bus structure of data warehouse [7]. New data mart can be added according to the urgency of the user's demand and gradually build into an integrated data warehouse. The whole process is a continuous change and development of iterative process. Including the creation of the data warehouse bus architecture (list data mart, list the various dimensions, identify intersection), the design of fact table (select a data mart, statement granularity, choice dimension, choose the truth), design unified dimension table, gather the design, etc. the design of bus structure table as shown in Table 2.

(2) Design of Fact Tables and Dimension Tables

As the main data items of data analysis, the fact table can store data in different granularity. If it is relatively simple theme, so a theme corresponds to a fact table. If it is a more complex analysis theme, probably a theme corresponds to multiple fact tables. The fact in the fact table that possess general data characteristics and additivity, this feature is very important for analytical applications [4].

All the fact table size belong to one of three categories: transaction, periodic snapshot and accumulating snapshot. Dimension table is the entrance into the fact table, the dimension table generally contains descriptive text information. This text information will become retrieval conditions of the fact table. Such as query sales information by region classification, or quarterly review drug sales trend, etc. D hierarchy level number depends on the granularity of the query. In the actual business environment, multidimensional data model is typically contains 4–15 D.

According to the bus structure matrix Table 2 shows, choosing a data mart, such as drug supply western medicine clinic library, design fact tables. When design a fact table, we need design its grain size, fact table has two typical particle size, a

Table 2 Bus structure table

Data mart	Dimension										User				
	Time	Drugs	Drug sector	Drug access type	Checkout	Drug properties	Drug classification	Drugs pharmacy ledger	Please collar	Prescription confirm		Period of limitation	Stock search	Invoicing inquiry	Aliquot
Drug supply western library	√	√	√	√	√	√	√			√	√	√	√	√	√
Herbal medicine supply library	√	√	√	√	√	√	√			√	√	√	√	√	√
Proprietary drug supply	√	√	√	√	√	√	√			√	√	√	√	√	√
Chinese medicine	√	√		√	√	√	√	√		√	√	√	√	√	√
Outpatient medicine library	√	√		√	√	√	√	√	√	√	√	√	√		√
Outpatient medicine library	√	√		√	√	√	√	√	√	√	√	√	√		√
Outpatient herbal library	√	√		√	√	√	√	√	√	√	√	√	√		√

(continued)

Table 2 (continued)

Data mart	Dimension										User				
	Time	Drugs	Drug sector	Drug access type	Checkout	Drug properties	Drug classification	Drugs pharmacy ledger	Please collar	Prescription confirm		Period of limitation	Stock search	Invoicing inquiry	Aliquot
Hospital medicine library	√	√		√	√	√	√	√	√		√	√	√		√
Hospital medicine library	√	√		√	√	√	√	√	√		√	√	√		√
Herbal hospital library	√	√		√	√	√	√	√	√		√	√	√		√
Pharmacy department	√								√			√	√		√
Outpatient pharmacy	√								√			√	√		√
Hospital pharmacy	√								√			√	√		√
Drug library	√								√			√	√		√
Preparation room	√				√							√	√		√

type is transactional, such as drug storage, a library form a record of the fact table, granularity is the time of the transaction happen, another is snapshot model, such as patient information, its drugs purchasing state change only when the transaction occurs, its particle size use the snapshot method. In type snapshot data, need to design size conversion calculation. there are two major steps on conversion step:

- Step 1: Every day collect data with the method of snapshot into snapshot fact table as far as possible using incremental acquisition method. If can judge the data source record change of circumstances, the only collect the change data, after the completion of the acquisition according to the data don't change data to generate a new day yesterday;
- Step 2: if a new day is the first day every month then process the data of the last month, to generate record of last month and put it to the corresponding fact table. If for the first day of the New Year again, to deal with last year's data, and to generated a record at the end of year and load into the fact table for particle size corresponds. And according to the detail record retention time to clean up records more than retention time.

(5) Design of Core Fact Table, Consistency Dimension Table and Gathered Fact Table

In the dimension model, fact tables show the many-to-many relationship between dimensions. The most using fact in fact table is the digital type and added type fact, so when designing the fact table, as long as possible, the facts in fact table should try to choose completely addition facts.

According to the analysis of business, the main fact table are warehousing drugs register, drug procurement plan, drug procurement plan; outbound drugs register, etc. Such as drug purchase plan: granularity for transactional, the dimension is drug purchasing plan, purchasing plan summary sheet, the optimal procurement analysis, date, trading, etc.

Consistency dimension, often referred to as granular (atomic) dimension, through intersection calculation of all these granular dimension, the various associations fact table can easily generate. Using consistency dimensions can avoid creating not compatibility data mart [3]. Through the above the fact table in logic design in detail, we get consistency dimension table: institutional dimension, the drug dimensions, billing dimension; Pharmacy drugs parameter dimensions, etc. For example the time dimension: the main attributes are day, holiday, festival, week, calendar week, calendar month, calendar quarters, calendar year, financial week, financial month, financial quarter, financial year.

Issues paid attention to the design is (1) dimension table design can not be too big. Larger dimensions appear not only very clumsy, but also can cause performance problems, and has no any user interface advantages. (2) suggesting use the time dimension. Each data warehouse fact table is the time series really after sorting. (3) using key words in the dimension model, that can not only buffer operation change for the data warehouse, but also support to revise dimension table attributes.

Gather mainly used to improve the query performance, when design the gathered strategy there two basic factors should to be consider. First of all, the business user's access mode need be considered. Second, the statistical distribution of data need been evaluated; as far as possible to make different types of users get better query performance; aggregating storing must be stored in the fact table, and separated from the underlying data. In addition, each different levels of gather must be with unique fact table. A complete set of gather might involve to ten several independent gathering of fact table, all these gather fact table reflect the original structure of the basic fact table.

Fully considering the need of statistical analysis is necessary in design of the gather table, according to the analysis of the business, we designed the gather tables is as follows: drugs for the summary table, warehouse drugs flow sheet, drug usage summary table. Such as drugs for the summary: the daily statistics shall be carried out in accordance with the pharmacy drugs, particle size for the day, the corresponding dimension of the pharmacy name: date, drug name, specification, the retail price, the initial inventory flow, flow quantity, amount, inventory, inventory amount, flow mode, head.

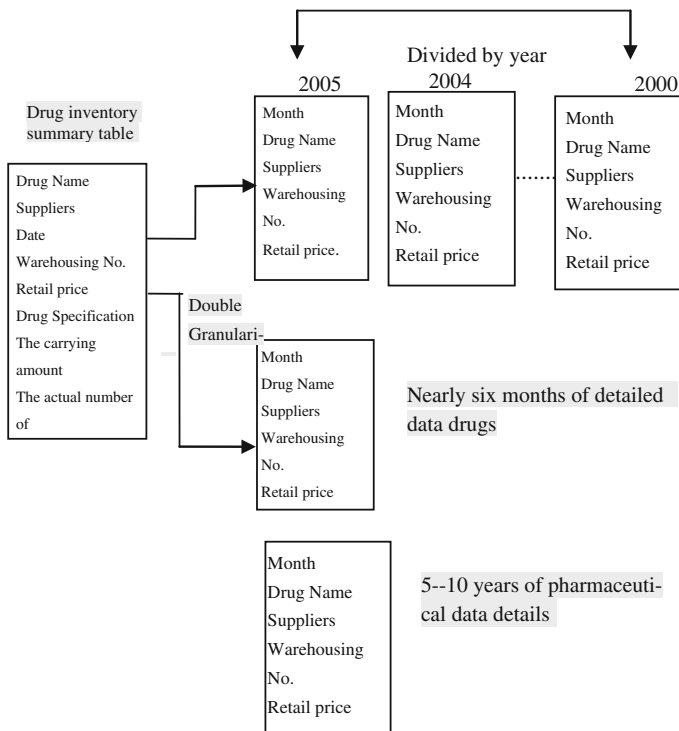


Fig. 4 Accordance with the time division of drug inventory summary

The physical design and implementation for data warehouse, you should solve data storage structure, data storage strategy, storage allocation optimization problem and so on. The principle used in which storage must use parallel storage structure data warehouse. Data storage strategy includes table merging, table segmentation, storage allocation optimization [5].

In the logical design phase often use segmentation to improve the performance of data warehouse and maintainability, the best way of dividing table is according to the date. Such as data segmentation according to monthly, quarterly or year. Usually only in the fact table, ultra-large dimension table and its indexes for segmentation. Figure 4 shows accordance with the time division of drug inventory summary table.

4 Research and Implement in Design Strategy

The below main discussed design strategy in the design of the fact table, dimension tables and gather consistent table.

4.1 Index Strategy

- (1) Date-key are should be placed at the primary key index. Using the star mode optimization function can improve the system performance Compared with using the traditional order link function. However, when the optimization function is enabled, a variety of special indexes should be created.
- (2) When the bitmap index of the dimension table is used with other indexes of the fact table, the query performance can be greatly improved. It should be created a single index for each column may will be used in the connection conditions, filter and grouping operation.
- (3) When the dimension table is large enough, creating indexes in multiple column that its dimension attributes often together with the use of the filter questions will be very useful.
- (4) Creating a single index for each key fact table, the rest of the things will been processed by the optimizer. B tree index should be created for the primary key of the fact table. The index plan for drugs sold is shown in the Table 3.

4.2 Table Design Strategy

The following is the design strategy in the design of fact table, consistent dimension tables and gather table these is used by dimensional modeling.

Table 3 Drugs sold index schedule

Index name	Index type	Whether only	Row	Location	Create a reason
<i>Drugs sold dimension table</i>					
Ypsc_detail_mkey	BitWise index	Y	Date_key Supplier_key Drug_key Pharmacy_key	The split in Ypsc_mkey_inx	Key index
Ypsc_detail_Date	Bitmap index	N	Date_key	The split in Ypsc_mkey_inx	Use the star join in user queries
Ypsc_detail_supply	BitWise index	N	Supplier_key	Ypsc_inx	Use the star join in user queries
Ypsc_detail_ypmc	BitWise index	N	Drug_key	Ypsc_inx	Use the star join in user queries
Ypsc_detail_yfmc	Bitmap index	N	Pharmacy_key	Ypsc_inx	Use the star join in user queries

- (1) The fact in fact table should try to choose completely additivity facts. This feature is very important for analytical applications.
- (2) In general, the particle size of the fact table as possible choose low levels of granularity. The lower level of granularity, the more robust design. Once determined the size of the fact table, the selection of the dimension becomes simple.
- (3) Empty key must been avoided in the fact table. Suitable design is including a line in the corresponding dimension table to identify the dimension of measurements is unavailable [8].
- (4) All the facts are sorted in order to eliminate duplication. They are grouped together based on the facts you need to put all the duplicate, together with core users to determine the basic facts and derived facts.
- (5) Using the agent key, which can not only to buffer the change of the operation data warehouse model, but also to support changing of the dimension table properties. Each of the connection between dimension and fact tables should be builded integer agent keywords with no clear meaning and to avoid using natural operational product coding.
- (6) The attribute in the dimension table should be specific, reflects the division of dimension hierarchies, and can become the constrain conditions of analytical queries. This is differences point of data warehouse and operational application in the data model design.

4.3 *Aggregate Design Strategy*

- (1) In the aggregation operation of the facts, either eliminate the dimension or link the fact and the accumulation dimension. Thus, the aggregation dimension table can be consistent with the basic dimension table.
- (2) The aggregation must be stored in its fact table and separated from the underlying data. In addition, each aggregation of different levels must be stored in individual tables with its unique fact tables, rather than stored in the original fact tables containing non clustered data.
- (3) In determining the content of the gathered, requirements documents need to be review, identify useful needs and find a higher level of demand. All dimensions should be reviewed to determine the properties that can be used.
- (4) A complete aggregation table may involve a dozen independent gathered the fact table, all of these gather the fact table must reflect the original structure basis of the basic fact table.

5 Discussion and Conclusion

In the light of the business situation of the hospital medicine management, this paper use the “dimension of business life cycle method” to finish building the hospital drug data warehouse system, and carried on the modeling design four aspects from topic model, logic modeling, physical design and data dump and development. Have studied detailed the question existing the design model of data warehouse and its application, and complete the design in detail of Indexing strategy, granularity conversion algorithm and form design strategy in the design process, through validation, overall design is clear, logical structure model building method is practical, For all building data warehouse of hospital drug has good reference value.

Bibliography

1. Tang DP, Wang XY (2003) Based on data warehouse to the pharmaceutical industry enterprise information portal. *J South China Univ Sci Technol* (7):13–17
2. Inmon WH (2004) *Building the Data Warehouse*, 3rd edn. Wiley Computer, USA
3. Dragon, Lei YJ, Zhang DP (2004) Based on the structure of the data warehouse with reference to the general framework and application. *Comput Eng Des* 25(3):148–150
4. Kimball R, Reeves L, Ross M, Thornthwaite W (2004) *The data warehouse life cycle toolkit: expert methods for designing, developing, and deploying data arehouses*. Wiley, USA, p 1
5. Zhu LY (2003) *Data warehouse principle and practice*. Beijing: people’s posts and telecommunications publishing house, pp 166–179
6. Adelman S (2003) *The difficulty of data warehouse solutions*, Beijing experts. Electronic Industry Press, p 1

7. Yan YY, Hai CW, Zhong SW, Fuqing Y (2002) Data warehouse star model and modeling tools. *J Softw* 12:23-25
8. Yeung S, Wang H (1999) Facing the theme of the data warehouse system structure. *Comput Appl* 19(10):107-109
9. Jiang XD, Zhou LZ (2001) The data warehouse inquires. In the processing of a table connection algorithm. *J Softw* 12(2):32-37

Question Recognition Based on Subject

Li-fang Huo, Li-ming Zhang and Xi-qing Zhao

Abstract Question analysis is an important component of a general Question Answering (QA) system. Question analysis has different angles and functions. The paper focuses on recognition of question subject in QA system. The goal of subject recognition is to identify given question according to special domain. We discuss three approaches to identify subjects of questions, then quantitatively evaluate effect of machine learning methods by a series of experiments. The results show that Naive Bayes gains the best accuracy and efficiency than other learning methods and two ways of feature extraction proposed by the paper improve accuracy for most of learning methods.

Keywords Question-answering · Recognition · Subject

1 Introduction

Question answering system, especially open-domain question answering is a retrieval task more challenging than common search engine tasks. Lots of research groups focus on solving the problems and have received much attention and research production [1–3]. A general QA system includes three important modules, that is, question analysis, passage extraction and answer integration. Question analysis module understands questions in nature language and user’s intention from different angles. Now, much research about question analysis focuses on classifying questions according to wh-word (what, which, who etc.) or answer type/class [4–8]. In fact, recognition of question subject/domain is the first step in question analysis. If question domain is known, a special domain QA system will decide whether to continue to parse the question and an independency-domain QA system can decide

L. Huo
Hebei Institute of Architecture and Civil Engineering, Zhangjiakou, China

L. Zhang · X. Zhao (✉)
Hebei North University, Zhangjiakou, China
e-mail: zxqlytqq@163.com

that which website or knowledge base will be used to collect useful information. The paper discusses questions analysis from different angles and focuses on recognition of question subject in QA system. We discuss three approaches to get subject of questions, then quantitatively evaluate effect of machine learning methods on auto identifying subject of questions by a series of experiments. The results show Naïve Bayes gains the best accuracy and efficiency than other learning methods and two ways of feature extraction proposed by the paper improve accuracy for much of learning methods.

The paper is organized as follows: Sect. 2 introduces question analysis. Question recognition based on subject will be described in Sect. 3. Section 4 discusses our experimental study. Related work will be introduced in Sect. 5. At last, Sect. 6 sums up conclusions and future work.

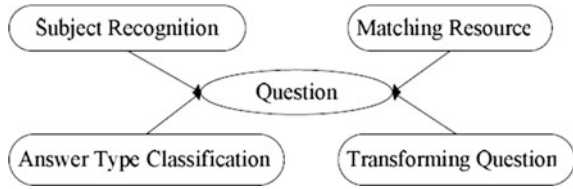
2 Question Analysis

Question analysis has different schemes and directions. The literature [9] summarized five general schemes. They are question word like “whom” “where” “when” and so on, subject of questions, functions of expected answers, forms of expected answers, types of sources respectively. In current QA field, much interest for lots of research groups focuses on question words classification and answer types/classes classification [5–8]. In fact, recognition of question subject and deciding types of source for QA system is also important.

Every question has a specific context, which is essential to find right answer. If a question dose not refer to context, for example “Why do you stay there?”, any QA system even the most intelligent QA system—people can’t give right answers. Recognition subject of questions is to understand context or domain of these questions. The direct aim of the work is locating suitable knowledge, decreasing irrelevant knowledge and improving accuracy and efficiency ultimately. For example, given the question “What makes crickets chirp at different speeds?”, if the QA system understands that the question belongs to “zoology” domain or subject, it can prefer to use knowledge from website related to zoology. It is such knowledge that is easier to answer the question. Lots of existing resource like websites, databases and Ontology is organized or linked based on topic or subject of knowledge in nature. More methods on the research will be used to cluster or integrate these source based on topic or domain. So, when these data, especially more and more reasonable languages such as OWL or rule ML become knowledge source used to answer questions, identifying subject of questions can be more important for improving efficiency of full QA system. Because the work largely decrease knowledge used to inference.

Research for types of source is to consider which format of knowledge is suitable for answering given question. Different users have different goals. Some users want to know practice approach or effective process, so they can ask the question like “How do I download images from certain experiments?”. In order to answer the

Fig. 1 Question analysis



question, a video about operation flow can be used besides text files. Other only want to know a simple name, they can ask the question like “Who is chairman in the China?” For the question, many sources can get right answer, for instance introduction file and Ontology instance. So matching resource is also an important research direction in question analysis.

Another important work is question transformation. Aiming at different knowledge source, there are all kinds of transformation formats. If knowledge source is web page, questions must be rewrite into different query phrases special for search engine according to manual rules or learned models. For Ontology source, question must be transformed into right format or triple that can directly reason or query on Ontology knowledge. In a word, question analysis has different directs as Fig. 1. We must select different work and combine them based on real requirement of QA system.

3 Question Recognition Based on Subject

The section discusses common methods of subject recognition and feature extraction. At last, it gives an abstract workflow of recognition subject in our QA system named Ask Me.

3.1 Subject Recognition

Many methods can be used to identify subjects of questions. Among, user response are the simplest method. When they propose questions, they propose the subjects of questions at the same time. However, the method has several problems. Much people are unwilling to fill additional item other than their questions, although this can lead some benefits. Some people don’t know which domains their questions belong to.

Another method is building manual rules based on analysis of lots of questions. We analyze 200 questions in TREC81 and build a rule for part of questions as following: if a question only includes a noun phrase, we extract it and find its hypernyms from Word Net as subject of the question. For example, given a question “Do worms sleep?” that satisfies the above rule, “worms” is extracted and

its hypernyms in Word Net “animal” is subject of the question. However, for lots of questions we can’t build right rules and this manual work is prohibitively expensive.

To identify subjects of questions by machine learning methods is a fascinating idea. The task belongs to supervised text classification in nature. On the one hand, it only needs questions enough in different domains, which is easy to acquire because there are lots of frequently asked questions² in many domains. On the other hand, there are lots of mature machine learning methods such as Naïve Bayes to be used to classify the text. We used a special machine learning tool Weka [10]. The Weka is implemented in Java and has a good GUI and convenient APIs. It includes all kinds of general Classifiers such as bayes series, functions series, tree series and so on. We will quantitatively evaluate some methods and select suitable methods to build subjects recognition model for our prototype QA system named Agile.

3.2 *Feature Extraction*

Though questions subjects classification is text classification in nature, feature extraction and vector representation is different from it in text classification. Generally, the most important distinction is domain vocabularies between different domains. Questions in same domain will more or less include the words in domain vocabularies. So, we can collect domain vocabularies from positive questions set as feature space, that is, unigram of every question in same subject can be used as main feature. Before key words extracting, we must remove stop words including wh-words such as who, where and so on, the reason is that these question words are independent from subjects, from these questions. After reediting stop word list, the method is both simple and straightforward. Corresponding feature set is a token set. However, the feature set includes lots of redundancy tokens. For instance, “computer”, “computing”, “computed” and “computers” in the same subject questions can appear in a feature set. From semantic angle, these words obviously represent similar meaning and can be substituted by one feature. Common grammar and spelling errors can also import another type of redundancy information. For example, the word of CD in computer domain has four type spelling as following: “Cd”, “CD”, “cd”, “cD”. To avoid effect of these syntactic formats, the paper uses two ways to amend candidate feature set acquired by key words extracting. One way is extracting stem from candidate tokens as new feature. For example, four tokens “computer”, “computing”, “computed” and “computers” are replaced by one stem “comput”. The way can decrease number of tokens in feature set and improve infection of every token on final classification. Another way is ignoring capital-sensitive. Through the process, one token “cd” is substituted for four different tokens such as “Cd”, “CD”, “cd”, “cD”. The way also can decrease redundancy tokens and is helpful to avoid some spelling errors. Deciding feature set of a domain, a binary feature vector can be used to represent each question

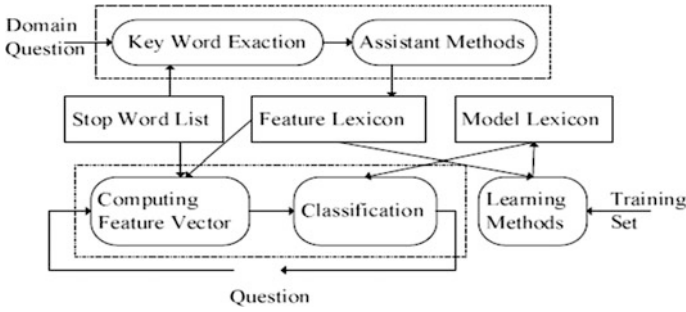


Fig. 2 Questions recognition of the agile

according to whether corresponding tokens in feature set appear in the question as the literature [6–8].

3.3 Recognition Workflow

Figure 2 is the full workflow of building model and subject recognition in our QA system named Agile. The first stage is building Feature Lexicon for given QA system involved question domains based on collecting different domain questions. The second stage is learning model of different domains from different training sets by machine learning methods. Finishing the two stages, inputting question can be classified into certain subject.

4 Experiments

4.1 Experimental Data

We collect three data sets of different type questions with different subjects from MadSci Circumnavigator3 as Table 1. MadSci is a science web network that represents a collective cranium of scientists providing answers to your questions that involves in a few domains. The questions are proposed during 1996–2005 and raw questions have some grammar errors and spelling errors. In order to compare these data with noise, we correct MadSci questions with lots of errors as another team of

Table 1 Questions from MadSci circumnavigator

Subject	Computer science	Medicine	Zoology
Raw questions	709	1787	2142
Correct questions	570	1358	1928

data. There are two reasons for recognizing subject of questions with noise. One is to improve QA system fault-tolerant ability. Another is to select user's general error words from learned feature words in given subject as lexicon that can be used to check user's some common errors and revise them. For instance, in zoology subject we find some spelling error such as "moquito", "mosquite", "mosiquotes" for "mosquito". If we build a lexicon for "mosquito" and these spelling errors, when users input similar error, the QA system can find them and revise them or feedback users to confirm. We randomly collected same number questions in other domains as negative instances in two training sets.

4.2 Experimental Results and Analysis

We conducted a series of experiments to compare the effectiveness of learning methods, and to show effect of ignoring uppercase-sensitive and extracting stem in feature extraction on accuracy and availability for questions with noise. We used Naïve Bayes, RBF Network, Voted Perceptron and SMO in the Weka as our experiment platforms. All methods are set up default options. To satisfy format of data file in the Weka, we implemented feature extracting and building vector methods in Java.

Table 2 is learning results for selected algorithms from the Weka. Feature only considers bag-of-word. From accuracy facet, we know Naïve Bayes, SMO are better methods as Table 2. From efficiency angle, the Naïve Bayes is the best method in the Weka as Table 3. The Voted Perceptron and SMO have got high accuracy in some subjects, but their efficiency is very bad. Another problem is that some methods such as SMO are not balance to positive instances and negative instances. They can precisely identify negative instances but for positive instances can't precisely recognize in same data of questions. For example, in 4-fold

Table 2 Accuracy of learned methods

Methods		NaiveBayes (%)	RBFNetwork (%)	VotedPerceptron (%)	SMO (%)
2-fold	Computer	74.38	72.51	75.89	74.91
	Medicine	75.03	70.42	71.78	69.64
	Zoology	76.22	70.92	73.52	75.49
4-fold	Computer	74.02	76.42	79.36	79.80
	Medicine	74.38	69.51	73.60	71.39
	Zoology	77.36	71.13	74.45	79.02

Table 3 Time of building model for computer domain

NaiveBayes (s)	RBFNetwork (s)	VotedPerceptron (s)	SMO (s)
2.19	9.17	21.36	55.44

Table 4 4-fold results about different feature extraction methods

Methods		NaiveBayes (%)	RBFNetwork (%)	VotedPerceptron (%)	SMO (%)
Computer	Case-sensitive	74.02	76.42	79.36	79.80
	Ignore case	81.85	76.87	81.76	81.94
	Stem	82.30	76.07	81.58	82.65
Medicine	Case-sensitive	74.38	69.51	73.60	71.39
	Ignore case	76.20	70.74	73.60	72.56
	Stem	77.24	68.79	74.18	73.67
Zoology	Case-sensitive	77.36	71.13	74.45	79.02
	Ignore case	80.37	72.38	77.15	80.11
	Stem	78.56	70.30	75.29	81.10

Table 5 Common errors in computer science

Wrong words	Right words	Wrong words	Right words
Intneret	Internet	Toknow	To know
Advantges	Advantages	Sotware	Software
Compuer, computor	Computer	Theorys	Theories
Physically	Physically	Transmited	Transmitted
Recomended	Recommended	Algoritms	Algorithm

computer experiment SMO identify 544 negative instances from 562 negative instances but only identify 353 positive instances from same number positive instances. Apart from methods in Table 2, we also test other methods such as J48 (C4.5 Decision Tree (DT) algorithm), their accuracy is too low to be suitable for the recognition of question subject.

In order to compare different feature extraction ways’ effect on accuracy of recognition subject of questions, we explored a series of experiments as the following Table 4. The results in Table 4 show that two ways remarkably enhance accuracy of recognition subject of questions special for Naïve Bayes. Both ways have similar function, which is reducing feature, so impact on accuracy also is similar.

We find part of common errors in computer science from its feature file and corresponding to right words as following Table 5.

5 Related Work

Questions recognition based on subjects is a task of text classification in nature. But including content in question is much less than general text file. So feature extraction and the most appropriate learning method is completely different from the literature [11, 12].

In QA research field, the best related work is question classification according to answer types [5–8]. We use same binary vector to represent feature in answer type classification for reference. However, two works focus on different feature component. For example, one important feature in answer type classification is question word such as who, how many, where. In our work, these words are not useful to identify subject, so they can be as stop words. The paper uses two different syntax ways in feature extraction based on characteristic of questions subject.

6 Conclusion

Question analysis is an important problem in building a QA system. The paper discusses questions analysis from different angles then quantitatively evaluate effect of machine learning methods on subjects recognition by a series of experiments. The results show that Naïve Bayes gains the best accuracy and efficiency than other learning methods and two ways of feature extraction proposed by the paper improve accuracy for most of learning methods.

The work is a part of our QA system named as Agile. Auto learning to identify suitable resource and question transformation are on the research.

References

1. Wang C, Xiong M, Zhou Q, Yu Y (2007) PANTO: a portable natural language interface to ontologies. In 4th European semantic web conference. Springer, Berlin
2. Lopez V, Pasin M, Motta E (2005) AquaLog: an ontology-portable question answering system for the semantic web, ESWC 2005, LNCS 3532, pp 546–562
3. Gao MX, Liu JM, Zhong N, Chen FR, Liu CN (2011) Semantic mapping from natural language questions to OWL queries. *Comput Intell* 27(2):280–314
4. Bu F, Zhu X, Hao Y, Zhu XY (2010) Function-based question classification for general QA. In: Proceedings of the 2010 conference on empirical methods in natural language processing, pp 1119–1128
5. Li X, Roth D (2002) Learning question classifiers. In: Proceedings of the 19th international conference on computational linguistics, pp 556–562
6. Metzler D, Croft WB (2005) Analysis of statistical question classification for fact-based questions. *Inf Retrieval* 8:481–504
7. Zhang D, Lee W (2003) Question classification using support vector machines, SIGIR'03, pp 26–32
8. Cheung Z, Phan K, Mahidadia A et al (2004) Feature extraction for learning to classify questions, AI 2004. LNAI 3339:1069–1075
9. Pomerantz J (2005) A linguistic analysis of question taxonomies. *J Am Soc Inform Sci Technol* 56(7):715–728
10. Witten I, Frank E (2000) WEKA machine learning algorithms in Java, data mining: practical machine learning tools and techniques with Java implementations. Morgan Kaufmann Publishers, Burlington

11. Chakrabarti, S, Roy S, Soundalgekar MV (2002) Fast and accurate text classification via multiple linear discriminant projections. In: Proceedings of the 28th VLDB conference, Hong Kong, China
12. Bloehdorn S, Hotho A (2004) Boosting for text classification with semantic features. In: Proceedings of the workshop on mining for and from the semantic web at the 10th ACM SIGKDD conference on knowledge discovery and data mining, pp. 70–87

Development of a Mobile Augmented Reality System to Facilitate Real-World Learning

Kai-Yi Chin, Ko-Fong Lee and Hsiang-Chin Hsieh

Abstract A number of research studies have explored the impact of applying Augmented Reality (AR) technology to real-world learning environments. These studies have asserted that AR can improve students' perceptions and enhance overall cognitive abilities when engaged in real-world learning activities. However, it is not easy for teachers to implement AR-based learning systems in classrooms because many teachers lack the skills and abilities of computer professionals or coding experts. In this study, we created an easy to use mobile augmented reality system that can support teachers in creating and designing AR materials. This mobile system provides teachers with the ability to combine course content and multimedia materials in a way that promotes learning within an engaging and intuitive AR-based environment. A quasi-experimental research design was used to evaluate the feasibility of using our proposed system to implement a variety of teaching activities. From the results of the questionnaire survey, we discovered that respondents rated the proposed system positively and were willing to formally incorporate mobile augmented reality into their future teaching plans. Therefore, we believe that teachers do regard our mobile augmented reality system as a useful tool that can supplement existing real-world learning activities with distinctive AR capabilities.

Keywords Augmented reality · Mobile learning · Mobile AR system

K.-Y. Chin (✉)

Department of Digital Humanities, Aletheia University, New Taipei City, Taiwan
e-mail: au0292@mail.au.edu.tw

K.-F. Lee

Department of Information Engineering and Computer Science, Feng Chia University, Taichung City, Taiwan
e-mail: kookyrational@hotmail.com

H.-C. Hsieh

Institute for Information Industry, Taipei City, Taiwan
e-mail: palapala@iii.org.tw

1 Introduction

In recent years, Augmented Reality (AR) technology has become more widely used in a number of different applications. This technology allows users to combine observable, real-world phenomena with animated graphics, textual information or inserted images, and creates an enhanced and augmented reality that can assist in amassing knowledge [1–3]. The exciting new possibilities of teaching traditional classroom concepts using AR are welcomed by educational researchers since the opportunities associated with enhanced learning is greatly fostered through technical advancements. As such, it has been demonstrated that AR enhances students' senses through the use of virtual or naturally invisible information superimposed on top of real-world objects or spaces [2, 4]. Therefore, it is accepted that AR can transfer educational experiences and knowledge from the classroom to applications in real-world learning environments. Such documented benefits to the field of education will promote AR as the key emerging technology preferred by educators for supplementing course content over the next decade [1].

Several studies have emphasized that AR can improve students' perceptions, knowledge, and interaction with the real-world, and it is known that AR has been used to enhance students' cognitive abilities during real-world learning activities and tasks [5, 6]. For example, Angela Di Serio and her colleagues proposed the implementation of an AR system to motivate middle-school students tasked with learning key facts about Italian Renaissance Art. In their study, AR was used to enhance each artistic masterpiece with added textual facts explaining specific details unique to the work of art, while simultaneously superimposing digital data (such as text, audio, 3D models, etc.) directly on top of the masterpiece [2]. The researchers also proved that the use of AR technology can promote higher levels of engagement in students utilizing enhanced learning tools, and thus it was found that educational activities with AR could confer a positive effect on students' learning outcomes. In another study, Chih-Ming Chen and his colleagues proposed an AR library instruction system that trained students to learn how to use libraries. This system integrated interactive 3D virtual technology with the physical library environment to generate a novel context-aware library instruction module. This learning module could enhance students' perception of reality, improve the effectiveness of new knowledge acquisition specific to their library and made the process of learning much more alluring for students using the interactive AR system [7].

However, like many educational and technological innovations introduced within the past century the incorporation of AR systems into classrooms may encounter resistance among teachers [1]. The opposition to widespread adoption of AR systems into classrooms stems from two main issues that educators face. First, most AR systems on the market today offer limited and non-customizable learning activities that come pre-programmed with integrated learning content. Thus, teachers usually do not have the opportunity to incorporate appropriate learning materials specific to their students into existing AR systems. Second, the development of educational programs is an arduous task that demands much effort, input

and skill from computer experts [8, 9]. It is commonly known that many teachers do not have sufficient programming knowledge and they lack the capability to create AR systems by themselves. If teachers want to be able to independently customize an AR system, specialized training and continuing education courses must be provided by computer professionals. For the above reasons, it is likely that AR systems cannot be strongly promoted or applied to a variety of educational endeavors.

To overcome the aforementioned limitations, we developed a mobile augmented reality system that was designed as an instructional adjunct to aid teachers in the creation of AR enhanced educational activities. This system not only provided a convenient authoring tool to support teachers in managing and designing multimedia materials from scratch, but also combined educational materials with AR technology to create an engaging AR-based learning environment. In other words, our mobile augmented reality system allows teachers to create AR-based multimedia learning systems without any previous experience with computers and prior coding knowledge is not required. Teachers can easily provide an interactive AR-based learning environment that fosters students' interest and engagement in real-world learning activities.

In addition, this study also included a trial experimental survey that collected insights from teachers' remarks on the overall usefulness of our proposed system. The questionnaire survey was used to assess the feasibility of implementing teaching tasks using mobile augmented reality systems. The results of the questionnaire demonstrated that teachers rated the system positively and were willing to accept and adopt our AR system into their classrooms. Teachers also agreed that our system was a useful tool that could enhance the quality of outdoor educational activities. All in all, these findings reveal that our proposed mobile augmented reality system could in fact give teachers the chance to add pedagogical value to their courses, and further help learners engage in authentic exploration of resources found in real-world environments.

2 Related Work

The term AR is regarded as a novel interactive technology that can add virtual information to users' physical contexts, and enhance sensory perceptions of the real-world with the addition of a computer-assisted contextual layer of information [2, 6]. According to the above mentioned studies, AR can be broadly defined as "a situation in which a real world context is dynamically overlaid with coherent location or context sensitive virtual information". From this definition, AR can assist users through technology-mediated immersive experiences in which real and virtual worlds are blended, and further augment the interactions and overall engagement of the user [10]. Additionally, AR could be initiated and implemented through a number of available technologies, including desktop computers, mobile devices, head-mounted displays, wearable computers and so on [1, 11]. This

technological compatibility across multiple platforms means that AR can be used in many domains and applications, such as medical visualization [12], maintenance and repair [13], robot path planning [14], and entertainment [15].

Nowadays, research teams are enthusiastic about applying AR technologies to different teaching and learning activities [2]. For example, Tsung-Yu Liu and his colleagues proposed an AR-supported mobile system that supplied interesting learning activities to increase students' motivation in learning English [11]. Researchers You and Neumann developed a mobile AR system that provided an efficient interactive online virtual assistant to enhance students' performance in their museum guide course [16]. Additionally, Utku Kose and colleagues proposed a mobile AR tool that led to improved learning experiences for students enrolled in abstract or technical courses [5]. Of note, Jose Manuel Andujar and his colleagues proposed the creation of augmented remote scientific laboratories that would allow students open access to the lab to practice development boards via an Internet connection [17]. The study by Andujar went on to prove that the interactive online laboratory platform could actually improve students' learning outcomes in the fields of science and engineering.

The above-mentioned studies all clearly indicate that AR in the educational context can be very valuable. However, we noticed that despite the demonstrated benefits of using AR within the classroom, few research studies examined the benefits of providing tools for teachers to actually create AR enhanced systems on their own. Thus, in contrast to the studies highlighted in this paper, our study specifically focused on supporting teachers in building and maintaining the AR-based learning environment. We anticipated that the proposed mobile augmented reality system could support teachers in creating and updating instructional materials, while also enabling them to combine relevant teaching content with AR technology to provide enhanced lessons to students.

3 System Overview

This study describes a proposed mobile augmented reality system that serves to improve the process of creating AR materials, enhance students' motivation during real-world learning activities, and develop authentic AR-based learning environments. As portability and mobility are necessary factors for using AR in real-world learning experiences, our proposed system requires the use of touchscreen mobile devices. Moreover, our mobile augmented reality system possesses cross-platform capabilities, so it has many exciting applications for enterprise, tourism and entertainment purposes.

As shown in Fig. 1, the mobile augmented reality system is composed of two major sub-systems: an AR-based mobile learning system and an AR materials remote server. The AR-based mobile learning system allows students to access AR materials via their mobile device, and has the ability to overlay virtual content on top of the QR code. The AR materials remote server enables teachers to combine

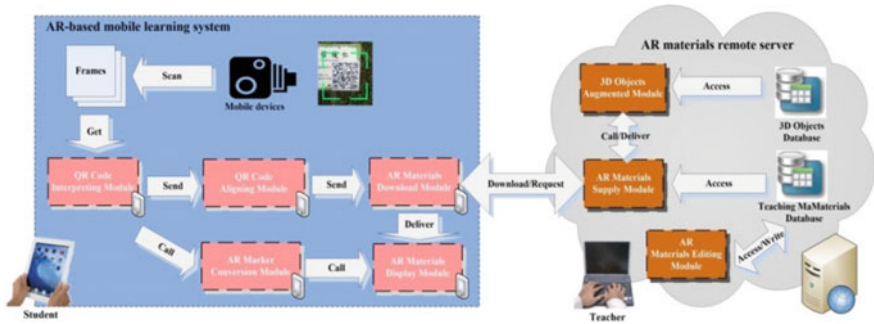


Fig. 1 The architecture of the mobile augmented reality system

course content and multimedia materials in a way that yields a complete AR learning package. Our design allows the two sub-systems to work together and enables teachers to create AR-based learning environments through a simple user interface. Together, the sub-systems make it possible to provide personalized learning opportunities for students and offers a more interesting learning experience. A detailed explanation of both sub-systems is presented in sections A and B of this paper.

3.1 The AR Materials Remote Server

As shown in Fig. 1, the AR materials remote server has three main modules (the AR Materials Supply Module, the 3D Objects Augmented Module and the AR Materials Editing Module) along with two databases (the Teaching Materials Database and the 3D Objects Database). When the AR Materials Supply Module receives a request to provide AR materials, the module will first access the related teaching content and resources from within the Teaching Materials Database and then follow up with a call to the 3D Objects Augmented Module. The 3D Objects Augmented Module is able to locate existing 3D virtual objects from the 3D Objects Database and subsequently instruct the delivery of virtual objects to the AR Materials Supply Module. Then, the AR Materials Supply Module can integrate all of the instructional materials into one AR enhanced material, which is defined as a lesson. The AR Materials Editing Module is equipped to provide a webpage-style visual editor from which teachers can create and maintain existing AR materials using just a personal computer and Internet connection.

Figure 2 shows screenshots of the visual editor that is an integral part of our proposed mobile AR system. Teachers must first select course content and multimedia materials using the visual editor upload page (see Fig. 2a). Then, teachers can choose one of many pre-programmed 3D virtual objects from the 3D Objects Database (see Fig. 2b) and select the corresponding QR code from the visual editor

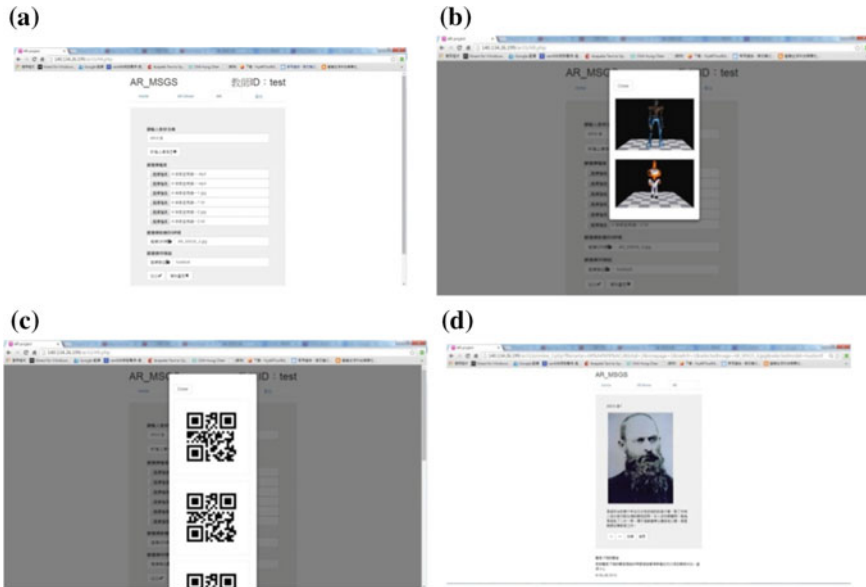


Fig. 2 Screenshots of the webpage-style visual editor

panel (see Fig. 2c). After the completion of the course, this visual editor also provides teachers with the option of reviewing the completed instructional content that was presented to the students (see Fig. 2d). Finally, the easy to use, and familiar webpage-style layout of the visual editor can help teachers more effectively integrate teaching content and resources into an interactive AR platform.

3.2 *The AR-Based Mobile Learning System*

Figure 1. shows how the five modules of the AR-based mobile learning system interact with each other to perform necessary tasks. Our proposed system consists of the following modules: the QR Code Interpreting Module, the QR Code Aligning Module, the AR Materials Download Module, the AR Marker Conversion Module and the AR Materials Display Module. When students scan the QR Code through a camera-equipped mobile device, the QR-based mobile learning system will automatically identify frames and get the embedded information contained within the QR code. Then, the QR Code Interpreting Module functions to interpret the QR code information and convert the coded information into text that will be passed to the QR Code Aligning Module. The QR Code Aligning Module is used to align QR code information, and then send a message to the AR Materials Download Module to correctly download the corresponding AR materials and 3D virtual objects., The AR Marker Conversion Module works in parallel to the QR Code

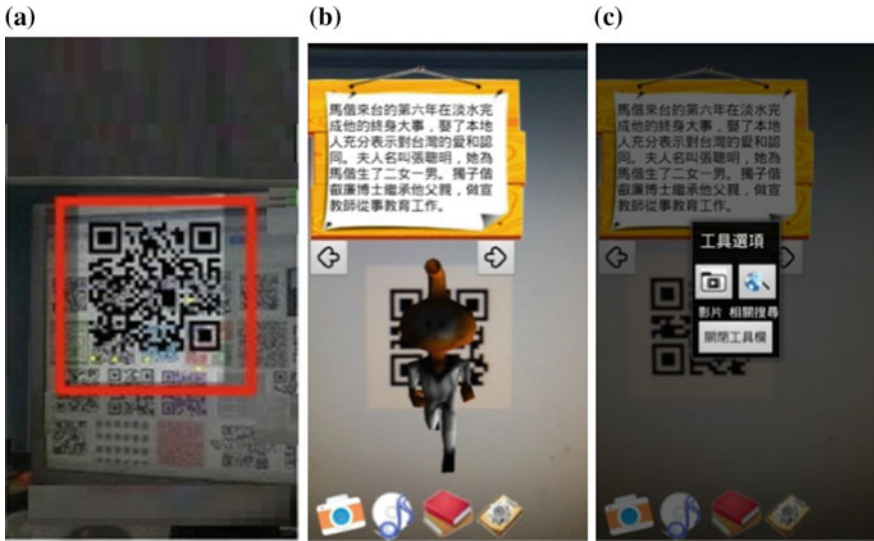


Fig. 3 Screen captures of the AR-based mobile learning system

Aligning Module to translate the information from the QR Code into registered images that can be identified and then played as 3D virtual objects on the mobile device. If the AR Materials Display Module receives AR materials from the AR Materials Download Module, relevant teaching information will become visible to students as 3D virtual objects that are superimposed on top of the real-world object that has contains the QR code.

Figure 3 shows screen captures of the AR-based mobile learning system as it is seen on a mobile device. When students scan a QR code using a mobile device, the AR-based mobile learning system will download corresponding AR materials from the AR materials remote server (see Fig. 3a). Once the information from the QR code is processed, the AR-based mobile learning system will combine course content with 3D virtual objects to present relevant course information (see Fig. 3b). If students want to learn about other related multimedia resources available for further exploration, they can open the related audio/video files through the use of pop-up buttons on the screen (see Fig. 3d).

4 Experimental Results

We invited five teachers to volunteer as formal evaluators of this study and the teachers were employed by Aletheia University in Taiwan for a number of years. Among the teacher participants, two out of five were male and the remaining teachers were female. Each teacher has prior experience with using computers and

Table 1 The attitudinal survey of teachers to the mobile augmented reality system

Question	Mean	SD
1. I believe that the mobile augmented reality system is a useful tool that supports my teaching	4.48	0.70
2. I believe that the mobile augmented reality system can better facilitate my course instruction	4.82	0.44
3. The webpage-style visual editor is an easy to operate method of creating AR materials	3.61	0.70
4. I think that the mobile augmented reality system is an easy-to-use tool	3.88	0.46
5. I feel more motivated to teach while using the mobile augmented reality system to create an AR-based learning environment	4.45	0.73
6. I wish to use the mobile augmented reality system within my formal courses	3.85	0.45
The overall average score	4.18	0.58

all possesses the basic skills for designing multimedia materials. At the beginning of the study, all teachers were requested to participate in a tutorial, which introduced our mobile augmented reality system to the participants in greater detail. When this training process was complete, all participating teachers were asked to create a simple AR enhanced course material that was presented using the AR-based mobile learning system. After completing the authoring work, the opinions and insights from participating teachers were collected using a questionnaire survey. The results of the questionnaire were measured using a 5-point Likert-scale, from which the answers provided were ranked from 1 (strongly disagree) to 5 (strongly agree).

Table 1 shows that the general mean value of consensus was 4.18 (SD = 0.58), indicating that the majority of the participating teachers were “satisfied” when using the mobile augmented reality system for course related endeavors. All mean values of the answers collected were higher than or equal to 3.50, which meant that the teachers held affirmative perceptions of our system. In fact, existing studies found that AR technology can render a learning system that is more entertaining or engaging for both teacher and student [1]. Therefore, we believe that the proposed system can be accepted as a useful tool for teachers interested in executing AR enhanced real-world learning activities.

5 Conclusions and Future Work

This study strives to support teachers in combining relevant instructional content with AR capabilities, by providing an easy to use mobile AR system that assists teachers in creating an engaging learning environment. Our development of a mobile augmented reality system to supplement real-world learning activities is unique when applied to the field of education. This advanced system provides

teachers with an AR materials remote server for writing course content and creating AR materials and the AR-based mobile learning system effectively presents customized course information to students. Moreover, a questionnaire survey was employed to measure the perceived usefulness of the proposed system from each of the participating teachers. The results of the survey showed that teachers were satisfied with the mobile augmented reality system, willing to continue using the system and strongly interested in integrating this system into their other formal courses.

In the future, we plan to use similar quantitative and qualitative methods to evaluate the effectiveness of our proposed mobile augmented reality system. All data from the statistical analyses must be collected to properly investigate the usability of the system in relation to current educational programs. The results obtained will be used to further demonstrate the effectiveness of using AR materials to enhance traditional classroom materials, and we hope to gain valuable insights from the student's perspective.

Acknowledgment This work was supported by the National Science Council of the Republic of China under Contract No. MOST 103-2221-E-156-007.

References

1. Wu H-K, Lee SW-Y, Chang H-Y, Liang J-C (2013) Current status, opportunities and challenges of augmented reality in education. *J Comput Educ* 62:41–49
2. Di Serio Á, Ibáñez MB, Kloos CD (2013) Impact of an augmented reality system on students' motivation for a visual art course. *J Comput Educ* 68:586–596
3. Platonov J, Heibel H, Meier P, Grollmann B (2006) A mobile markerless AR system for maintenance and repair. In: *Proceedings of the 5th IEEE and ACM international symposium on mixed and augmented reality*, pp 105–108, Oct 2006
4. Azuma RT (1997) A survey of augmented reality. *Presence: teleoperators and virtual environments* 6(4):355–385
5. Kose U, Koc D, Yucesoy SA (2013) An augmented reality based mobile software to support learning experiences in computer science courses. *Procedia Comput Sci* 25:370–374
6. Azuma R, Baillet Y, Behringer R, Feiner S, Julier S, MacIntyre B (2001) Recent advances in augmented reality. *IEEE Comput Gr Appl Bridges Theor Practic Comput Gr* 21(6):34–47
7. Chen C-M, Tsai Y-N (2012) Interactive augmented reality system for enhancing library instruction in elementary schools. *J Comput Educ* 59(2):638–652
8. Virvou M, Alepis E (2005) Mobile educational features in authoring tools for personalized tutoring. *J Comput Educ* 44(1):53–68
9. Tsou W, Wang W, Li H-Y (2002) How computers facilitate English foreign language learners acquire English abstract words. *J Comput Educ* 39(4):415–428
10. Klopfer E, Sheldon J (2010) Augmenting your own reality: student authoring of science-based augmented reality games. *New Dir Youth Dev* 128:85–94
11. Liu T-Y, Tan T-H, Chu Y-L (2009) Outdoor natural science learning with an RFID-supported immersive ubiquitous learning environment. *Educ Technol Soc* 12(4):161–175
12. Bajura M, Fuchs H, Ryutarou Ohbuchi M (1992) Merging virtual objects with the real world: seeing ultrasound imagery within the patient. In: *Proceedings of the 19th annual conference on computer graphics and interactive techniques*, vol 26, no 2, pp 203–210

13. Feiner S, Macintyre B, Seligmann D (1993) Knowledge-based augmented reality. *Commun ACM* 36(7):53–62 (special issue on computer augmented environments: back to the real world)
14. Ong SK, Chong JWS, Nee AYC (2010) A novel AR-based robot programming and path planning methodology. *Rob Comput-Integr Manufact* 26(3):240–249
15. Oda O, Lister LJ, White S, Feiner S (2008) Developing an augmented reality racing game. In: *Proceedings of the 2nd international conference on intelligent technologies for interactive entertainment*, Jan 2008
16. You S, Neumann U (2010) Mobile augmented reality for enhancing e-learning and e-business. In: *2010 international conference on Internet technology and applications*, pp 1–4, Aug 2010
17. Andujar JM, Mejias A, Marquez MA (2011) Augmented reality for the improvement of remote laboratories: an augmented remote laboratory. *IEEE Trans Educ* 54(3):492–500

A Simple Randomized Algorithm for Complete Target Coverage Problem in Sensor Wireless Networks

Weizhong Luo, Zhaoquan Cai and Zhi Zeng

Abstract Achieving energy efficient monitoring of targets is a critical issue in sensor networks and, thus various power efficient coverage algorithms have been proposed. These algorithms divide the sensors into monitor sets, where each monitor set is able to cover all the targets. However, even the number of monitor sets is set to 3, the considered problem has been proven to be NP-hard. In this paper, we propose a randomized and efficient coverage algorithm that produces disjoint monitor sets, i.e., monitor sets with no common sensors. The monitor sets are activated successively and only the sensor nodes from the current active set are responsible for monitoring all the target nodes, while all other nodes are in a low-energy sleep mode. Our algorithm can generate a solution with guaranteed probability $1 - \varepsilon$ ($0 < \varepsilon < 1$). Simulation results are presented to verify our approach.

Keywords Wireless sensor network · Energy efficiency · Sensor scheduling · Monitor sets

1 Introduction

Wireless sensor networks (WSNs) provide new applications for military applications, environmental monitoring, target surveillance and disaster prevention. Sensor nodes are small devices equipped with one or more sensors, one or more transceivers, storage resources and processing. The characteristics of sensor network include limited resources, a dynamic topology and, dense and large networks.

W. Luo (✉) · Z. Cai · Z. Zeng
Department of Computer Science, Huizhou University, Huizhou, China
e-mail: lwz@hzu.edu.cn

Z. Cai
e-mail: cai@hzu.edu.cn

Z. Zeng
e-mail: zengzhi@hzu.edu.cn

Depending on the application and the environment, sensor node deployment and placement can be realized by a predefined way or a random way. In hostile environments, sensor nodes are dropped from an aeroplane, resulting in a random placement, where the node density cannot be guaranteed and some areas may contain more sensors than others.

A sensor node can be in one of the following three modes of operation: active mode, sleep mode and off mode. In the active mode, a sensor can communicate with other sensors. In sleep mode, a sensor cannot monitor or transmit data. Obviously, in the sleep mode, the sensor consumes much less energy than in the active mode. In the off mode, the sensor nodes are completely turned off.

Sensors have limited battery life, and thus, a critical issue in sensor networks is power efficiency. Power saving techniques can generally be classified into four categories: (1) schedule the sensors to alternate between active and sleep mode; (2) power control by the way of adjusting the transmission range of sensors; (3) power efficient routing; (4) reducing the amount of data transmitted. Herein, we address the first method, that is, we design a mechanism that allows redundant nodes to enter sleep mode. The set of sensors are divided into subsets, called *monitor set*, where each monitor set is capable of monitoring all the targets. Sensors belonging to one monitor set are in active, while other sensors are in sleep mode.

Sensor coverage is an important issue in wireless sensor networks. This issue is centered around a fundamental question “How well do the sensor nodes observe the monitored space?” The goal is to let each target in the physical space of interest within the sensing scope of at least one sensor.

In this paper, we propose to save energy by dividing the sensors into a number of disjoint monitor sets, such that each monitor set completely covers all the targets. Herein, “disjoint” demand that each sensor is allowed to participate only in one monitor set. We propose an efficient randomized algorithm for the considered problem and, prove the guaranteed bound on the probability of generating problem solution.

The rest of the paper is organized as follows. In Sect. 2 we provide energy efficient coverage related works. Section 3 describes the system model and the considered problem. Next, in Sect. 4, we provide our randomized algorithm for the considered problem. Section 5 presents the simulation results for our algorithm, and Sect. 6 concludes our paper.

2 Related Work

Slijepcevic and Potkonjak [1] propose a centralized algorithm for the area disjoint coverage problem. They considered the field as set of targets. The author provided a heuristic algorithm with time complexity $O(n^2)$, where n is the total number of sensors. Cardei et al. [2] provide an algorithm to solve the problem by converting the considered problem into a graph problem. They state that the algorithm computing the disjoint sets from the transformed graph is $O(n^3)$. Abrams et al. [3] considered the

partial coverage where each monitor set is not necessary to cover all the target. They present a heuristic algorithm with time $O(nm|P_m|)$, where n is the number of sensors available, m the number of generated cover sets and $|P_m|$ the maximum number of fields that a sensor covers. Berman et al. [4] consider the non-disjoint coverage problem, where any given sensor may participate in more than one monitor sets. They design algorithm for the problem based on the method of Linear Programming. Their algorithm first compute a series of cover sets and then deduce the optimal lifetime for each cover set. Cheng et al. [5] provided a linear programming-based optimal algorithm and an approximation algorithm. Dimitrios Zorbas et al. [6] present a novel and efficient coverage algorithm for both disjoint and non-disjoint coverage. More results could be found in the literature [7–10].

Other line of research focuses on the design of distributed coverage algorithms [11, 12]. They use a localized and distributed scheduling scheme in order to disseminate the scheduling information rapidly. Due to the increased overhead in message exchanges between participating nodes, there are extra costs associated with this kind of algorithms. Note that, herein, we concentrate on centralized algorithm problems and, thus, distributed algorithms are outside the scope of our work.

3 System Model and Problem Definition

We consider a set of targets with known locations which need to be continuously monitored and a large number of sensors randomly deployed closed to the targets. Also we assume that the number of sensor nodes deployed in the field is greater than the optimum needed to perform the cover task. Thus, an power-efficient method consists in scheduling the sensors activity to alternate between active and sleep state. A sensor can go to the sleep mode if the sensor is not scheduled to perform the sensing task.

Let $T = \{t_1, t_2, \dots, t_m\}$ be the set of targets and $S = \{s_1, s_2, \dots, s_n\}$ the set of sensors. Assume that any target in T is covered by at least one sensor in S . Assume that any target lying within the circle defined by the circle center of a sensor and the minimum sensing range of the sensor can be monitored by this sensor. Let N_i be the set of neighbor-sensors of target node t_i where each $s_j \in N_i$ is capable of monitoring the target t_i .

Definition 1 Complete Target Coverage Problem: Given a set T of m targets, a set S of n sensors, and a set of tuples $\{(t_1, N_1), \dots, (t_m, N_m)\}$, $t_i \in T$ and an integer k , the task is to product a collection $C = \{C_1, \dots, C_k\}$ of k monitor sets. Each monitor set C_i is a subset of the sensors set S , and it must cover all targets in T . Moreover, each sensor is allowed to participate only one monitor set.

Herein, we concern only with designing the node scheduling mechanism, and do not address the issue of selecting protocol for data gathering or node synchronization.

4 Our Solution

In this section we propose a randomized approach for the Complete Target Coverage problem. Our algorithm takes as the input parameters T —the set of targets, S —the set of sensors, and I —the set of tuples $I = \{(t_1, N_1), \dots, (t_m, N_m)\}$, $t_i \in T$. The algorithm returns the set of covers C_1, \dots, C_k .

Random-CTC algorithm (T, S, I)

Input: set of targets T , set of sensors S , set of tuples I ;

Output: A partition C_1, \dots, C_k of S such that every C_i covers all the targets in T , or report “no such partition exists”.

1. for $i = 1$ to M do // M is non-negative integer
 - 1.1 for each sensor $s \in S$ do
 - choose a random number $i \in \{1, \dots, k\}$, and assign s to C_i ;
 - end for
 - 1.2 if every C_i covers all the targets in T then return (C_1, \dots, C_k) ;
 - 1.3 end for
2. return “no such partition exists”.

The above randomized algorithm assigns each sensor node to a cover chosen at random. It has few assumptions about the network, and is simple enough to facilitate implementation. Its performance is affected greatly by the loop times M . In the following, we prove that, if M is set to Nk^{km} where $N = \lceil \ln 1/\varepsilon \rceil$ (ε is a non-negative real number $0 < \varepsilon < 1$), then the Random-CTC algorithm generate a solution with probability at least $1 - \varepsilon$ if the input instance is a yes instance.

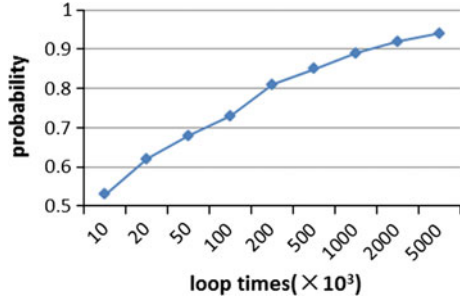
Theorem 1 *Let $M = Nk^{km}$. The Random-CTC algorithm generates a solution with probability at least $1 - \varepsilon$ if the input instance is a yes instance.*

Proof Let (C_1, \dots, C_k) be a partition of S such that every C_i covers all targets in T . Observe that, for every C_i , there exists a subset $|L| \subseteq C_i$ such that $|L| \leq m$ and L covers all the targets in T . Let $K_i \subseteq C_i$ with $|K_i| \leq m$ be some set of sensors that cover all targets in T . Let $K = K_1 \cup K_2 \cup \dots \cup K_k$. We say that the nodes in K is properly partitioned by a k -partition (S_1, \dots, S_k) of S , if for two nodes $v_1, v_2 \in K$, the following conditions hold: (1) If $v_1, v_2 \in K_i$, then $v_1, v_2 \in S_{i'}$ ($1 \leq i, i' \leq k$); (2) If $v_1 \in K_i$ and $v_2 \in K_j$ with $i \neq j$, then $v_1 \in S_{i'}$ and $v_2 \in S_{j'}$ with $i' \neq j'$ ($1 \leq i, i', j, j' \leq k$).

If we use a random process to partition the nodes in S into k disjoint subsets, then the probability that the nodes in K is properly partitioned is not less than $k!/k^{km}$.

The Random-CTC algorithm implements the above ideas. According to the above discussion, each random k -partition of S has a probability of at least $k!/k^{km}$ to get a solution. Since Step 1 loops Nk^{km} times, with a probability of at least $1 - (1 - k!/k^{km})^{Nk^{km}}$, the partition constructed by Step 1.1 is a solution. Since $\lim_{m \rightarrow +\infty} (1 + 1/m)^m = e$, $(1 - k!/k^{km})^{Nk^{km}} < e^{-N}$. Note that $N = \lceil \ln 1/\varepsilon \rceil$. It follows that $1 - (1 - k!/k^{km})^{Nk^{km}} > 1 - \varepsilon$.

Fig. 1 probability of generating solution when M changes



5 Simulation Results

In this section we evaluate the performance of Random-CTC algorithm. We simulate a network with sensor nodes and target points randomly located in a $200\text{ m} \times 200\text{ m}$ area. Assume that the sensing range is equal for all the sensor nodes. Our simulation environment consists of two families of programs. The first family of program is responsible for generating network topology, while the second is responsible for executing the desired algorithm on the generated topologies.

Initially, sensors and targets are scattered randomly in the $200\text{ m} \times 200\text{ m}$ area. Targets covered by less than k sensors are ignored, and sensors not covering any target are ignored.

In the first experiment, we evaluate how the probability of generating a solution for given yes-instances is affected by the parameter M . Figure 1 plots probability for generating solution when M changes. As we can see, the probability of generating solution increases when the parameter M increases.

Next, we evaluate the performance of running time. We assume that M is set to Nk^{km} . Simulation results demonstrate that our algorithm exhibits a rapid increase in execution time as the value of m becomes larger. The reason is that the running time of our algorithm is actually exponential with respect to m if M is set to Nk^{km} .

6 Conclusion

Wireless sensor networks are battery powered, therefore saving the network energy by a power aware node organization is highly desirable. An efficient way for power saving is to schedule the sensors to alternate between active and sleep mode. The sensor nodes are divided into disjoint monitor sets, and every monitor set completely monitors all the targets. These monitor sets are activated in turn. The problem is modeled as the Complete Target Coverage problem. We proposed an efficient randomized algorithm by using the method of partition. Furthermore, we proved the guaranteed bound on probability of generating solution. Simulation results were provided to verify our approach.

Acknowledgments This work is supported by the Ph.D. Research Startup Foundation of Huizhou University under Grant (C₅13.0211), and the National Natural Science Foundation of China under Grant (61370185, 61170193).

References

1. Slijepcevic S, Potkonjak M (2001) Power efficient organization of wireless sensor networks. In: Proceedings of international conference on communications (ICC'01), IEEE, pp 472–476
2. Cardei M, MacCallum D, Cheng MX, Min M, Jia X, Li D, Du D-Z (2002) Wireless sensor networks with energy efficient organization. *J Interconnect Netw* 3(3–4):213–229
3. Abrams Z, Goel A, Plotkin S (2004) Set k-cover algorithms for energy efficient monitoring in wireless sensor networks. In: Proceedings of third international symposium on information processing in sensor networks, ACM, pp 424–432
4. Berman P, Calinescu G, Shah C, Zelikovsky A (2004) Power efficient monitoring management in sensor networks. In: Proceedings of wireless communications and networking conference, vol 4. IEEE, pp 2329–2334
5. Cheng MX, Gong X (2011) Maximum lifetime coverage preserving scheduling algorithms in sensor networks. *J Global Optim* 51:447–462
6. Zorbas D, Glynos D, Kotzanikolaou P, Douligeris C (2010) Solving coverage problems in wireless sensor networks using cover sets. *Ad Hoc Netw* 8:400–415
7. Wang B (2011) Coverage problems in sensor networks: a survey. *ACM Comput Surveys*, pp 1–53
8. Xu X, Song M (2014) Restricted coverage in wireless networks. In: Proceedings of international conference on computer communications (INFOCOM). IEEE, pp 558–564
9. Yang Q, He S, Li J, Chen J, Sun Y (2015) Energy-efficient probabilistic area coverage in wireless sensor networks. *IEEE Trans Veh Technol* 64:367–377
10. Gorain B, Mandal PS (2013) Point and area sweep coverage in wireless sensor networks. In: Proceedings of 11th international symposium on modeling and optimization in mobile, Ad Hoc and Wireless Networks (WiOpt). IEEE, pp 140–145
11. Zhang H, Hou J (2005) Maintaining sensing coverage and connectivity in large sensor networks. *Ad Hoc Sens Wirel Netw* 1:89–123
12. Gallais A, Carle J, Simplot-Ryl D, Stojmenovic I (2008) Localized sensor area coverage with low communication overhead. *IEEE Trans Mobile Comput* 7(5):661–672

A Novel Enveloped-Form Feature Extraction Technique for Heart Murmur Classification

HaoDong Yao, BinBin Fu, MingChui Dong and Mang I. Vai

Abstract Analysis of heart sound (HS) signal is a significant approach for detecting cardiovascular diseases (CVDs). Specifically, heart murmurs are regarded as the first indication of pathological occurrences and carry important diagnostic information. With the aids of computer and artificial intelligence technologies, a lot of HS analysis methods are suggested, which principally fall into two kinds: acoustic analysis and time-frequency analysis. However, most of existing methods are associated poorly with diagnostic information in heart murmurs, which restricts severely further interpretations. Aiming to handle this bottleneck problem, a novel enveloped-form heart murmur feature extraction methods is proposed, which extracts features merely and directly from heart murmurs. Initially, the influences of fundamental HSs are eliminated and the envelopes of heart murmurs are acquired, by employing discrete wavelet transform, Shannon envelope, as well as detecting and selecting peaks of heart murmurs. Thereafter, two key features SP and TS (the ratios of start position and time span of the envelopes of heart murmurs to the length of a HS cycle respectively) are extracted directly from the envelopes of heart murmurs, which are according to that the envelopes of different heart murmurs are of diverse shapes. By applying the key features to artificial neural network for classification and CVD diagnosis, the diagnostic accuracy is up to 96 %, which significantly validates the practicability and effectiveness of the proposed method.

Keywords Automatic auscultation • Discrete wavelet transform • Enveloped-form feature extraction • Heart sound analysis • Shannon envelope

H. Yao (✉) · B. Fu · M. Dong · M.I. Vai
Faculty of Science and Technology, Department of Electrical
and Computer Engineering, University of Macau, Macau S.A.R, China
e-mail: yhd1992@126.com

B. Fu
e-mail: ariespleo51@gmail.com

M. Dong
e-mail: mcdong@umac.mo

M.I. Vai
e-mail: fstmiv@umac.mo

1 Introduction

Heart sound (HS) is the periodic signal generated by the mechanical activities of heart valves. Normal HS signals merely contain fundamental HSs (FHSs): the first and second heart sounds, S1 and S2 [1]. Besides FHSs, sounds produced by turbulent blood flows are termed as heart murmurs, which are regarded as the first indication of pathological occurrences in heart valves. Heart murmurs usually appear in abnormal HS signals and carry tremendous pathological information. With the aids of computer and artificial intelligence technologies, HS is recorded as digital signals and used for diagnosing cardiovascular diseases (CVDs).

Aiming to analyze HS automatically, plenty of HS feature extraction methods have been suggested and these methods principally fall into two kinds. The first kind of methods are oriented from the acoustic aspect. Mel-frequency cepstral coefficient (MFCC) [2] and timbre analysis [3] are two representative techniques for HS feature extraction. However, as Kamarulafizam tested and described, the analysis accuracy of MFCC is worse than that of time-frequency (TF) analysis method [4]. The timbre analysis method is recently suggested and only experimented with a small amount of HS signals [3]. The second kind of methods take TF parameters as features which comprise time and/or frequency by TF transforms [5]. Such as, in [6] the characteristic waveforms of S1 and S2 in time domain are utilized directly. And a series of TF transform coefficients of discrete wavelet transform (DWT) [7], short-time Fourier transform (STFT), continuous wavelet transform (CWT) [8] are selected. In spite of their performances, the majority of features in existing algorithms are associated poorly with diagnostic information in heart murmurs. As aforementioned that a high proportion of diagnostic information is possessed by heart murmurs, feature parameters which are stemmed from a complete cycle or FHSs but not directly from heart murmurs are with congenital limitations in subsequent analysis.

Virtually, the envelopes of heart murmurs of different CVDs exhibit diverse shapes [9–11], which show promising for further intelligent interpretation. However, there are rarely researches exploring morphological feature of HS. Although in [12] the envelope-concept is mentioned by taking frequency and temporal widths of heart murmurs as features, the following HS classification requires the assists of other feature parameters as well. Intending to dispose this bottleneck problem, this paper creatively proposes an enveloped-form heart murmur feature extraction method. The detailed methodology description and its realization are demonstrated as follows.

2 Methodology

The envelopes of heart murmurs are acquired by applying the original HS signal to the flow consisted of DWT, Shannon envelope, and detecting and selecting peaks of heart murmurs. After that, two diagnostic parameters are extracted grounded on

enveloped-form of heart murmurs, which will be applied as inputs of classifier for evaluation.

2.1 Discrete Wavelet Transform

DWT is a worldwide TF analysis method, which displays superiority in gaining good frequency resolution at low frequency band along with good time resolution at high frequency band. DWT decomposes a tested signal into a group of sub-band signals in leveled frequency ranges [13, 14]. In each level, DWT invokes two set of functions termed as wavelet functions and scaling functions, which are related with half band high pass and half band low pass filters. Therefore, decomposing the analyzed signal into different frequency bands is able to be accomplished by continuing low pass and high pass filtering. Moreover, the result of low pass filter is named as approximation coefficients, while the output of high pass filter is termed as detail coefficients. As in (1) and (2), $x[i]$ is the analyzed signal, after passing through half band low pass filter $h[i]$ and half band high pass filter $g[i]$, the outputs are obtained as:

$$Y_{low}[m] = \sum_{i=-\infty}^{\infty} x[i]h[2m - i] \quad (1)$$

$$Y_{high}[m] = \sum_{i=-\infty}^{\infty} x[i]g[2m - i] \quad (2)$$

where $Y_{low}[m]$ and $Y_{high}[m]$ are the approximation coefficients and detail coefficients respectively, m and i denote the numbers of points in signals.

According to former researches [13, 15], the frequencies of FHSs distribute below 500 Hz, while the frequencies of heart murmurs are generally greater than that of FHSs. Thereafter, 7 levels DWT is applied to decompose the analyzed HS signals into 8 sub-band signals. Then, the 6th level detail component (D6) which is in the range of 512–1024 Hz is selected out and considered as heart murmur.

2.2 Shannon Envelope

Due to the amplitude of HS signal is influenced by some factors such as sex, age, and physiology of subject, normalization to regulate the amplitude of selected D6 from -1 to 1 is implemented necessarily.

After normalization, the normalized average Shannon energy is utilized to obtain the Shannon envelope of selected sub-band signal [16]. The preliminary step is to segment sub-band signal into continuous 0.01 s windows having 0.0025 s overlaps

with adjacent windows at both ends, so that the average Shannon energy of normalized sub-band signal is achieved by:

$$E_s(k) = -\frac{1}{N} \sum_{n=1}^N S_{norm}^2(n) \log S_{norm}^2(n) \tag{3}$$

where k is the number of window ($k = 1, 2, \dots, K$), $E_s(k)$ is the average Shannon energy of k th window, N is window length, and $s_{norm}(n)$ is the normalized D6 sub-band signal.

On the basis of E_s 's, normalized average Shannon energy $P(k)$ is calculated by:

$$P(k) = \frac{E_s(k) - M_E}{S_E} \tag{4}$$

where M_E is the mean value of all E_s 's, and S_E is the standard deviation of that, separately. Lastly, Shannon envelope vector \mathbf{P} is determined according to (5).

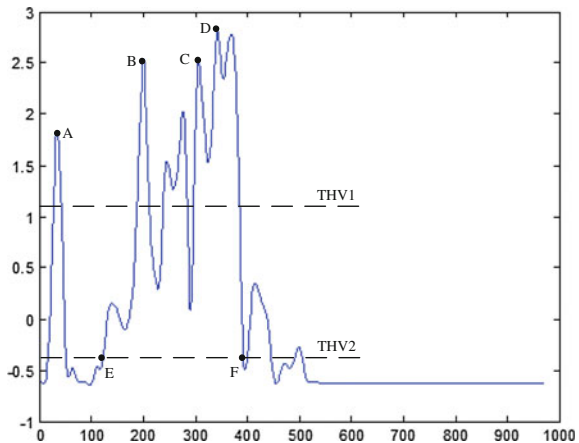
$$\mathbf{P} = [P(1), P(2), \dots, P(k)]^T \tag{5}$$

In addition, cubic spline interpolation algorithm is evoked by inserting 9 points between two abutting $P(k)$'s, so as to smooth the envelope and promote the accuracy in feature extraction.

2.3 Detecting and Selecting Peaks of Heart Murmurs

It is notable that a wave of any heart murmur is always with several peaks, as depicted in Fig. 1, peak B, C, D are in a same wave of heart murmur. Hereby, in this

Fig. 1 The envelopes of heart murmur and residual S1



research, the peak of the greatest value is recognized as the exact peak of the relevant wave and others are deemed as void peaks. Besides, although D6 sub-band signal should only contain heart murmurs in theory, a fraction of FHSs components are residual in selected D6 after DWT in some signals. Through Shannon envelope algorithm, the envelopes of surplus FHSs exist with the envelopes of heart murmurs, such as peak A in Fig. 1 is the envelope of residual S1 in an abnormal HS signal and the two peaks in Fig. 2 are the envelopes of residual FHSs in normal HS signal.

Due to those void peaks and the peaks of residual FHSs components disturb severely the feature extraction of heart murmurs, detecting and selecting exact peaks of heart murmurs should be operated before feature extraction. The workflow of such an operation is exhibited in (a), (b), (c) of Fig. 3, where $elp(n)$ represents the envelope point at index n and $Yelp(n)$ is the magnitude value of $elp(n)$, THV1 is an optimal threshold defined as the value multiplying the maximum value of envelope by 0.5, and THV2 is defined as the value multiplying the maximum value of envelope by 0.05.

In addition, the operation of eliminating the influence of residual FHSs is required. In principal, when compared to the length of one HS cycle, the index positions of S1 and S2 are always within certain percentage ranges. Moreover, compared to heart murmurs, both edges of FHSs are much steeper, which results in the both edge slopes of FHSs are much greater than that of heart murmurs as the two peaks of Fig. 2 exhibited. Hereby, the detailed steps of eliminating the peaks of residual FHSs are listed in Fig. 3d.

Since the length and maximum value of each signal envelope are different, normalizations in x -axis and y -axis are implemented individually. In x -axis, normalization is given as the indexes of exact peaks and edge points are divided by the length of a cycle, while the normalization in y -axis is the resulted value of each

Fig. 2 The envelopes of D6 component of a normal HS

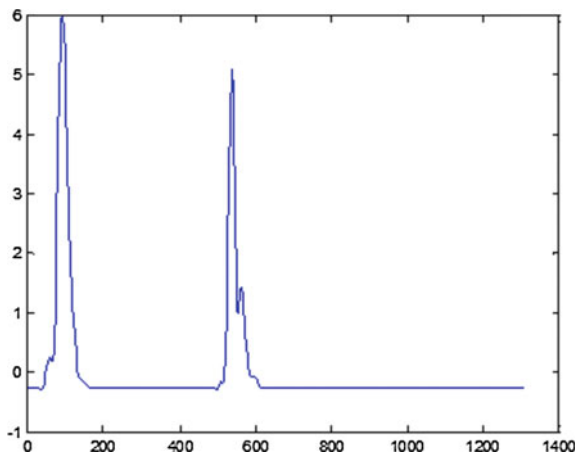
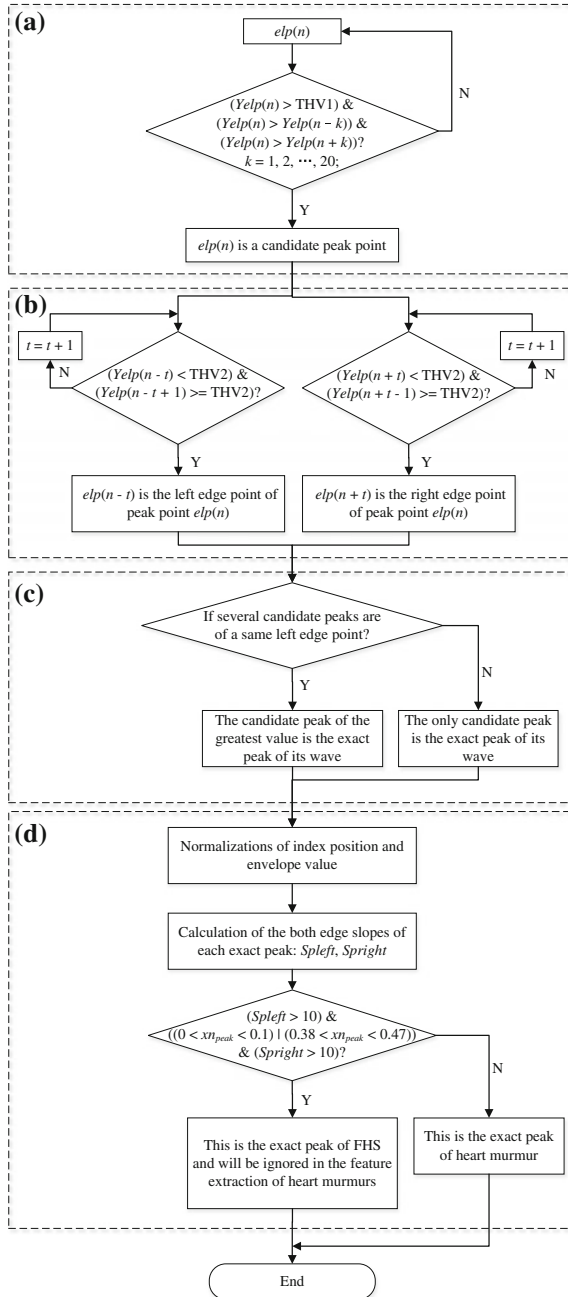


Fig. 3 Workflow of detecting and selecting peaks of heart murmurs



point compared to the greatest value of the envelope. The both edge slopes, *Spleft* and *Spright*, are defined as follows:

$$Spleft = (yn_{peak}(j) - yn_{left}(j)) / (xn_{peak}(j) - xn_{left}(j)) \quad (6)$$

$$Spright = (yn_{peak}(j) - yn_{right}(j)) / (xn_{right}(j) - xn_{peak}(j)) \quad (7)$$

where $yn_{peak}(j)$, $yn_{left}(j)$, and $yn_{right}(j)$ are the normalized envelope values of j th exact peak and corresponded left and right edge points, while $xn_{peak}(j)$, $xn_{left}(j)$, and $xn_{right}(j)$ are the normalized index positions of j th exact peak and corresponded left and right edge points.

2.4 Feature Extraction

Inspired by the previous researches [9–11], the ratios of start position and time span of the envelopes of heart murmurs at an appropriate threshold to the length of a HS cycle can be treated as effective feature parameters, which are defined as SP and TS. Through a train of experiments and optimizations, start position is gained as the first position which exceeds THV1 in the first wave of heart murmurs, and time span is acquired as the summation of each time span on THV1 of each wave of heart murmurs.

3 Experimental Results

3.1 Experimental Settings

In order to evaluate the proposed enveloped-form heart murmur feature extraction method, a set of HS signals from eGeneral Medical benchmark database and other online databases are involved [10, 17, 18]. These HSs are categorized into 5 types: normal (N), aortic regurgitation (AR), mitral stenosis (MS), tricuspid regurgitation (TR), and pulmonic stenosis (PS). There are total 60 HS records selected for HS analysis, including 7 records of each type for training and 5 records of each type for test.

Due to that the recruited HS signals are from diversified databases with various sampling frequencies, resampling all signals with unified sampling frequency of 32,768 Hz is carried out as pre-processing before running proposed method.

After feature extraction, classification using artificial neural networks (ANN) is implemented. As the typical and generally used ANN, back-propagation neural networks (BPNN) with input, hidden, and output layers [19] are served as the classifier in this research. All processes of the proposed heart murmur feature extraction and HS analysis method are implemented on MATLAB platform.

Table 1 Results of experiments

	SP	TS	Accuracy
N	0 ± 0	0 ± 0	5/5
AR	0.3778 ± 0.0423	0.3549 ± 0.0752	5/5
MS	0.5475 ± 0.0480	0.1572 ± 0.0363	4/5
TR	0.0349 ± 0.0236	0.2843 ± 0.0588	5/5
PS	0.1638 ± 0.0502	0.1610 ± 0.0162	5/5
Total			24/25

3.2 Results

By applying recruited HS signals to proposed method, two designated features SP and TS, as well as the tested accuracy are acquired. The mean value and standard derivation of extracted features are exhibited in Table 1.

Conspicuously, SP and TS of normal HS signals all equal to zero, which indicates that residual FHS peaks of normal HS signals are all successfully eliminated by the proposed method.

Furthermore, it is apparent to learn that HS signals of a same type are with pretty similar features, while signals of different types show out extremely different features. The tested result based on diagnostic features reaches a high accuracy which is up to 96 %.

In a word, the proposed enveloped-form heart murmur feature extraction method is able to discriminate various types of HS signals for carrying CVD diagnosis successfully.

4 Conclusion

A novel enveloped-form heart murmur feature extraction method is proposed for automatic HS interpretation in this paper, which extracts features merely and directly from heart murmurs. Initially, this method effectively eliminates the influence of FHSs and obtains the envelopes of heart murmurs. Thereafter, two parameters are extracted directly from the envelopes of heart murmurs, which are according to that the envelopes of different heart murmurs are of diverse shapes. By applying the features to BPNN, the experimental results markedly validate that the enveloped-form heart murmur feature extraction method is able to extract effective and sufficient features for HS analysis.

Acknowledgement This work was supported in part by the Research Committee of University of Macau under Grant No. MYRG2014-00060-FST, and in part by the Science and Technology Development Fund (FDCT) of Macau under Grant No. 016/2012/A1, respectively.

References

1. Debbal SM, Bereksi-Reguig F (2007) Time-frequency analysis of the first and the second heartbeat sounds. *Appl Math Comput* 184(2):1041–1052
2. Chauhan S, Wang P, Lim CS, Anantharaman V (2008) A computer-aided MFCC-based HMM system for automatic auscultation. *Comput Biol Med* 38:221–233
3. Wang HY, Li GP, Fu BB, Huang J, Dong MC (2014) Multidimensional feature extraction based on timbre model for heart sound analysis. *Int J Biosci Biochem Bioinf* 4:318–321
4. Kamarulafizam I, Salleh S, Najeb JM, Ariff AK, Chowdhury A (2007) Heart sound analysis using MFCC and time frequency distribution. In: *World congress on medical physics and biomedical engineering 2006*. Springer, Berlin, pp 946–949
5. Yao HD, Ma JL, Dong MC (2014) A study of heart sound analysis techniques for embedded-link e-health application. In: *International conference on intelligent system, data mining and information technology*, pp 87–91, Bangkok, Thailand
6. Jiang Z, Choi S (2006) A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope. *Expert Syst Appl* 31:286–298
7. Ölmez T, Dokur Z (2003) Classification of heart sounds using an artificial neural network. *Pattern Recogn Lett* 24:617–629
8. Debbal SM, Bereksi-Reguig F (2008) Computerized heart sounds analysis. *Comput Biol Med* 38:263–280
9. Learn the heart, <http://www.learntheheart.com/cardiology-review/heart-murmurs/>
10. Easy auscultation, <http://www.easyauscultation.com/>
11. Medscape cardiology, <http://emedicine.medscape.com/cardiology>
12. Meziani F, Debbal SM, Atbi A (2012) Analysis of phonocardiogram signals using wavelet transform. *J Med Eng Technol* 36:283–302
13. Singh J, Anand RS (2007) Computer aided analysis of phonocardiogram. *J Med Eng Technol* 31:319–323
14. The Wavelet Tutorial, http://person.hst.aau.dk/enk/ST8/wavelet_tutorial.pdf
15. Liang H, Sakari L, Iiro H (1997) A heart sound segmentation algorithm using wavelet decomposition and reconstruction. In: *19th annual international conference of the IEEE*, pp 1630–1633, Chicago
16. Choi S, Jiang Z (2008) Comparison of envelope extraction algorithms for cardiac sound signal segmentation. *Expert Syst Appl* 34:1056–1069
17. eGeneral Medical, <http://www.egeneralmedical.com/listohearmur.html>
18. Heart sounds and murmurs database, <http://depts.washington.edu/physdx/heart/tech.html>
19. DeGroff CG, Bhatikar S, Hertzberg J, Shandas R, Valdes-Cruz L, Mahajan RL (2001) Artificial neural network-based method of screening heart murmurs in children. *Circulation* 103:2711–2716

Research on Network Security Strategy Model

Anyi Lan, Bo Li, Rongsheng Huang, Xiao Zhang and Guiliang Feng

Abstract Nowadays integrated network ensures the security of information transmission by adding encrypt, simple authentication and so on, but it lacks effective means to secure the verification, authorization, confidentiality and completeness of the information, especially in the wireless network, and as a result of the openness of the transmission medium, it is particularly important to guarantee its security. This paper focuses on the modern cryptosystem to establish and realize a safe and practical integrated network security strategy model. The architecture of the model consists of three portions, namely security system, secure connection of network and security transmission of data, key management.

Keywords Integration · Network security · Network security strategy

1 Introduction

Integrated network is a system composed of multiple LAN with each LAN responsible for its own module. Data, files and instructions should be transmitted safely among each LAN, so that the command decision-making can be carried out smoothly. This paper focuses on the security of the information transmission of integrated network, adopts different safety techniques making best use of the advantages and bypassing the disadvantages, develops and realizes a network security strategy system.

A. Lan · R. Huang · X. Zhang · G. Feng (✉)
College of Information Science and Engineering,
Hebei North University, Hebei, China
e-mail: 6838710@qq.com

B. Li
College of Science, Hebei North University, Hebei, China

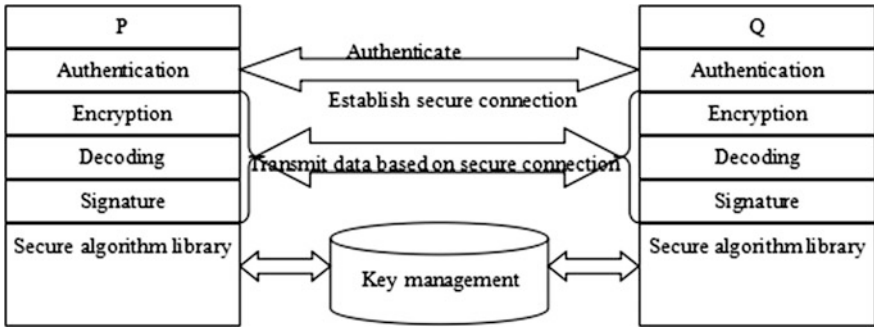


Fig. 1 Architecture of the model

2 Architecture of the Model

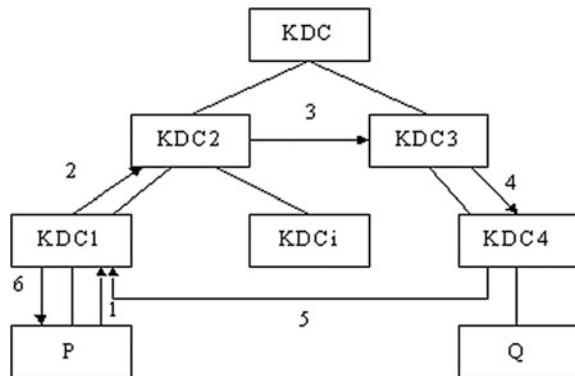
The model consists of three portions: security system, secure connection of network and security transmission of data, security key management. The architecture of the model is shown in Fig. 1.

3 Workflow of the Model

3.1 Acquisition of Public Key for Both Sides

If user P wants to correspond with user Q, User P needs to know the public key of User Q first, and the model acquires the public key from the distributed KDC setting which produces security key. The logical structure of the KDC is shown in Fig. 2.

Fig. 2 Acquisition of public key



P sends message to local KDC to apply for public key first, and then KDC checks the address of Q in its own workstation address table, if found, we say P and Q are in the same user group in local KDC management; at the same time, KDC sends replied message about Q's public key to applicant P, also sends P's public key to Q. If KDC doesn't find Q's address in its own workstation address table, it shows that Q belongs to other user group.

Now, the procedure that P gains Q's public key can be described as follows:

1. P applies KDC1 for non-local communication.
2. Checks with KDC1 whether the public key application is related to the KDCi which has the same superior as the local KDC, namely whether Q is below to the vis-a-vis KDCi; if unrelated, sends application to upper KDC (KDC2).
3. If KDC2 receives the application of public key transmission from the subordinate, it will handle as step 2, finding KDC3 and transmitting application to it.
4. When KDC3 receives the application, it will check whether Q is direct managed, if not, it will find relevant subordinate KDC4.
5. When KDC4 receives the application, it will handle as step 4. If Q is direct managed, it will send the public key message to KDC1, at the same time, KDC4 sends the public key of P which is in the received public key application message to Q.
6. KDC1 sends Q's public key message to user P, so P and Q can intercommunicate with each other.

In the above processes of acquiring public key, public key application transmits among each KDC until it finds the local KDC which manages the communication object directly. After both communication sides acquire each other's public key, they can establish communication connection and transmit data.

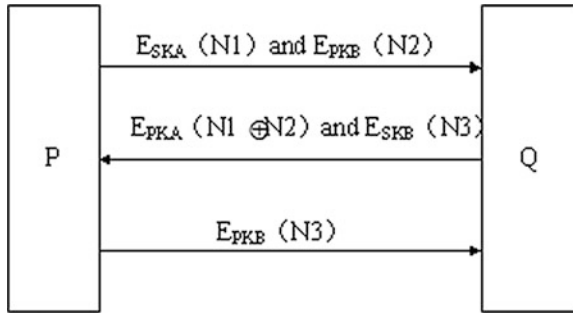
3.2 Establishment of the Secure Connection

In order to identify both side's identification, initiator P and recipient Q need to adopt public key infrastructure to transmit credentials and complete authentication process of three-time handshake. The precondition of adopting public key infrastructure is both sides have acquired public key of each other, and according to the last step, they complete the transmission within both public keys. The procedure of the three-time handshake between initiator P and recipient Q is show in Fig. 3.

Firstly, initiator P produces random number N1 and N2, encrypts N1 using its private key SKA and gets $ESKA(N1)$, then encrypts N2 using Q's public key PBK and gets $EPKB(N2)$, then sends $ESKA(N1)$ and $EPKB(N2)$.

Secondly, after recipient Q has received $ESKA(N1)$ and $EPKB(N2)$, it decodes $ESKA(N1)$ to $DPKA(ESKA(N1)) = N1$ using the public key of P PKA, decodes $EPKB(N2)$ to $DSKB(EPKB(N2)) = N2$ using its own private key SKB, produces

Fig. 3 Establishment of the secure connection



random number $N3$, using the public key of P, PKA encrypts $(N1 \oplus N2)$ to $EPKA(N1 \oplus N2)$, then using its own private key SKB encrypts $N3$ to $ESKB(N3)$, sends $EPKA(N1 \oplus N2)$ and $ESKB(N3)$.

Finally, P decodes $EPKA(N1 \oplus N2)$ using private key SKA, then decodes $ESKB(N3)$ to $N3$ using the public key of Q PKS, checks whether the data of decoded $EPKA(N1 \oplus N2)$ is $(N1 \oplus N2)$, if it is, using the public key of Q encrypts $N3$ to $EPKB(N3)$ and sends $EPKB(N3)$; if not, there's something wrong with Q or it isn't the one that P wants to communicate with, then reports error and disconnect. Q decodes $EPKB(N3)$, checks whether the decoded data is $N3$; if it is, we say that the identification of P is correct and the connection is successful, or disconnects and releases the socket.

3.3 Transmission of Data

The purpose of acquisition of each other's public key and establishment of secure connection is to transmit data in network securely. Data transmission is based on the TCP which is reliable, including the functions of encryption, signature, decoding, verification of signature and so on, guarantees the confidentiality, integrity, authenticity and non-repudiation of achieving users' action. The procedure is as follows:

Firstly, using MD5, P calculates the 128-bit digital-fingerprint D of the data needs to be sent, and adds its identification and timestamp information T , in case of the replay attachment. Using its private key SKA encrypts them and forms the signature information $S = ESKA(D + T)$. Then produces the random conversational security key K , and using K encrypts data M and signature information S to form ciphertext $C = EK(M + S)$, then using Q's public key PKB encrypts conversational security key K and gets $CK = EPKB(K)$, then encapsulates the ciphertext C and encrypted conversational security key CK into a data envelope, and sends the data envelope to Q.

After Q has received the data envelope, unpacks it in the adverse sequence. The concrete process is as follows: Using its own private key SBK decodes the cryptographic conversational security key CK, namely $DSBK(CK) = DSBK(EPKB(K)) = K$; Using the restored conversational security key K decodes the received ciphertext C, the procedure is $DK(C) = DK(EK(M + S)) = M + S$, at the same time, using P's public key SPK decodes signature information, $DSPK(S) = DSPK(ESKA(D + T)) = D + T$, and gets the digital fingerprint D of the original data. On the other hand, using MDS recalculates the digital fingerprint D1 of the received data M, and judges whether the integrity of the data has been damaged via whether the two digital fingerprint D and D1 equals with each other and using the information T identifies the other's identification and prevents replay attack.

In this process of data transmission, the combination of single key and two-key system not only guarantees the security but also improves the efficiency of the encryption algorithm. Every time encrypting the data, it produces different conversational security key randomly which is transmitted with the data. This can avoid the handshake protocol of exchanging security key, and achieve the effect of one key a time. At the process of all data transmission, data is transmitted at the form of encryption to avoid data leakage, while signature the digital fingerprint of the data guarantees the integrity of data and improves the reliability.

4 Realization of the Model

According to the architecture of the model and the function to be accomplished, the realization of the model consists of three portions: security system, data transmission and key management.

4.1 *Main Algorithm Principle of the Model's Security System*

The security system of this model is a complexed system, it mainly includes four password elements, namely two-key password, single key sectionalization password, one-way hash function and a random generation algorithm. In the process of model implementation, algorithm of public key adopts RSA and single key adopts the improved algorithm of DES—triple DES, their working modes all adopt CBC patten. Hash algorithm adopts MD5 (also retaining DES, MD2 and MD4, which can be adopted when there's no high request about data security, but high data processing rate). These algorithms are all included in a dynamic linking library Crypto.dll, forming a secure algorithm library.

4.2 *Realization of the Data Transmission*

Based on Windows Platform, this model establishes programming interface between Windows environment and network by adopting Windows Socket. In VC++ environment, realisation of data transmission mainly relies on CSocket in MFC. Establishes sending socket in client-side and establishes listening and receiving socket in server-side. Of course, sending and receiving here is concerned with ciphertext data transmitted in network. Actually, socket in client-side and server-side can not only send but also receive data.

4.3 *Realization of Key Management*

Every KDC maintains the KDC address table to offer KDC addressing. The local workstation also maintains user's address table and public key table. The user's address table is used to deposit the local user's address table which is registered. User's public key table is used to deposit the public key of local users, forming the user public key database for query. When maintains the KDC address table, the administrator can take the KDC address table into the system in the initial installation of the system, using a fixed routing scheme. User's address table and public key table can be dynamically maintained by KDC after receiving the corresponding message.

Security key is the variable part of the encryption algorithm, its security depends on the protection of security key rather than the protection of algorithm or hardware themselves. In a symmetric encryption algorithm, the importance of protecting the privacy of security key is obvious. In the security system using of public key algorithm, in addition to protect the privacy of security key, ensuring the public key is not replaced by a third party was the key problem to ensure the security of system.

Acknowledgement Supported by Hebei North University (Q2013002) and Zhangjiakou Technology Bureau (13110038I-12).

Investigating on Radioactivity of LBE and Pb in ADS Spallation Target

Yaling Zhang, Xuesong Yan, Xunchao Zhang, Jianqi Chen,
Qingguo Zhou and Lei Yang

Abstract Using the Fluka Monte Carlo method, we argue the activity of lead-bismuth eutectic (LBE) and lead (Pb) spallation targets in Accelerator Driven Subcritical System (ADS). Results reveal that the radioactivity of Pb target is lower than that of LBE target when they have the same beam energy and target diameter, the accumulation activity of two target is enhanced as the increase of proton energy and target diameter. In addition, we also discuss the respective contribution of prominent radionuclides of LBE target under various cooling time, it is specified that the used LBE target can be seen as a permanent radioactive waste. For the present works, we expect that it can offer some valuable hints for the future experiment and the assessment of safety hazards of nuclear facilities.

Keywords ADS · Spallation products · Induced activity · FLUKA

Y. Zhang · X. Yan · X. Zhang · J. Chen · L. Yang (✉)
Institute of Modern Physics, Chinese Academy of Science, 509 Nanchang Road,
Lanzhou 730000, Gansu, China
e-mail: lyang@impcas.ac.cn

Y. Zhang
e-mail: zhangyl@impcas.ac.cn

X. Yan
e-mail: xuesongy@impcas.ac.cn

X. Zhang
e-mail: zhangxunchao@impcas.ac.cn

J. Chen
e-mail: chenjianqi@impcas.ac.cn

J. Chen
University of Chinese Academy of Science, No. 19A Yuquan Road,
Beijing 100049, China

Q. Zhou
Lanzhou University, No. 222 Tianshuinan Road, Lanzhou 730000, Gansu, China
e-mail: kinggo@gmail.com

1 Introduction

Accelerator Driven Subcritical Systems (ADS) are a new option of nuclear energy due to their good cost efficiency, environmental benefits, security benefits, etc. [1, 2]. In such a system, an important component is the spallation target, which is the interface between the two principal elements, the high-power accelerator and the subcritical core. Choice of target material and design of target are dependent strongly on several factors like high neutron yield, ease of cooling, low chemical and radio-toxicity production, low risk of fire hazard in operating system, relatively low running cost of the system [3–6]. Presently, lead-bismuth eutectic (LBE) and lead (Pb) are widely used as a spallation target material for neutron production due to its several merits, such as good liquid coolant, high neutron yield, lower thermal neutron capture cross-section, high boiling point (of Pb 1737 °C, LBE 1670 °C) and low vapor pressure (parts per million of mercury in the operating temperature respectively) [7, 8]. These advantages mentioned above will bring great convenience of actual operation in ADS. Furthermore, the higher operating temperature of liquid metal target, which associates with the melting point of target material, is avoided in realistic applications. For the Pb and LBE targets, there is a melting point up to 327 °C for Pb and up to 123.5 °C for LBE, respectively [7], indicating that the Pb target has much higher operation temperature than LBE one. Hence, liquid Pb target requires special structural material that can endure higher temperature than LBE target. From the requirement point of view for target container, LBE has more advantages than Pb. Unfortunately, the LBE has its flaws as target material, the amount of polonium production in the LBE arouses a particular concern because of its high radiotoxicity and the resulting handling problems. Noble gases and the gaseous phase of ^{210}Po are also known to occur, and the alpha decay $^{210\text{m}}\text{Bi}$ are formed as a result of the reaction $^{209}\text{Bi}(n, \gamma)^{210\text{m}}\text{Bi}$ in LBE spallation target [9, 10].

As we all know, nuclear wastes are transmuted in ADS, meanwhile radioactive spallation residues are produced constantly through the primary interactions and secondary particles [11, 12]. Up to now, there are many experimental studies on the production of radionuclides [13–18], however, because of safety restrictions on the dose rate, targets were cooled for dozens of hours before the first measurement. So, it will lost information on the production of short-lived radionuclides, and experimental measurement could not identify $^{208-210}\text{Po}$ radioisotopes due to the absence of their suitable characteristic photo peaks. Therefore, it is necessary to study the nuclide production and discuss the respective contribution of prominent radionuclides in spallation target based on nuclear physics calculations, the aim is to give some valuable hints for the future experiment and the assessment of safety hazards of nuclear facilities, including also options for intermediate and final disposal of the target material. In present work, we have not only calculated the induced activity in an ADS employing LBE and Pb targets at beam energies of 0.25, 0.5 and 1.0 GeV, but also surveyed the effects of LBE and Pb targets geometry on the generation of induced activity. Finally, we discuss the residual activity of long-life spallation

products in LBE target after one year irradiation with proton beam current of 1.0 mA, beam energy of 1.0 GeV, and the cooling time up to 10^5 years.

2 Method of Calculation

In this manuscript, all calculations were performed with the Fluka Monte Carlo code [19]. The average proton current is 1.0 mA for 1 year of irradiation. Except for the discussion of the effects in target geometry, the rest of calculations always use a 10.2 cm diameter and 60.0 cm length spallation target.

2.1 Fluka

Fluka is a general purpose tool for calculations of particle transport and interactions with matter, covering an extended range of applications spanning from proton and electron accelerator shielding to target design, calorimetry, activation, dosimetry, detector design, Accelerator Driven Systems, cosmic rays, neutrino physics, radiotherapy etc. It can simulate with high accuracy the interaction and propagation in matter of about 60 different particles, including photons, electrons, neutrinos, muons, hadrons and all the corresponding antiparticles. The program can also transport polarised photons and optical photons. Time evolution and tracking of emitted radiation from unstable residual nuclei can be performed on line. Fluka can handle even very complex geometries, using an improved version of the well-known Combinatorial Geometry (CG) package. Various visualization and debugging tools are also available.

2.2 Induced Activity Calculation

In an ADS system, we always use thick and large mass target to ensure fully stop the projectile beam. In such a thick target, an isotope is produced by the interaction of the incident particle at different degrading energies from incident energy down to reaction threshold. The total yield of isotope is obtained by summing over the yield of the isotope over this entire energy range. The activity of a radioisotope 'i' is defined by [3]:

$$A_i = \sum_j N \sigma_i(E_j) \phi_j(E_j) (1 - e^{-\lambda t}) \quad (1)$$

where A_i represents activity of radioisotope 'i', N is the number of target atoms available for reaction, $\sigma_i(E_j)$ is cross-section for production of radioisotope 'i' at

projectile energy E_j , $\phi_j(E_j)$ is projectile flux at energy E_j , λ is decay constants of radioisotope 'i' and t stands for irradiation time.

3 Results and Discussion

Many types of nuclides are produced in a spallation target due to spallation, fission, (n, gamma), (n, xn) and other nuclear reactions. Which can be identified in three classes, one is represented by sharp peak of the light products dominated by tritium and helium. The second comprises a wide list of intermediate products resulted from either fission and evaporation, and the third is formed by extended range of nuclides sharply peaked at the mass numbers of initial target nuclei [20]. In the following description, we only consider some of the higher radiotoxic isotopes.

3.1 Induced Activity in Spallation Target

The generation of induced activity in an ADS employing LBE and Pb target has been studied. In Table 1, we make a comparison on the radioactivity of the LBE and pure Pb target after one year irradiation at beam energies of 0.25, 0.5 and 1.0 GeV. It can be seen distinctly that activity of the LBE target is higher than that of the Pb target at the same beam energy. Especially, the isotopes of polonium and bismuth are about two orders of magnitude larger than pure Pb target.

The reason results in the formation of ^{210m}Bi and ^{208}Bi are the reaction ^{209}Bi (n, gamma) $^{210m}\text{Bi}/^{210}\text{Bi}$ and ^{209}Bi (n, 2n) ^{208}Bi , respectively. While ^{210}Bi decay through beta emission to form ^{210}Po , ^{210}Po is the most important isotope of polonium that needs to be quantified from the radio-toxicity point of view [11]. ^{205}Pb formed by reaction of ^{204}Pb with ^{204}Pb (n, gamma) ^{205}Pb plays a very important contribution to long-lived component of the activity of pure Pb target, also for the LBE target. For pure Pb target, there is no ^{210m}Bi production. We also observed that the accumulation activity of spallation products decreases remarkably with the reduction of proton energy, especially these nuclide with mass number around 150, including ^{150}Gd , ^{148}Gd and ^{154}Dy , among them, the nuclide ^{148}Gd attends widely attention due to its long-lived as well as alpha emitter, and releases significant amount of dose in case of inhalation [20]. The contribution of alpha emitting rare earth element will not be considered when beam energy is 0.25 GeV. This results are agree reasonably well with the Artisyuk's results [20]. Therefore, pure Pb target is better than LBE target in terms of radioactivity and radiotoxicity.

In Table 2, we make a comparison on the radioactivity of the LBE and Pb target with different diameter after one year irradiation at beam energy of 1.0 GeV. It can be seen that the activity of spallation products improves significantly as increasing

Table 1 Comparison on the radioactivity of the LBE and Pb target in different beam energy (Bq)

Nuclide	LBE			Pb		
	0.25 GeV	0.5 GeV	1.0 GeV	0.25 GeV	0.5 GeV	1.0 GeV
²¹⁰ Po	1.91E + 12	6.40E + 12	1.50E + 13	0	4.19E + 09	3.46E + 10
²⁰⁹ Po	3.50E + 10	4.58E + 10	4.27E + 10	0	1.69E + 07	1.78E + 08
²⁰⁸ Po	5.57E + 12	6.74E + 12	7.84E + 12	2.66E + 08	7.97E + 08	3.72E + 09
^{210m} Bi	5.23E + 05	1.75E + 06	4.09E + 06	0	0	0
²⁰⁸ Bi	3.99E + 08	1.31E + 09	3.49E + 09	9.49E + 06	1.16E + 07	1.24E + 07
²⁰⁷ Bi	4.03E + 12	9.71E + 12	2.10E + 13	5.93E + 11	7.34E + 11	8.70E + 11
²⁰⁶ Bi	1.68E + 14	3.49E + 14	6.62E + 14	4.51E + 13	5.45E + 13	6.29E + 13
²⁰⁵ Bi	1.74E + 14	3.30E + 14	5.73E + 14	7.29E + 13	8.73E + 13	9.60E + 13
²⁰⁵ Pb	1.26E + 07	2.96E + 07	6.06E + 07	1.40E + 07	3.60E + 07	8.22E + 07
²⁰² Pb	2.79E + 09	5.65E + 09	8.82E + 09	3.02E + 09	6.08E + 09	1.05E + 10
²⁰⁶ Tl	7.69E + 12	3.05E + 13	6.95E + 13	1.76E + 13	6.91E + 13	1.59E + 14
²⁰⁴ Tl	1.73E + 12	6.61E + 12	1.45E + 13	3.80E + 12	1.42E + 13	3.11E + 13
²⁰² Tl	1.61E + 13	5.64E + 13	1.16E + 14	3.27E + 13	1.11E + 14	2.26E + 14
¹⁹⁴ Hg	4.35E + 10	2.12E + 11	3.53E + 11	5.27E + 10	2.47E + 11	4.05E + 11
¹⁹⁵ Au	2.87E + 13	1.28E + 14	2.12E + 14	3.99E + 13	1.60E + 14	2.60E + 14
¹⁹⁴ Au	3.77E + 11	3.42E + 12	1.03E + 13	6.67E + 11	5.98E + 12	1.80E + 13
¹⁹³ Pt	4.90E + 11	2.24E + 12	3.93E + 12	7.02E + 11	2.80E + 12	4.67E + 12
¹⁵⁰ Gd	456.3	1.19E + 04	4.28E + 05	483.4	1.02E + 04	4.65E + 05
¹⁴⁸ Gd	0	1.62E + 08	1.14E + 10	0	1.04E + 08	1.25E + 10
¹⁵⁴ Dy	288.4	3751	3.81E + 05	0	1444	4.22E + 05
⁹⁹ Tc	1.60E + 07	5.78E + 07	1.09E + 08	1.10E + 07	4.11E + 07	8.20E + 07
³ H	1.32E + 11	1.41E + 12	1.29E + 13	1.30E + 11	1.41E + 12	1.29E + 13

target diameter. In other word, spallation products are highly dependent on the mass of the target material. Only for the activity, lower beam energy and smaller target diameter are expected, but this effect will confront with the overall decrease in neutron production. Whereas the main application of spallation target in ADS is as a source for generation of neutron, thus the balance must be taken into account between the neutron yield and the cumulative activity.

3.2 Residual Activity in LBE Target

Many types of nuclides, including alpha-active polonium isotopes (²¹⁰, ²⁰⁹, ²⁰⁸, ²⁰⁶Po) and ^{210m}Bi as well as hard gamma emitting bismuth isotopes (²¹⁰, ²⁰⁸, ²⁰⁷, ²⁰⁶, ²⁰⁵Bi), are produced in a LBE target due to spallation, (n, gamma), (n, xn), (p, xn) and other nuclear reactions. These alpha and hard gamma emitter are the most hazardous due to their relatively long-lived radiowaste. In Figs. 1 and 2, we display the contribution of the isotopes of bismuth and polonium to total activity after

Table 2 Comparison on the radioactivity of the LBE and Pb target with different diameter (Bq)

Nuclide	LBE			Pb		
	10.2 cm	20 cm	30 cm	10.2 cm	20 cm	30 cm
²¹⁰ Po	1.51E + 13	4.75E + 13	9.56E + 13	3.46E + 10	3.72E + 10	3.77E + 10
²⁰⁹ Po	4.23E + 10	5.31E + 10	5.63E + 10	1.78E + 08	1.56E + 08	1.48E + 08
²⁰⁸ Po	7.90E + 12	9.74E + 12	1.04E + 13	3.72E + 09	4.25E + 09	3.72E + 09
^{210m} Bi	4.11E + 06	1.30E + 07	2.62E + 07	0	0	0
²⁰⁸ Bi	3.50E + 09	6.28E + 09	8.28E + 09	1.24E + 07	1.54E + 07	1.62E + 07
²⁰⁷ Bi	2.10E + 13	3.60E + 13	4.66E + 13	8.70E + 11	1.05E + 12	1.12E + 12
²⁰⁶ Bi	6.63E + 14	1.09E + 15	1.37E + 15	6.29E + 13	7.67E + 13	8.14E + 13
²⁰⁵ Bi	5.75E + 14	9.18E + 14	1.15E + 15	9.60E + 13	1.18E + 14	1.25E + 14
²⁰⁵ Pb	6.07E + 07	1.01E + 08	1.30E + 08	8.22E + 07	1.40E + 08	1.80E + 08
²⁰² Pb	8.85E + 09	1.34E + 10	1.62E + 10	1.05E + 10	1.64E + 10	2.03E + 10
²⁰⁶ Tl	6.89E + 13	1.03E + 14	1.27E + 14	1.59E + 14	2.38E + 14	2.88E + 14
²⁰⁴ Tl	1.45E + 13	2.13E + 13	2.57E + 13	3.11E + 13	4.62E + 13	5.59E + 13
²⁰² Tl	1.17E + 14	1.69E + 14	2.02E + 14	2.26E + 14	3.31E + 14	3.96E + 14
¹⁹⁴ Hg	3.54E + 11	4.16E + 11	4.44E + 11	4.05E + 11	4.83E + 11	5.18E + 11
¹⁹⁵ Au	2.12E + 14	2.56E + 14	2.75E + 14	2.60E + 14	3.19E + 14	3.45E + 14
¹⁹⁴ Au	1.03E + 13	1.19E + 13	1.25E + 13	1.80E + 13	2.06E + 13	2.15E + 13
¹⁹³ Pt	3.94E + 12	4.69E + 12	5.04E + 12	4.67E + 12	5.64E + 12	6.10E + 12
¹⁵⁰ Gd	4.16E + 05	4.14E + 05	4.27E + 05	4.65E + 05	4.73E + 05	4.62E + 05
¹⁴⁸ Gd	1.07E + 10	1.12E + 10	1.11E + 10	1.25E + 10	1.23E + 10	1.25E + 10
¹⁵⁴ Dy	3.81E + 05	3.79E + 05	3.62E + 05	4.22E + 05	4.43E + 05	4.39E + 05
⁹⁹ Tc	1.09E + 08	1.24E + 08	1.30E + 08	8.20E + 07	9.13E + 07	9.51E + 07
³ H	1.29E + 13	1.34E + 13	1.35E + 13	1.29E + 13	1.33E + 13	1.35E + 13

1 year irradiation at a beam current of 1 mA, beam energy of 1 GeV, and a cooling time up to 10^5 years. It is observed in Fig. 1 that during the first month of cooling ²⁰⁵Bi ($T_{1/2} = 15.31$ days), ²⁰⁶Bi ($T_{1/2} = 6.24$ days) and ²¹⁰Bi ($T_{1/2} = 5.012$ days) give the prominent contribution to total bismuth activity, and then ²⁰⁷Bi ($T_{1/2} = 32.9$ years) plays a major part in the activity. After 300 years of cooling, the activity of ²⁰⁷Bi decreases to $3.16E + 10$ Bq, and then ²⁰⁸Bi ($T_{1/2} = 3.68E + 05$ years) and ^{210m}Bi ($T_{1/2} = 3.04E + 06$ years) dominate the total inventory. Therefore, we conclude that for bismuth isotopes long-lived radionuclides ²⁰⁷Bi, ²⁰⁸Bi, and ^{210m}Bi are the most hazardous. Particularly, the ^{210m}Bi decays by alpha particle emission that needs to be quantified from the radio-toxicity point of view.

In Fig. 2, we show the activity of polonium isotopes with a half live greater than one day, such as ²⁰⁶Po (lifetime of 8.8 days), ²⁰⁸Po (lifetime of 2.9 years), ²⁰⁹Po (lifetime of 102 years) and ²¹⁰Po (lifetime of 138 days). All of these nuclides decay almost by alpha emission except that ²⁰⁶Po has only about 5.45 % probability of alpha emission. In addition, we can see in Fig. 2 that in the initial days of decay, ²⁰⁶Po dominates an essential activity of polonium isotopes. During the period of

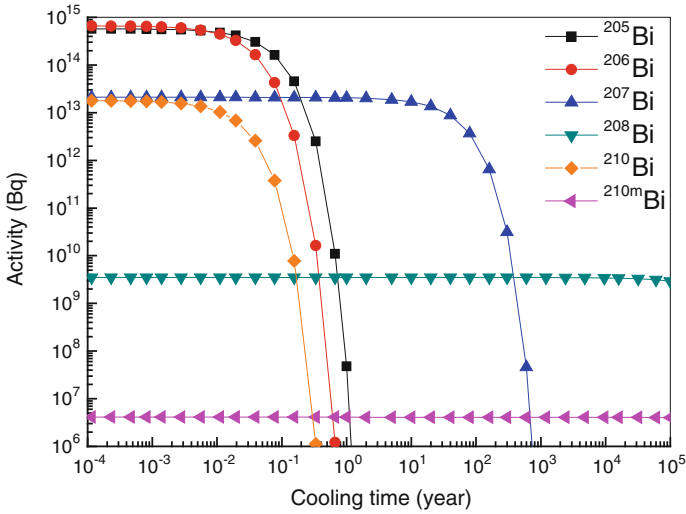


Fig. 1 Residual activity of the isotopes of bismuth produced in LBE target

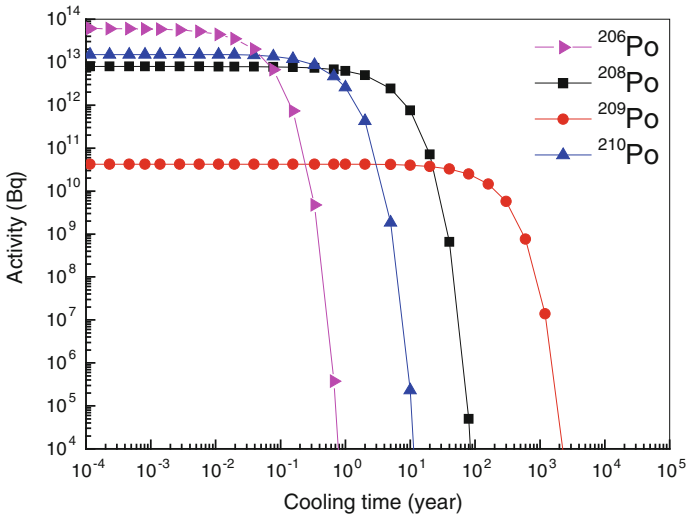


Fig. 2 Residual activity of the isotopes of polonium produced in LBE target

cooling from a month to about five years, ²¹⁰Po as the major radionuclide, which needs to be quantified from the radiotoxicity point of view because of the gaseous phase. After that ²⁰⁸Po and ²⁰⁹Po become the predominant radionuclide in the total polonium activity.

Figure 3 presents the total activity generated in the LBE target and a few prominent radionuclides, except for the isotopes of bismuth and polonium, the

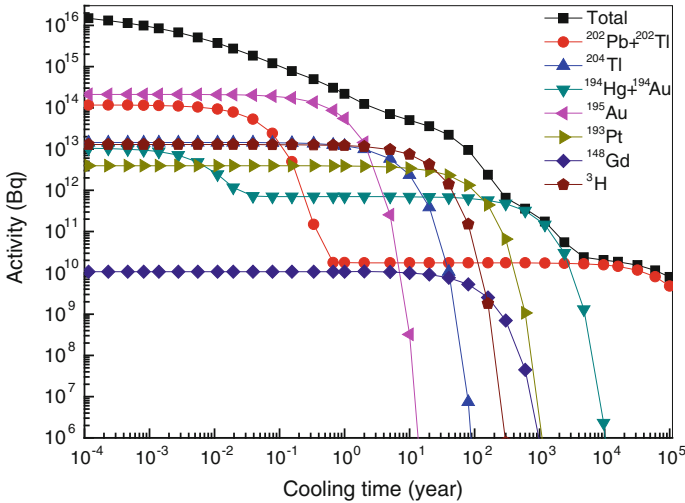


Fig. 3 Residual activity of the other nuclide produced in LBE target

contribution of these isotopes to the total activity is larger than 5 % at various cooling times, the activity of long-lived radionuclide ^{148}Gd which decay via alpha particle emission is also given to make a comparison with other isotopes. In addition, as the major radionuclides ^{195}Au , there is a cooling time up to 2 years and then the activity decreases quickly as increasing time, the activity of ^{195}Au devotes to the total activity up to 29 % after one year of cooling, tritium and ^{204}Tl can be retained during 20 years of cooling. After about 300 years of cooling, the nuclides of ^{194}Hg , ^{194}Au , ^{202}Pb and ^{202}Tl dominate in the total inventory, the contribution of ^{194}Hg and ^{194}Au to total activity approaches 90 % when cooling time is up to 600 years. After that, ^{202}Pb together with ^{202}Tl make the main contribution to the total activity. These results are consistent with research by Lemaire [21]. Therefore, we can conclude that LBE as spallation target material, the necessary measures must be taken to deal with used target as radioactive waste.

4 Summary and Conclusions

The Fluka Monte Carlo method is applied to explore the activity of LBE and Pb spallation target, it is found that the radioactivity of Pb target is significantly lower than that of LBE target under the same beam energy and target diameter, and there is no $^{210\text{m}}\text{Bi}$ production in pure Pb target. The accumulation activity of spallation products decreases remarkably with the reduction of proton energy, especially these nuclides with mass number around 150, when beam energy is 0.25 GeV, the contribution of alpha emitting rare earth element will not be considered. In addition, the respective contribution of prominent radionuclides of LBE target are surveyed

at various cooling time, the results reveal that the alpha-active polonium isotopes ($^{210}, ^{209}, ^{208}, ^{206}\text{Po}$), $^{210\text{m}}\text{Bi}$ and hard gamma emitting bismuth isotopes ($^{210}, ^{208}, ^{207}, ^{206}, ^{205}\text{Bi}$) are the most important radionuclides from the radiotoxicity and long term storage point of view. Moreover, ^{195}Au , ^{204}Tl , ^3H , ^{193}Pt , ^{194}Hg , ^{194}Au , ^{202}Pb and ^{202}Tl also have the important contribution to total activity in various cooling time, and then LBE as a target material must be classified as a radioactive waste permanently, for the Pb target, the situation is not so highlighted. Therefore, we hope that our results can give some valuable hints for the choice of target material and the future experiment.

Acknowledgements This work is supported by the National Magnetic Confinement Fusion Science Program of China (Grant No. 2014GB104002), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA03030100) and CAS 125 Informatization Project (No. XXH12503-02-03-2).

References

1. Rubbia C, Aleixandre J, Andriamonje S (2001) A European roadmap for developing accelerator driven systems (ADS) for nuclear waste incineration, ENEA Report, 88
2. Gokhale PA, Deokattey S, Kumar V (2006) Accelerator driven systems (ADS) for energy production and waste transmutation: international trends in R&D. *Prog Nucl Energy* 48:91
3. Nandy M, Lahiri C (2012) Radioactivity generation in Pb target by protons—a comparative study from MeV to GeV. *Indian J Pure Appl Phys* 50:761
4. Zhao Z, Luo Z, Xu Y, Ding D (2003) Study on ADS Pb (Pb/Bi) spallation target. Power reactors and sub-critical blanket systems with lead and lead-bismuth as coolant and/or target material, IAEA-TECDOC-1348, 211
5. Bauer GS (2010) Overview on spallation target design concepts and related materials issues. *J Nucl Mater* 398:19–27
6. Xu CC, Ye YL, Chen T, Ying J, Ma JG, Liu HT (1999) Study on the parameters of the lead spallation target. *High Energy Phys Nuc* 23:402–408
7. Orlov VV, Leonov VN, Sila-novitski AG, Smirnov VS, Filin AI, Tsikunov VS (2003) Lead coolant as a natural safety component. Power reactors and sub-critical blanket systems with lead and lead-bismuth as coolant and/or target material, IAEA-TECDOC-1348, 89
8. Orlov YI, Martynov PN, Gulevsky VA (2003) Issues of lead coolant technology. Power reactors and sub-critical blanket systems with lead and lead-bismuth as coolant and/or target material, IAEA-TECDOC-1348, 95
9. Yefimov E, Gromov B, Leonchuk M, Orlov Y, Pankratov D (1999) Problems of molten lead-bismuth target development for accelerator-driven systems. In: Proceedings of the 3rd international conference on accelerator driven transmutation technologies and applications-ADTTA 99, 7
10. Usanov VI, Pankratov DV, Popov EP, Markelov PI, Ryabaya LD, Zabrodskaya SV (1999) Long-lived radionuclides of sodium, lead-bismuth, and lead coolants in fast-neutron reactors. *At Energy* 87:658–662
11. Sunil C, Biju K, Sarkar PK (2013) Estimation of prompt neutron and residual gamma dose rates and induced activity from 0.1–1.0 GeV protons incident on lead-bismuth target. *Nucl Instrum Methods Phys Res, Sect A* 719:29–38
12. Agosteo S, Magistris M, Silari M (2005) Radiological considerations on multi-MW targets. Part I: induced radioactivity. *Nucl Instrum Methods Phys Res, Sect A* 545:813–822

13. Maiti M, Ghosh K, Mendonca TM, Stora T, Lahiri S (2014) Comparison on the production of radionuclides in 1.4 GeV proton irradiated LBE targets of different thickness. *J Radioanal Nucl Chem* 302:1003–1011
14. Hammer B, Neuhausen J, Boutellier V, Linder HP, Shcherbina N, Wohlmuther M, Turler A, Schumann D (2014) Analysis of the ^{207}Bi , $^{194}\text{Hg}/\text{Au}$ and ^{173}Lu distribution in the irradiated MEGAPIE target. *J Nucl Mater* 450:278–286
15. Schumann D, Neuhausen J, Michel R, Alfimov V, Synal HA, David JC, Wallner A (2011) Excitation functions for the production of long-lived residue nuclides in the reaction $\text{nat Bi}(p; xn, yp)Z$. *J Phys G: Nucl Part Phys* 38:065103
16. Lorenz T, Dai Y, Schumann D, Turler A (2014) Proton-induced polonium production in lead. *Nucl Data Sheets* 119:284–287
17. Gloris M, Michel R, Sudbrock F, Herpers U, Malmberg P, Holmquist B (2001) Proton-induced production of residual radionuclides in lead at intermediate energies. *Nucl Instr Methods A* 463:593–633
18. Hammer B, Schumann D, Neuhausen J, Wohlmuther M, Turler A (2014) Radiochemical determination of polonium in liquid metal spallation targets. *Nuclear Data Sheets* 119: 280–283
19. Ferrari A, Sala PR, Fasso A, Ranft J (2011) *Fluka: a multi-particle transport code (Program version 2011)*
20. Artisyuk V, Saito M, Stankovskii A, Korovin Yu, Shmelev A (2002) Radiological hazard of long-lived spallation products in accelerator-driven system. *Prog Nucl Energy* 40:637–645
21. Lemaire S, David JC, Leray S (2007) Simulation of helium and residue production in the Megapie target. In: *International conference on nuclear data for science and technology*

Design of Farmland Environment Remote Monitoring System Based on ZigBee Wireless Sensor Network

Yongfei Ye, Li Hao, Minghe Liu, Hongxi Wu, Xiao Zhang
and Zhisheng Zhao

Abstract To change the traditional management of agricultural production, using ZigBee technology for short distance wireless transmission to design intelligent farmland environment remote monitoring system, which integrated communication, computer and network all aspects of technology. The real-time, accurate data collection of farmland soil PH value, temperature and humidity surrounding the plant, light intensity, crop growth and bacteria occur posture, provide reliable data for the intelligent agricultural production, thereby increasing the level of intelligence of agricultural management, and promote modernization of agricultural production process.

Keywords Farmland environment remote monitoring · Zigbee technology · Wireless sensor network · Precise agriculture

1 Introduction

Agricultural production plays an important role in human society. Since the birth of human society, it experienced primitive agriculture, traditional agriculture two stages and now entered the era of modern agricultural development. Modern agriculture with times characteristics, mainly using information technology in agricultural production, through the acquisition and analysis of spatial data, builds a knowledge-based agricultural management system [1].

Y. Ye · L. Hao · H. Wu · X. Zhang (✉) · Z. Zhao
School of Information Science and Engineering, Hebei North University,
Zhangjiakou, Hebei, China
e-mail: 780117251@qq.com

Y. Ye
e-mail: yeyongfei005@126.com

M. Liu
School of Economics and Management, Hebei North University,
Zhangjiakou, Hebei, China

ZigBee technology is a mature technology for wireless communication, with its low consumption, low cost, low speed, simple operation and other advantages, in the construction of wireless sensor network favored. ZigBee Alliance to develop communication network layer and application layer protocols and performs networking, security, authentication, etc., to protect the different ZigBee devices compatible with work. The topological structures of the network based on ZigBee technology are star shaped, tree (the cluster shape) and network three kinds [2]. When building a modern agricultural environment monitoring system, a lot of low speed sensor nodes should be arranged in the farmland for data acquisition. Adapting the small probability of network environment changing with time, the system selects ZigBee technology for network formation. Remote monitoring system for agricultural environment can realize real-time and accurate data acquisition, provide reliable data for the intelligent agricultural production decision, guide agricultural production and solve the key issues in precision agriculture.

2 Overall Design Framework of Farmland Environment Data Monitoring System

2.1 Goals of System Design

Crop growth environmental data is collected remotely by using the farmland environment data monitoring system, including soil pH value, plants surrounding temperature, humidity, light intensity, crop growth and bacteria occur posture and so on. The system will provide reliable data for the intelligent decision system, by using expert knowledge system to guide production and determine the amount of production of irrigation, weed pest agents and the amount of fertilizer.

2.2 Overall System Structure

The overall framework of the system based on ZigBee is shown in Fig. 1.

In the information collection system, we should synthesize a variety of techniques to collect data. Firstly, the hierarchical (cluster) topology is used to lay the sensor in the farmland; there are three types of nodes in the network: information gathering node, coordinating node and gateway node. In the system, the gateway node is mainly used for information exchange with the server, and the coordination node plays a link between the gateway and the information gathering node. The system is used to collect data, such as soil pH value, temperature and humidity, light intensity, crop growth, and bacteria occurring in the crop growing environment.

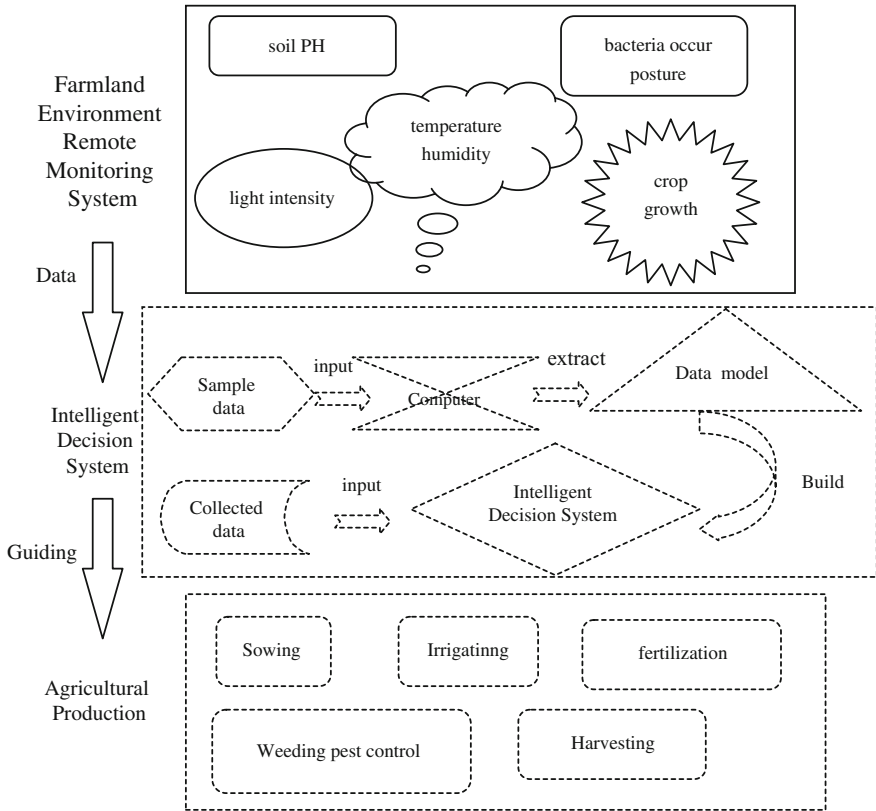


Fig. 1 The overall framework of the intelligent agricultural monitoring system

The formation of intelligent decision system needs to study by many times. Firstly, input the sample data onto computer, then the model is extracted after study repeatedly; finally the intelligent decision system is constructed according to the model. After the formation of the mature decision system, the data collected by the information acquisition system could be inputted in the system, that strategy guide agricultural production activities is formed. This module uses the precision agricultural prescription intelligent generation system proposed by Chen [3].

Agricultural production is the core of agricultural activities, and affects the crop yield directly. In this module, it involves many operation steps of sowing, irrigation, fertilization, weeding, and pest. According to the guidance of the intelligent decision support system, the agricultural production will reflect the precise operation and change the experience operation of the traditional operation. At this stage, the operation needs support of intelligent agricultural. The valve of farm machinery control sowing density and fertilizer quantity control as the node to join wireless

sensor networks based on ZigBee can be realized variable seeding, fertilization according to the current farm environment [4].

This design mainly realizes the function of farmland environment remote monitoring system, and then sends the data to the intelligent decision system for statistical analysis proposed by Chen Yunping, finally to guide agricultural production according to the statistical analysis results.

3 Design of Farmland Environment Remote Monitoring System

3.1 Overall System Design

In the information acquisition system, the sensor is used as the data acquisition node and the information is transmitted through wireless channel. The system is based on ARM processors, ZigBee module, WIFI module for the hardware platform and embedded Linux system with Boa server for the software platform, using C language and HTML language to develop web applications, using SQLite3 database to store data. The system frame is shown in Fig. 2.

3.2 System Hardware

The hardware platform of the monitoring system is based on the TQ2440 development board. TQ2440 is a professional tool for developing a variety of embedded application systems; it uses the S3C2440 of ARM9 chip made by Samsung Corp as CPU. The development board has been widely used in the vehicle holding, network

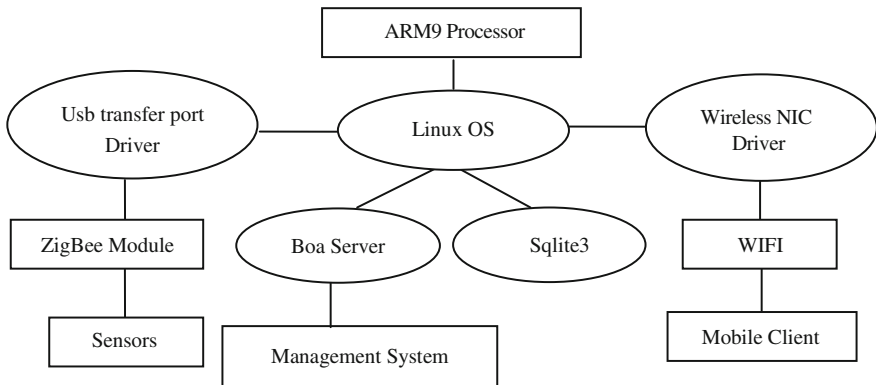


Fig. 2 System framework

monitoring, industrial control, detection equipment, instrumentation, intelligent terminals, medical devices and other products embedded high-end products. The wireless sensing network is based on the ZigBee module with the CC2530 chip made in TI, because CC2530 can establish a strong network node with very low cost. The monitoring system uses the WIFI module to access the Internet.

3.3 System Software Design

System software design includes the transplant of application, Web module, system database designing and ZigBee module four aspects.

3.3.1 The Application Transplant Design

- Embedded System Transplantation

The embedded Linux system transplantation is divided into Bootloader transplantation, kernel transplantation, and file system transplantation. Bootloader is a program that running before the embedded system starting, helping to initialize the hardware device, and preparing for the kernel and file system transplantation.

BosyBox is a software that integrates more than one hundred common Linux commands and tools. Using BusyBox software for transplantation aimed to product a basic file system, and then add the corresponding application to it based on the need. The whole transplantation system flow chart is shown in Fig. 3.

Fig. 3 The transplantation system flow chart

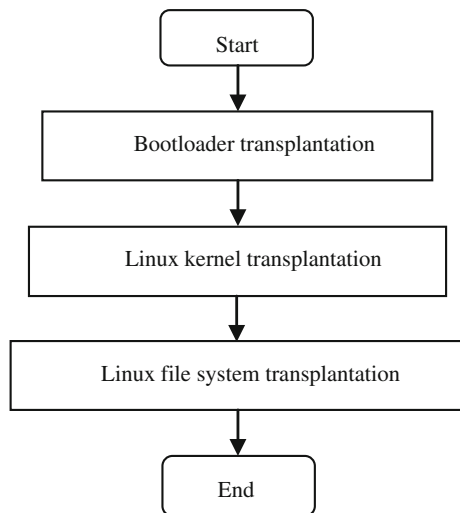
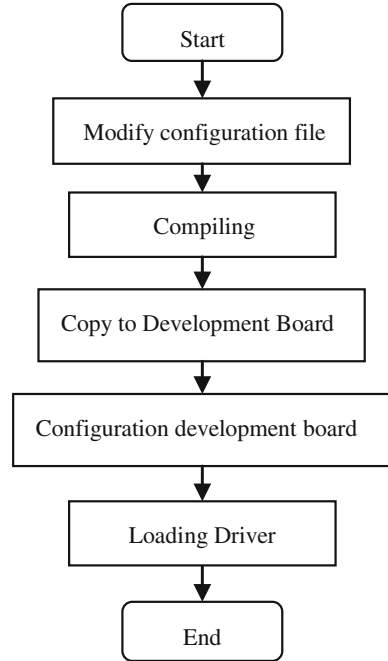


Fig. 4 Wireless network NIC driver porting flow chart



- **WIFI Wireless NIC Driver Porting Design**

After the transplantation of embedded Linux system, the driver of the WIFI wireless NIC is modified and loaded into the ARM9 development board after the successful compiling. When the WIFI wireless NIC loaded successfully, the WIFI signal is generated, and users are connected to the network by devices. Wireless network NIC driver porting flow chart is shown in Fig. 4.

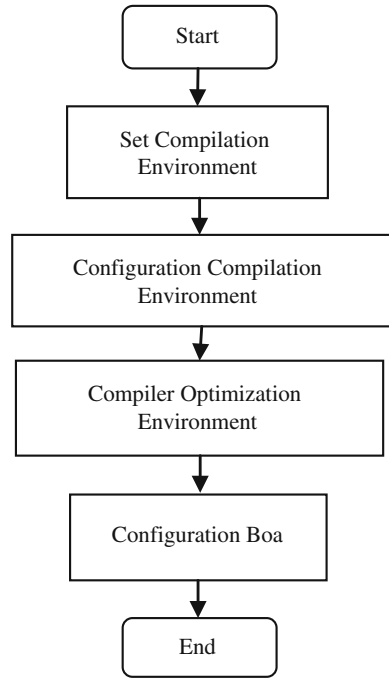
- **Boa Server Porting Design**

Boa is a compact Web server. It runs on Linux and supports CGI programs. Boa server porting flow chart is shown in Fig. 5.

- **Sqlite3 Database Management System Porting Design**

Serial port monitoring program stores the data into the database, such as temperature, humidity, light intensity, soil pH value, crop growth and bacteria occurring that are collected in the field, and also stores the new nodes data in the database. Sqlite3 database management system porting flow chart is shown in Fig. 6.

Fig. 5 Boa server porting flow chart



3.3.2 Web Module Design

User enters the login page by using Web browser after connected WIFI signal successful. There are running state, basic settings, remote display and remote control four functions on the main interface. User chooses a function and requests the Web server according to their own needs, Web server processes the user's request and returns the required data when receiving the request.

Through the serial port, the web sends commands to the coordinator, and coordinator sends it to the terminal node. If the terminal node receives a reading command, the sensor data is read and stored in the SQLite3 database and transmitted to the coordinator. If the terminal node receives the control command, it will control the signal light.

3.3.3 System Database Design

In order to storage the system data, the ZigBee node and ZigBee data collection two tables are built in SQLite3. ZigBee node table is about node's information in the network, mainly recording node's address in the network (addr), the node's physical address (paddr), node's status (status), node's type, the structure is shown in Table 1. ZigBee data collection table stores related data, including the node's

Fig. 6 Sqlite3 database management system porting flow chart

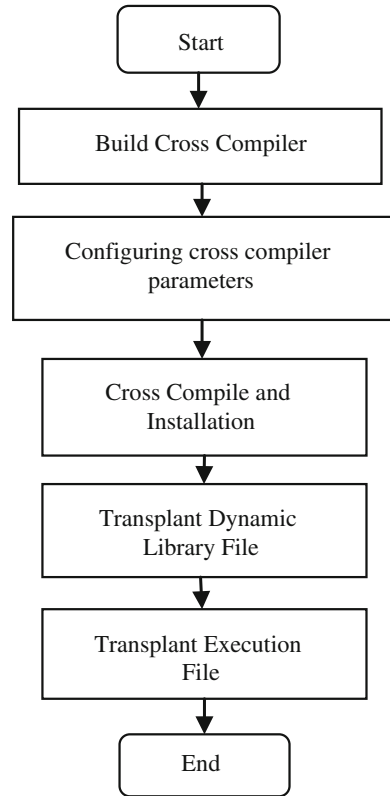


Table 1 ZigBee node table

Primary key	Column name	Type	Null
Yes	addr	Integer	No
Yes	paddr	Integer	No
No	status	Integer	Yes
No	type	Char	Yes

address in the network (addr), temperature (temp), humidity (humidity), illumination (light), soil pH value (PH), crop growth (croGrowth) and bacteria occurring (bacState) in the crop growing environment, Table 2 showing the structure.

3.3.4 ZigBee Module Design

ZigBee wireless transmit-receive module adopts a star network topology. There are two kinds of devices in the network, the coordinate node and the terminal node. The coordinator is responsible for establishing the network, after the system is powered up, the application layer is initialized at first, and then the network is built. Once the

Table 2 ZigBee data collection table

Primary key	Column name	Type	Null
Yes	addr	Integer	Yes
No	temp	Char	No
No	humidity	Char	No
No	light	Char	No
No	PH	Char	No
No	crGrowth	Char	No
No	bacState	Char	No

network is established, the functions of coordinator same as the router, and allows the nodes to join the network and select the data routing. After the formation of the network, terminal nodes in a waiting state, when receiving the reading command sent by coordinator, terminal nodes began acquisition of temperature, humidity, light intensity, the soil pH value, crop growth and bacteria occurring, and then transmit the data to the coordinator. Coordinator receives terminal node data and transmits the data to the ARM9 processor. When the coordinator sends the command of control LED, the terminal node will control the LED turn on, and the status of the LED can be used to test the equipment.

4 Realization of Farmland Environment Remote Monitoring System

The realization of the remote monitoring system mainly includes the transplantation of application, Web module, the system database, the ZigBee module four parts.

The transplantation of the application is mainly from the embedded system transplantation, the WIFI wireless NIC driver transplantation, the Boa server migration, the database management system transplant four aspects to realize.

Web module mainly realizes the functions of user login, basic information display and setting, remote control and data display and serial function module.

Serial reading and writing is the communication key between the web and ZigBee to read data in a loop process. If the serial listener process receives the command head of RP, the reading data is inserted into the ZigBee data acquisition table. If the serial listener process receives the command head of JO, which indicates that a new node joins the network, the new node information is stored in the ZigBee node table.

Pipe monitoring process receive the data is sent by page through a pipe and write it to the serial port. If receiving a reading command, the coordinator sends reading command to the terminal node. After the terminal node receives the command, reads the relevant sensor and returns the reading data to the page. If the terminal node receives a control command, the corresponding operation of the LED is performed.

After creating the database, two tables need to manually create. The terminal node table for terminal node information and the ZigBee data acquisition table for storing sensor data.

The system takes the ZigBee network nodes as the main module. CC2530 coordination node communicates with the ARM9 processor by serial. CC2530 terminal nodes are connected to all kinds of sensors and acquisition of temperature, humidity, light intensity, the soil pH value, crop growth and bacteria occurring dynamic data. CC2530 coordinator and CC2530 terminal node communicate via ZigBee protocol. The realization process is as follows:

- CC2530 Coordination Node Implementation

The coordinator node mainly completes the establishment and management of ZigBee network, and carries on the data transmission with the terminal node, and performs the data exchange with the Web server by serial port and so on. After the system is powered up, the coordinator completes the initialization and establishes the network, waiting for the new terminal node to join the network. The CC2530 coordinator has two monitoring functions, one is the data sending request of the terminal node, and the other is the control command of the Web server.

When the terminal node requests the network connection, the coordinator receives `ZDO_STATE_CHANGE` message. After receiving the message, the coordinator node packages the frame header, physical address, network address, and transmits it to the Web server by the serial port and stores the data in the Sqlite3 database.

When the coordinator received the data from web server by serial port, the application layer will received message `SPI_INCOMING_ZTOOL_PORT`. By calling the function `UartRxComCallBack` to analysis and sends the command to a terminal, and terminal nodes will make corresponding processing to send data to the coordinator. After a series processing by other protocol layers, the application layer only receive `AF_INCOMING_MSG_CMD MSG_CMD` message. The coordinator sends the received information frame to the Web server via the serial port, and stores the data in the Sqlite3 database.

ZigBee protocol requires each node should have physical and network two addresses. Each node must have a unique physical address in the network, and the network address of the node is automatically assigned by the gateway, and will change with the route when adding the network. After nodes join the network, the network address and physical address will be matched to the node.

- CC2530 Terminal Node Implementation

Terminal nodes respond to coordinator node.

Considering the low-power, the terminal node uses battery as the power supply, and in sleeping status when no work to do. After powered up, terminal node completes a series task of initialization and find coordinated network node that startup successfully, waiting for joining the network. When joined the network successfully, terminal nodes will send their information (node type, node address) to the coordinator node and monitors if there is data sent back. When received

Table 3 The communication protocol table

Task type	Command head	Father's address	Network address	Nodes type	Data buffer
2 Bytes	2 Bytes	2 Bytes	2 Bytes	3 Bytes	7 Bytes

AF_INCOMING_MSG_CMD MSG CMD message, terminal nodes analyze the received data. If it's a reading command, the relevant data is transmitted to the coordinator node. If it's a control command, the corresponding operation of the LED light is performed. When completed all task, the terminal node enters the sleeping status.

- Serial Communication Implementation

The coordinator needs to interact with the Web server via the serial port, so the coordinator needs to initialize and process the data by using MT_UartInit function.

We should define the receiving buffer and sending buffer of sending serial port, and the buffer size can be defined according to needs. By calling the HalUARTRead function to read the serial data. The communication between Web server and coordinator is bidirectional, including Web server sending data to coordinator and coordinator returning data to Web server. In order to facilitate data communication, serial communication protocol is needed. The communication protocol table is shown in Table 3.

5 System Test Results

For reliable transmission distance of ZigBee is 10–75 m, so farmland is divided into regions by 50 m * 50 m interval and in each region placed ZigBee terminal node, ARM9 development board, ZigBee coordinator and the WiFi module to construct wireless network and all regions are connected together.

Mobile phone users login server by connecting WiFi and select anyone module to enter the interface. If chose the “remote display” module and selected the region number, all data collected in the region are shown in the page, such as temperature, humidity, light intensity, soil pH, bacteria occurred dynamics and crop growth information. Login server on the mobile phone, then clicking “remote control” and selecting region 1, can control the LED turn on or off in region 1. Clicking “add” button can add another region, if put the other ZigBee terminal node into the area, you can also achieve analog lights control. From the experimental results, this system has good scalability, can change the capacity of the network according to the need.

In the experiment, the multi user mobile phone is chosen to connect with WIFI and login the server. Experimental results show that multi users can achieve concurrent access to the network and acquire the relative data. After testing, the modules of the system are normal to complete all tasks.

6 Conclusion

Based on the analysis of current situation of agriculture development, design the farmland environment remote monitoring system with the help of ZigBee technology. The client communicates with control center by WiFi and remote monitoring related parameters about crop growth in the field. In the system, through the establishment of cross compiler environment, transplant Linux system, choose Boa as the Web server, develop CGI program, realize the interaction with users. Design remote web monitoring platform and remote monitoring temperature, humidity, light intensity of illumination, soil pH, bacteria occurred dynamics and crop growth information on the Web and remote monitoring terminal equipment is working properly through control LED.

Experimental results show that the system realized remote real-time and accurate monitoring of farmland environment data on mobile phone, notebook, PC and so on. The data can be further conveyed to the intelligent decision-making system and guide agricultural production according to the decision-making results. All these what we do will promote the development of Chinese precision agriculture.

Acknowledgments (1) Major Research Projects of Hebei North University (ZD201303); (2) Hebei Province Population Health Information Engineering Technology Research Center.

References

1. Sun Y-w, Shen M-x (2008) A summary of precision agriculture and the “3S” technologies. *Gansu Agric Sci Technol* 12:39–42
2. Pang N, Cheng D-f (2010) Design of greenhouse monitoring system based on ZigBee wireless sensor networks. *J Jilin Univ (Inf Sci Ed)*, 2010, 28(1):55–60
3. Chen Y-p, Zhao C-j, Wang X et al (2007) Prescription map generation intelligent system of precision agriculture based on knowledge model and WebGIS. *Sci Agric Sin* 40(6):1190–1197
4. Yu H, Luo H, Ren S et al (2012) Application progress and prospect of ZigBee technology in precision agriculture. *J Agric Mechanization Res* 8:1–6

Attractions and Monuments Touring System Based on Cloud Computing and Augmented Reality

Deqiang Han, Zongxia Wang and Qiang Zhang

Abstract This paper discusses the design and implementation of an interactive attractions and monuments touring system based on cloud computing and augmented reality technology of Google Earth. The system accesses the user's motion information through the wireless motion detection module, and then transforms the three-dimensional scenes of real objects correspondingly. It also provides text, sound and other forms of information to make an immersive touring experience possible. With the help of system tools, the content managed by the system (information about attractions, monuments and recommended tourist routes, etc.) can be added, deleted, and modified to improve system extensibility and usability.

Keywords Augmented reality · Attractions and monuments tours · Motion detection · Cloud computing · Google Earth

1 Introduction

Attractions and monuments are the best records and reflections of ancient culture, visit attractions and monuments is the one of the most common ways to feel and learn the ancient culture. However, the traditional way of travelling takes a lot of time and money, thus for those who are busy and budgeted the opportunity to visit and learn ancient culture is limited, over time the lacking of ancient cultural phenomenon become obvious [1]. In addition, traditional forms of tourism will

D. Han (✉) · Z. Wang · Q. Zhang
Beijing University of Technology, Beijing, People's Republic of China
e-mail: handq@bjut.edu.cn

Z. Wang
e-mail: wzongxia@bjut.edu.cn

Q. Zhang
e-mail: Johnson9009@163.com

inevitably result in different levels of vandalism, and increasing difficulties of cultural heritage's restoration and protection. So monuments electronic tour system came into being. We found that most of monuments electronic tour systems have many defects, such as monotonous screen, limited interaction and lack of user attraction especially for younger generation. Therefore, we use self-designed motion detection module, and cloud computing and augmented reality technology of Google Earth, developed a monuments touring system that collects somatosensory tour, intelligent guide, audio narration, three-dimensional scene in it, to help people access the world without leaving home. With attractions and monuments recovery capabilities offered by the system, visitors can vividly feel verve and grand of the damaged ancient buildings.

2 System Overview

The high-sensitivity acceleration sensor and gyroscope are used in our touring system to collect human motion data. The data are transmitted to a Bluetooth receiving end wirelessly and finally arrive at the most important part of the whole system, the Atom processor platform, through USB port. Then the software installed on the Atom processor platform analyzes and processes the motion data and transform three-dimensional scene in real time. In order to enrich our touring system, graphics, text and audio information are added into many of the iconic attractions in the three-dimensional scene. The added information will be displayed intelligently when a tourist passes by. Touring system architecture is shown in Fig. 1.

The angle measurement nodes tied around the head to measure the rotation angle of the head. When user's head is turning, three-dimensional scene is rotating accordingly, which gives users a real touring experience. Motion detection node equipped around the leg detects leg action made by the user. In this way, the user can interact with the touring system, such as walking, running, switching places. Our touring system contains a lot of representative buildings, text and audio information of introductions. We must consider how to update the content that the system provides. After several trials and errors, we decide to build and manage a database containing all the introduction information on server. Therefore, as long as the database on the server is updated, the client can get the most up-to-date services without changing anything.

3 Design of Motion Detection Module

As the front-end of motion detection system, wireless motion detection module must have these features: small size, low power consumption, high stability, and convenience of data transmission. Shown in Fig. 2, the module selects low-power

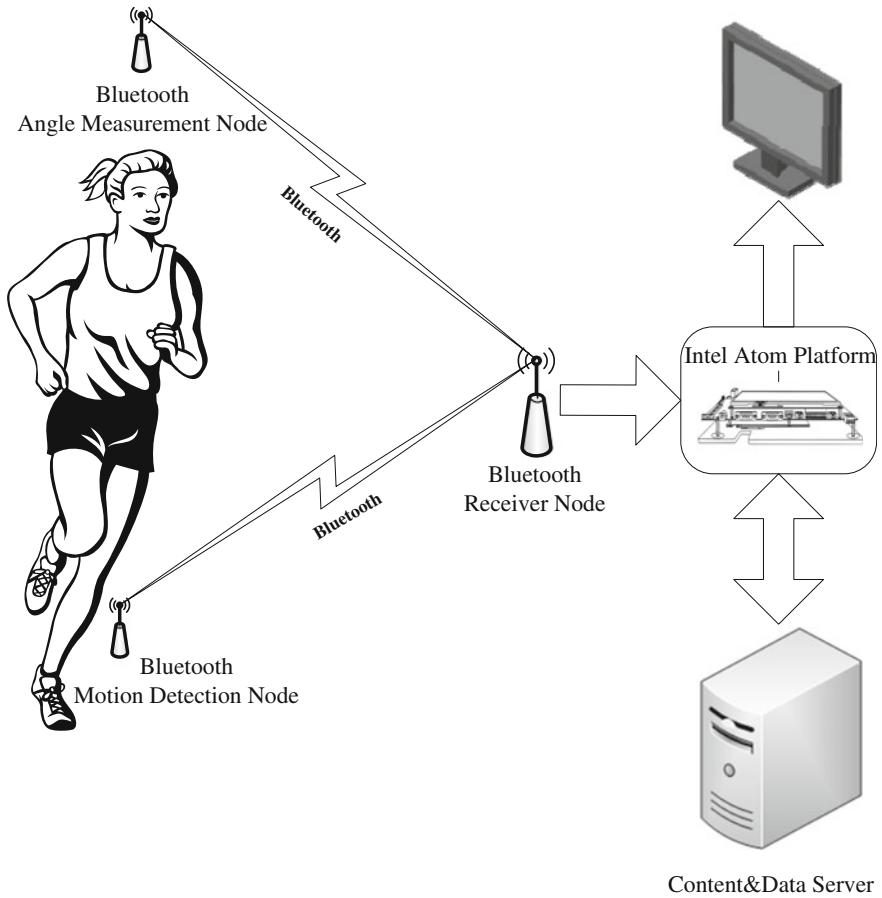
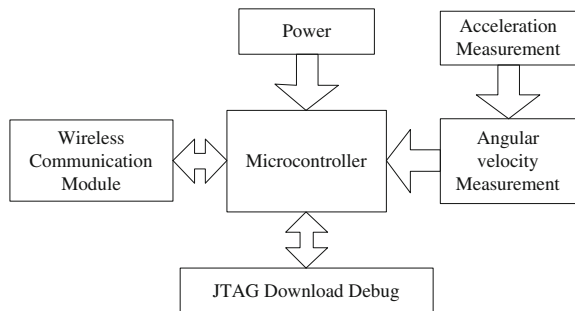


Fig. 1 Touring system architecture diagram

Fig. 2 Wireless motion detection module block diagram



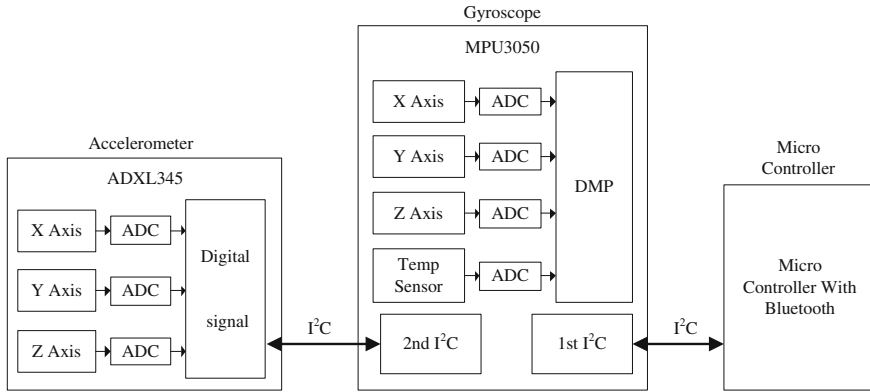


Fig. 3 Six degrees of freedom detection system block diagram

microcontroller as controller and introduces gyroscope which has a better dynamic performance of angle measurements than acceleration sensor. Although acceleration can be measured only use the acceleration sensor, but because of poor dynamic performance, acceleration sensor will lead to the acceleration of gravity component removing incomplete and coordinate transformation abnormal phenomenon. Therefore, gyroscope is introduced to this system, using gyro angle output information to help remove the gravitational acceleration component and transform the coordinate information, thereby enhancing the stability of the module.

The sensor module consists of acceleration sensors and gyroscopes. As MPU3050 contains a digital motion processor (DMP) that integrates the raw data of the gyro and acceleration sensors, we select MPU3050 as the gyroscope chip. Below is the block diagram which illustrates construction of six degrees of freedom detection system using MPU3050 shown in Fig. 3.

Shown in Fig. 3, the data of Accelerometer can be automatically read by MPU3050's second I²C interface [2], so Accelerometer can be connected to gyroscope directly. Gyroscope gets acceleration data from the accelerometer, and then put it into the built-in data fusion DMP with its own angular velocity data. Finally, microcontroller gets the result of data fusion via the first I²C interface of MPU3050.

4 Establishment of Spatial Movement Model

4.1 Relative Rotation Using Quaternion

Use angular velocity data outputted by gyro and acceleration data of accelerometer, through Kalman filter, numerical integration methods, eventually we can obtain accurate rotation angle of rigid body in three-dimensional space. DMP can

complete angle calculation, by DMP, we can get the absolute rotation angle which is represented by quaternion. However, in most applications, it is more convenient to use a relative angle of rotation, combining quaternion equations related formula, absolute rotation angle can be further processed to a relative rotation angle.

Quaternion math concept is Hamilton (Hamilton, W.R., 1805–1865) presented in 1843, it was originally designed to take advantage of two complexes to build three-dimensional spaces. After no success, he proposed expanding the imaginary unit, and ultimately formed Hypercomplex which included four real elements w, x, y, z and consist of a solid number of units 1 and three imaginary unit i, j, k composition, said quaternion [3].

Rotation conversion formula is represented by quaternion as follows.

$$p' = qpq^{-1} \tag{1}$$

where p is quaternion of the point before rotation, q represents quaternion of rotation angle, q^{-1} represents the inverse of q , p' is quaternion of the point after rotation. Assume that quaternion of the object before rotation is p , quaternion after rotation is p' , quaternions of two consecutive absolute rotation angle are A and B , after rotate the object by the A angle, we can get the following relationship.

$$p' = ApA^{-1} \tag{2}$$

After then rotate the object by the B angle, result as follows.

$$p' = BpB^{-1} \tag{3}$$

$$\therefore A^{-1}A = 1 \quad \therefore p' = BA^{-1}ApA^{-1}AB^{-1} \tag{4}$$

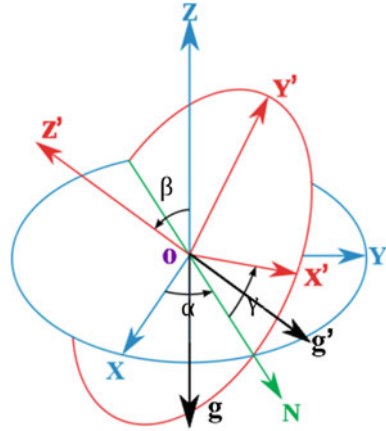
Difference of two quaternions of consecutive absolute rotation angle is BA^{-1} , Actually, it is quaternion of relative rotation. In summary, the formula by which we get quaternions of relative rotation $Q_1 \dots Q_n$ from quaternions of absolute rotation angle $q_0, q_1, q_2 \dots q_{(n-2)}, q_{(n-1)}, q_n$ as follows.

$$Q_n = \begin{cases} q_0 & (n = 0) \\ q_n q_{(n-1)}^{-1} & (n \geq 1) \end{cases} \tag{5}$$

Characteristics of quaternion decided that it can represent rotation only in normalized conditions, so it must be normalized by formula (6).

$$i = \frac{i}{\sqrt{w^2 + x^2 + y^2 + z^2}} \quad (i = w, x, y, z) \tag{6}$$

Fig. 4 The principle of removing the component of gravity acceleration



4.2 Coordinate Transformations and Gravity Component Removal

In the process of measuring object motion information, the rotation of the object itself leads to misalignment between measuring module’s own coordinate and world coordinate system. The components of gravitational acceleration at the measurement axis will be superimposed on the acceleration of the output [4]. If the speed and displacement information are calculated by speed and the calculation formula of absolute or relative displacement, they will inevitably increase the error of the calculation results.

As shown in Fig. 4, OXYZ is the coordinate system before rotation. g is gravitational acceleration. $OX'Y'Z'$ is the coordinate system after rotation. g' is the equivalent gravitational acceleration in the $OX'Y'Z'$ coordinate. To get the components of gravitational acceleration g on the axis OX' , OY' and OZ' can be equivalent to evaluate the components of equivalent acceleration of gravity g' in the OXYZ coordinate system. The gravitational acceleration is perpendicular to the plane OXY before rotation. So the components on the axis OX and OZ is 0, and the component on the axis OZ is g . The components of gravitational acceleration on the three axes can be expressed by a three-dimensional vector as follows:

$$g_{OXYZ}(g_x, g_y, g_z) = (0, 0, g) \tag{7}$$

Equation (7) states that we can find a dot, $G(0, 0, g)$, in the OXYZ coordinate. The line between the dot and origin point is perpendicular to the plane OXY, and overlaps with the axis OZ. Three coordinate values of the dot are the components of gravity acceleration on each axis. Because the coordinate $OX'Y'Z'$ is derivate from the rotation of the coordinate OXYZ, we can obtain the equivalent dot G' of the dot G in the coordinate $OX'Y'Z'$ according to the quaternion $Q(w, x, y, z)$, which represents the rotation information, and Eq. (8).

$$g_{OX'Y'Z'}(0, g'_x, g'_y, g'_z) = QQ_GQ^{-1} \tag{8}$$

In Eq. (8), Q_G is the quaternion constructed by dot G , that is $(0, 0, 0, g)$. We suppose the quaternions $Q_1(w_1, x_1, y_1, z_1)$, $Q_2(w_2, x_2, y_2, z_2)$, $Q_3(w_3, x_3, y_3, z_3)$, and $Q_3 = Q_1Q_2$. According to the multiplication formula of quaternions, we can give

$$\begin{cases} w_3 = w_1w_2 - x_1x_2 - y_1y_2 - z_1z_2 \\ x_3 = w_1x_2 + x_1w_2 + y_1z_2 - z_1y_2 \\ y_3 = w_1y_2 - x_1z_2 + y_1w_2 + z_1x_2 \\ z_3 = w_1z_2 + x_1y_2 - y_1x_2 + z_1w_2 \end{cases} \tag{9}$$

Consequently, the component of gravitational acceleration on the axis OX' is calculated as $2(x \cdot z + y \cdot w)g$, the axis OY' is $2(z \cdot y - x \cdot w)g$, and the axis OZ' is $(1 - 2(x^2 + y^2))g$.

The measurement axis of acceleration sensor is fixed. The acceleration measuring after rotation is not relative to world coordinate system. It cannot reflex the real motion of a moving object, even though the components of gravitational acceleration are removed. In practice, we expect a measured acceleration value that is independent of own rotation of moving object. So, we need to transform the coordinate of acceleration [5].

The key is the correction of acceleration direction for resolving the affect to acceleration by rotation. It is easier to understand analyzing this problem from the acceleration sensor. It seems to acceleration sensor that the world only have three measurement axis. The judgment to itself motion state can only depend on these data measuring from three axis.

When move acceleration is occurred under no rotation shown in Fig. 5, acceleration sensor detects acceleration only on the axis Z' . It is regarded as moving up. That confirms to the fact. But under the circumstance as shown in Fig. 6, the acceleration sensor believe that it is oblique moving upward under the action of an oblique upward force. The face is the measured object is moving up. There is a mistake here. The solution can be interpreted as rotating acceleration a in Fig. 5 in accordance with transformation from the coordinate XYZ to $X'Y'Z'$. Then acceleration a and the axis Z' coincide again, and there is not a component on the axis Y' . Finally, the sensor only detects acceleration on the axis Z' . It thinks that itself is moving upward as the fact.

Fig. 5 The relationship of two coordinate systems before rotation

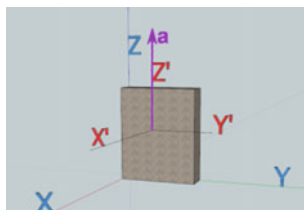
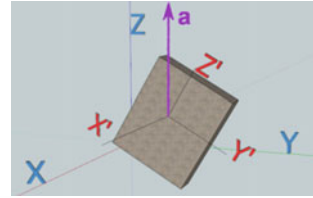


Fig. 6 The relationship of two coordinate systems after rotation



According to this theory and quaternion transformation formula, the coordinate transformation formula of acceleration data is given by

$$a' = QaQ^{-1} \quad (10)$$

where a' is the transformed acceleration. Q and Q^{-1} are the quaternion of rotation angle and its reciprocal. a is the acceleration removed the component of gravity acceleration.

5 The Implementation of Google Earth Virtual Touring

The program of touring scenic spots and historic sites needs huge amount of landforms and architectural models information. The sheer size of the information is needed is too large for our small team. Thus we work with Google Earth to further develop the real and entire geographic information and models. That avoided the redundancy of data, and also greatly improved the reusability of software.

Data of Google Earth includes images in the public domain, aviation images with permissions, KeyHole spy satellite images, and a lot of urban images taken by other satellites. It runs cloud computing technology to combine satellite images from multiple sources, aerial images from more than one source and images uploaded by users into one. Ultimately, they adjust each other and compose a three-dimensional scene, which makes users to have an immersive experience. As shown in Fig. 7, the program of touring scenic spots and historic sites is composed by several different sub-modules. It ensures development efficiency and reusability of programs.

5.1 Data Processing

After users motion data received by the program of touring scenic spots and historic sites, the system will analyze and process the data in the first place. Through analyzing, we found out there are few interference from other frequency signals to acceleration data. Hence we mainly need to filter the signals based on signal amplitude. Because of the acceleration sensor's feature, the sensor's output will not return to zero after using it for some time. An offset will appear as shown in Fig. 8a.

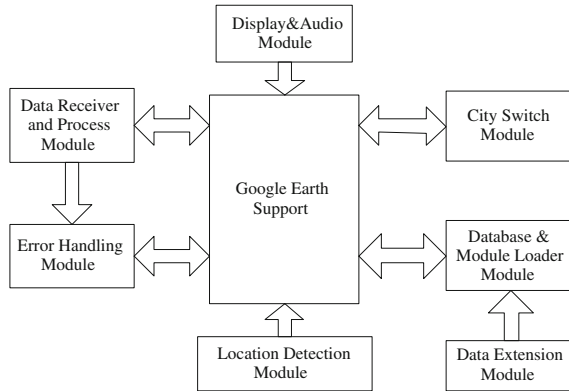


Fig. 7 The programming framework of touring scenic spots and historic sites

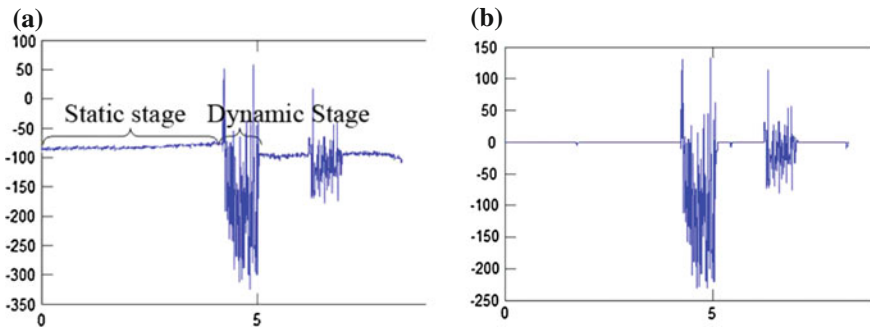


Fig. 8 The waveforms of self-learning baseline filter. a Before, b after

In addition to the nonlinearity of motion, size and direction of the offset (intensity and direction of motion) cannot be known.

This system designed and implemented a filter that could learn acceleration data on idle state by itself. It can filter out small disturbance signal. Meanwhile it can also adjust the baseline of waveform to zero. The result is shown as Fig. 8b.

Our algorithm mainly includes data adjustment and judgment for static stage or dynamic stage. The methods of adjustment are different at different stages. At dynamic stage, the method is subtracting the average value at static stage from every acceleration data. At static stage, the method is setting it to zero directly. The static or dynamic stage judgment has two cases, one is from static stage to dynamic stage. We should set a static threshold value. When the absolute value of acceleration data is greater than the static threshold value, we consider it has entered the dynamic stage. The other case is from dynamic to static. We should also set a dynamic threshold value. When the absolute value of the difference between several continuous acceleration data (the number is determined by static inspection number) and

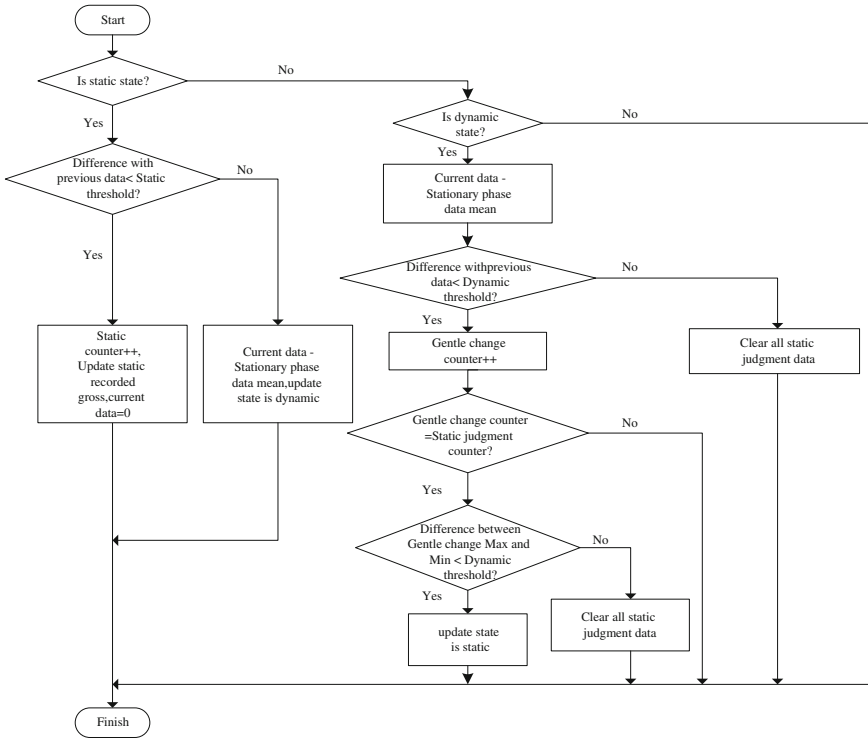


Fig. 9 The diagram of self-learning baseline filter

previous acceleration data is less than the dynamic threshold value, and the absolute value of the difference between the maximum and the minimum is less than the dynamic threshold value, the acceleration data change gently and it has been the static stage. The flow diagram of self-learning baseline filter is shown as Fig. 9.

5.2 Data Management of Geographic Information

After analysis and processing movement data, the program of touring scenic spots and historic sites can realize these functions such as movement, perspective transformation and towns switching over, etc. There are two defects if all necessary functions of the system (recommend touring route, introduce tourist spots in text, graphics and audio, design model files by users, and so on) are implemented in one program. Firstly, it is not easy to modify. Every modification needs to rebuild system. Secondly, the system will be over complicated.

KML is a language that uses XML grammars and format to describe and save geographic information and can be recognized and displayed by Google Earth and

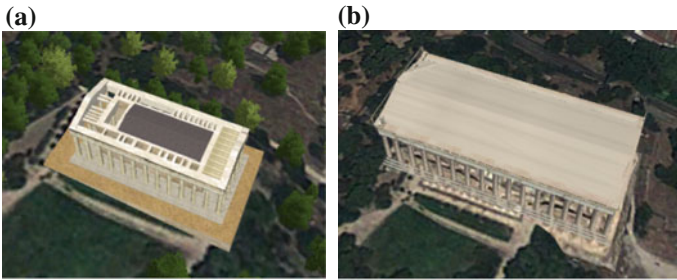


Fig. 10 The sketches of the Parthenon in Athens. **a** Before repaired, **b** after repaired

Google Maps. KMZ is compressed KML files. So, it can pack all of other files that were required by KML files [6]. KML and KMZ files offered by Google Earth can describe these essential contents efficiently and normatively. With the addition of Google Earth's native support to parse these files efficiently, it can relieve the burden of programs and expand system conveniently that these required contents are written into KML or KMZ files.

Furthermore, in Google Earth there are numerous damaged historical buildings. This program can add repaired historic building models into the system. It gives users a chance to feel the fun of visiting a non-existent scene. On the other hand, we can view the beauty of destroyed historic buildings. The working sketch of the Parthenon in Athens before and after repaired are shown in Fig. 10.

6 Conclusions

Our system integrates multiple advanced technologies such as sensor, wireless transmission technology, data fusion, Google Earth, KML, and so on into one system. It combines ancient culture and modern technology perfectly. The system is different from any previous virtual games. In virtue of mass real geographic data of Google Earth, this system presents a virtual scene the same as the real scenic spot. Working together with wireless motion detection module, users can enjoy the tour just as personally on the scene. User can begin to visit effortlessly and freely in the living room. The original appearance of destroyed historic buildings even can be visited. Also user will feel better if he or she wears VGA video glasses to use this system.

References

1. Richards G (Ed) (2001) Cultural attractions and European tourism. CABI
2. InvenSense (2011) MPU-3050 motion processing unit product specification Rev 2.9
3. Shoemake K (1985) Animating rotation with quaternion curves. In: ACM SIGGRAPH computer graphics, vol 19(3). ACM, pp 245–254

4. Gal-Chen T, Somerville RC (1975) On the use of a coordinate transformation for the solution of the Navier-Stokes equations. *J Comput Phys* 17(2):209–228
5. Ohkawara K, Oshima Y, Hikiyama Y, Ishikawa-Takata K, Tabata I, Tanaka S (2011) Real-time estimation of daily physical activity intensity by a triaxial accelerometer and a gravity-removal classification algorithm. *Br J Nutr* 105(11):1681–1691
6. Ying-jun D, Yu CC, Jie L (2009) A study of GIS development based on KML and Google Earth. In: Fifth international joint conference on INC, IMS and IDC, 2009. NCM'09. IEEE, pp 1581–1585

Constructing Weighted Gene Correlation Network on GPUs

Guanghui Yang, Sheng Zhang, Yuan Tian, Ping Lin,
Jiang-Feng Wan, Qingguo Zhou and Lei Yang

Abstract Here we constructed a weighted gene correlation network in human glioblastoma cells by developing the graphics processing units (GPUs) algorithm. The strength distributions of entire network, housekeeping genes and hubs in protein interaction network were calculated and the differences between them were found. Six definitions of clustering coefficient previously proposed for weighted networks were calculated in this paper and behaved quite differently. Interestingly, the clustering coefficient distributions of housekeeping genes and hubs are similar to that of entire network, as the strengths of them are generally bigger. This work explored how to calculate the network indices in weighted biological networks on GPUs and whether these indices can reflect the characteristics of biological networks.

G. Yang · S. Zhang · Y. Tian · P. Lin · J.-F. Wan · L. Yang (✉)
Institute of Modern Physics, Chinese Academy of Science,
509 Nanchang Road, Lanzhou 730000, Gansu, China
e-mail: lyang@impcas.ac.cn

G. Yang
e-mail: yangguanghui@impcas.ac.cn

S. Zhang
e-mail: halifax@gmail.com

Y. Tian
e-mail: tianyuan08@impcas.ac.cn

P. Lin
e-mail: pinglin@impcas.ac.cn

G. Yang · J.-F. Wan
University of Chinese Academy of Science, No. 19A Yuquan Road,
Beijing 100049, China
e-mail: jiangfengwan@impcas.ac.cn

Q. Zhou
Lanzhou University, No. 222 Tianshuinan Road, Lanzhou 730000, Gansu, China
e-mail: kinggo@gmail.com

Keywords Weighted network · GPU · Gene correlation expression · Biological data analysis

1 Introduction

Graphics processing units (GPUs) originate as specialized hardware useful only for accelerating graphical operations, but in the following they grew into exceptionally powerful equipment, especially for the supercomputing in sciences and engineering. Recently, GPUs have become programmable to the point where they are a viable general purpose programming platform. Contrast to the CPUs, the main advantages of GPUs are the large-scale parallelism ability and the less computing power [1]. As a result, the applications on GPUs have grown explosively in recent years. Many important simulation methods, no matter based on the probability computation, molecular dynamics (MD) or Monte Carlo (MC) algorithms, have been implemented on GPUs, leading to a much better performance than the best CPU implementations usually. For analyzing large-scale biology and medicine data, the computational and systems biology approaches which based on GPUs have been recently developed [2–4]. These application has a great potential for understanding biology and disease, drug design or synthetic biology.

In living cells, the biological regulations can be divided into several different levels, such as the genetic, protein or metabolic levels. At each level there are several regulatory networks, the functions of which include the maintenance of cell viability or responds to stimuli [5–7]. The protein interaction network (PIN) and metabolic network are widely studied as the complex networks so far and a series of instructive results have been obtained to uncover the complexity in biology [8, 9]. Due to the popularity of microarray technology, there are numerous high-throughput genome-wide expression profiles for various cancers and species in public databases. In order to utilize these data, weighted correlation network analysis (WGCNA) has been developed to construct a weighted gene correlation network including thousands of genes [10] and the modified WGCNA has been employed in neuroscience, immunology and cancer research [11–13]. In this type of networks the value of each edge is not a Boolean value but a real number between 0 and 1 [14, 15] and for the weighted complex networks some topological indices, such as clustering coefficient and betweenness, don't have a unique mathematical definition [16, 17]. Because of the various data sources, different definitions of clustering coefficient in weighted networks have been proposed and all definitions can reduce to the classical clustering coefficient for unweighted networks [16, 18].

Here we used WGCNA to construct a weighted gene correlation network in glioblastoma (GBM) on GPUs and compared the calculation results of clustering coefficient by employing different definitions. The housekeeping genes and hubs in PIN are also introduced here, and then the distributions of strength and different clustering coefficients of them were calculated. We found that in the correlation

networks the results were quite different if we used different definitions of clustering coefficient. Moreover, the important genes can be distinguished from the strength distribution and they have different behavior in the clustering coefficient distribution. These results showed the applicability of clustering coefficient in weighted biological networks and reflected some characteristics of them.

2 Methods

We used the following definitions and notations, and introduced some basic indices used to characterize the topology of a network. A network consists of N nodes and L links, and is usually described by the so-called adjacency matrix $A = [a_{ij}]$, where $a_{ij} = 1$ if the nodes i and j are connected, 0 otherwise ($a_{ii} = 0, \forall i$). The degree (or connectivity) of node i is defined as $k_i = \sum_j a_{ij}$. The degree distribution $P(k)$, defined as the fraction of nodes in the network having k links, is the most basic topological characteristic of a network and affects many physical properties. In real life, many networks, including social networks, airline networks and biological networks, display the scale-free property that the degree distribution follows a power law: $P(k) \sim k^{-\gamma}$. Moreover, for many real networks not all links have the same capacity or intensity. For instance, the strength of relationships between individuals in social networks may either strong or weak; the carbon flow between species in food webs is diverse; and the amount of data communication along connections on the Internet can be different. All these systems can be better described by weighted networks, where each link carries a weight (or value) measuring the strength of the connection. Following Barrat et al. [14], a weighted network is described by the weights matrix $W = [w_{ij}]$, where entry w_{ij} is the weight of the link between nodes i and j , and $w_{ij} = 0$ if the nodes i and j are not connected ($w_{ii} = 0 \forall i$). In a weighted network, the strength of node i is defined as $s_i = \sum_j w_{ij}$, and the “disparity” of node i is evaluated by $Y_i = \sum_j (w_{ij}/s_i)^2$. While the definitions of some indices for weighted networks can generalize from similar definitions for unweighted networks, there has been no consensus on the definitions of some other indices which have had well definitions in unweighted networks. For example, there were six versions for the definition of the clustering coefficient (as shown in Table 1). The five weighted definitions can degenerate into the unweighted clustering coefficient C_i , when the weights are replaced by the corresponding adjacency matrix elements. In addition, all these clustering coefficients equal to 0 if there are no links between the neighbors of node i .

Here, we constructed a weighted correlation network deriving from glioblastoma (GBM) microarray data. The gene expression data including 117 GBM samples were downloaded from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>), the pre-processing of microarray data were performed by using the package affy

Table 1 Definition and application for six different clustering coefficients

Author and Reference	Definition ^a	Application
Watts et al. [23]	$C_i = \frac{\sum_{jk} a_{ij} a_{jk} a_{ki}}{k_i(k_i-1)}$	Friendship networks
Zhang et al. [20]	$C_{w,i}^z = \frac{\sum_{jk} w_{ij} w_{jk} w_{ki}}{\left(\sum_j w_{ij}\right)^2 - \sum_j w_{ij}^2}$	Gene correlation networks
Lopez et al. [24]	$C_{w,i}^L = \frac{\sum_{j,k \in N(i)} w_{jk}}{k_i(k_i-1)}$	Social networks
Onnela et al. [25]	$C_{w,i}^O = \frac{\sum_{jk} (w_{ij} w_{jk} w_{ki})^{1/3}}{k_i(k_i-1)}$	Financial and metabolic networks
Barrat et al. [14]	$C_{w,i}^B = \frac{1}{s_i(k_i-1)} \sum_{jk} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{jk} a_{ki}$	Airport and scientific collaboration networks
Serrano et al. [26]	$C_{w,i}^S = \frac{1}{s_i^2(1-Y_i)} \sum_{jk} w_{ij} w_{ik} a_{jk}$	Airport, trade and scientific collaboration networks

^aThe subscript *w* indicates weighted; *N*(*i*) is the set of neighbors of node *i*

[19] under R environment). Due to computational limitation, the 7973 most varying genes were used in this study. Affecting by Zhang et al. [20], we defined the gene similarity weights as the power of absolute value of the Pearson correlation, namely:

$$w_{ij} = |cor(x_i, x_j)|^\beta \tag{1}$$

where x_i and x_j are the *i*-th and *j*-th gene expression profiles across 117 samples, β represents the power parameter whose value is chosen with satisfying certain criterion. In this paper, we only considered two criteria (i.e., scale-free and non-scale-free topology). For scale-free topology criterion, we set the parameter β equal to 7 (using the pickSoftThreshold function in the package WGCNA [10]) that led to the network satisfying scale-free property. For non-scale-free topology criterion, the parameter β was given the value 1.

A large w_{ij} indicates that genes *i* and *j* are highly correlated. By the definition, the 7973×7973 weights matrix *W*, with $w_{ij} = w_{ji} \in [0, 1]$ and $w_{ii} = 0$, encoded the strength of connection between pairs of genes. To filter the significant relationships, we considered two different situations: one was that we supposed there is always a link between any two genes (i.e., a complete connected network), the other was that whether there is a link depends on the *P*-value of the correlation coefficient. As for the latter, *P*-values were used to determine the binary adjacency matrix *A*, where $a_{ij} = 1$ (i.e., existing a link between genes *i* and *j*) if *P*-value ≤ 0.05 (corresponding to correlation ≥ 0.182), 0 otherwise ($a_{ii} = 0 \forall i$). In addition, the weight w_{ij} replaces with the value 0 when there is no link between genes *i* and *j* (i.e., $a_{ij} = 0$) for consistency.

Aware of the fact that there is a hierarchy of genes in living cells, we consequently compared the effect of different sets of genes on the distributions of strength and clustering coefficients. Therefore, we selected two sets of genes in this paper: one was the 101 hub genes identified by Rambaldi et al. [21] within the human cancer PIN, the other was the 2064 housekeeping genes identified by Chang et al. [22]. 78 and 1370 genes were among the 7973 most varying genes respectively, which were used in subsequent comparison analysis.

To calculate the clustering coefficients, matrix multiplication is needed. There is an pseudocode on GPUs for calculating the product of 7000×7000 matrix and 7000×1 vector as follows (The speed up of this GPU code is 70 times than CPU):

```
x ← blockIdx.x * blockDim.x + threadIdx.x
y ← blockIdx.y
shared_result[threadIdx.x] ← 0;
Call __syncthreads()
If x < 7000 Then
shared_result[threadIdx.x] ← global_matrix1[x] × glob-
al_matrix2[y×7000+x]
End If
Call __syncthreads()
For i = 256 And i > 0 Step i >> 1 Do
  If threadIdx.x < i Then
    shared_array[threadIdx.x] ← shared_array[threadIdx.x]
+ shared_array[threadIdx.x+i]
  End if
  Call __syncthreads()
End For
If threadIdx.x=0 Then
  Call atomicAdd(&result[y], shared_array[threadIdx.x])
End if
```

3 Results and Discussions

The data processing of data samples and construction of weighted correlation network are shown in Fig. 1. A previous study by Kalna et al. [27] compared the strength and weighted clustering coefficient distributions between normal and tumor networks, which was quite different from ours. We only focused on the tumor network (i.e., GBM gene network) and considered six different clustering coefficients. Besides, we compared three sets of genes: (1) all 7973 genes; (2) 78 hub genes; (3) 1370 housekeeping genes for four different network topologies: (1) Scale-free; (2) Non-scale-free; (3) Scale-free and complete-connected; (4) Non-scale-free and complete-connected (see details in Methods).

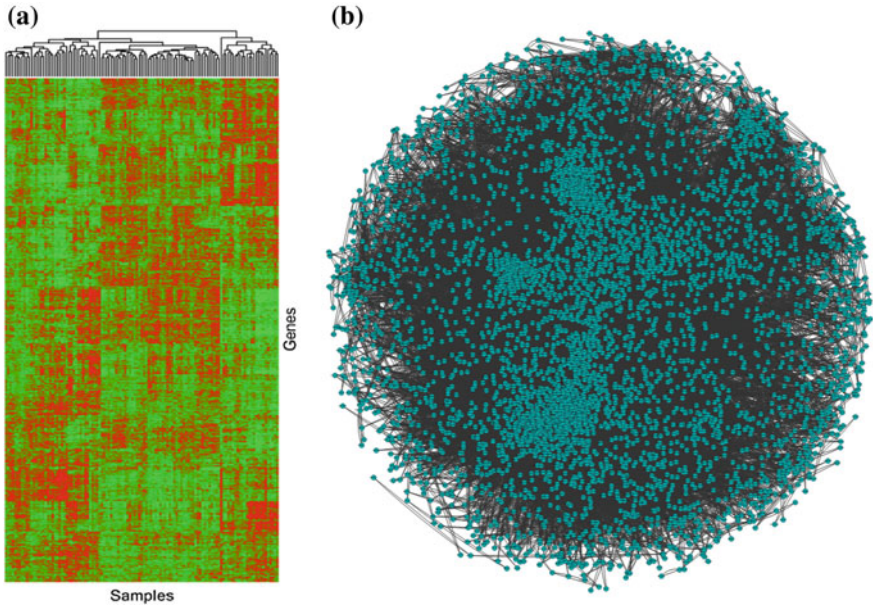


Fig. 1 The data processing of GBM samples and construction of weighted correlation network. **a** The gene expression profiles of 117 GBM samples. **b** The scale-free and non-complete-connected network deriving from 117 GBM expression data and existing a link between two genes if their correlation coefficient P -value ≤ 0.05

Figure 2 shows the rescaled strength $s/\langle s \rangle$ distributions of three sets of genes for four different network topologies. For either the scale-free [panels (1) and (3)] or non-scale-free cases [panels (2) and (4)], the distributions of the rescaled strength in the same set of genes almost coincide, regardless of whether the network is completely connected or not. The four panels have a common characteristic, the average strengths of sets of hubs and housekeeping genes are larger than that of all 7973 genes, which indicates the importance of housekeeping genes can be reflected in strength distribution. For instance, the average rescaled strengths $s/\langle s \rangle$ of hubs and housekeeping genes are 1.63 and 1.21 respectively in the scale-free cases [panels (1) and (3)], which are larger than that of all genes (average strength equals to 1). Since the hub genes themselves have higher degrees in PIN, this result implies there are consistency between PIN and weighted gene correlation networks. The former is a network at protein level and the latter is at genetic level.

Figure 3 shows the rescaled clustering coefficient $c/\langle c \rangle$ distributions of six different definitions for four different network topologies and three sets of genes. It is remarkable that all clustering coefficient distributions are independent of these three sets of genes. In many unweighted biological networks there is a power law between degree and clustering coefficient [28–31], this result implies that these

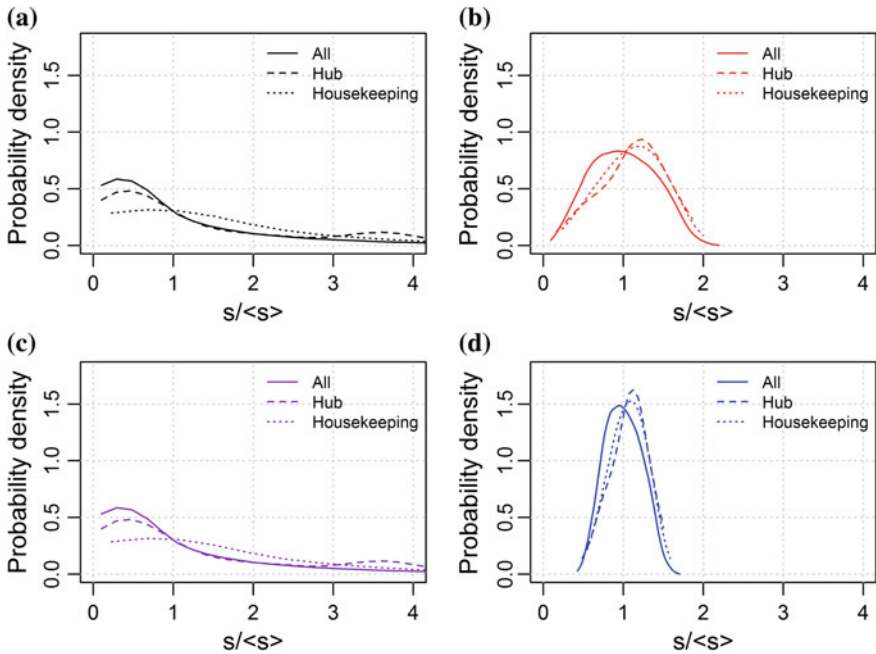


Fig. 2 Rescaled strength $s/\langle s \rangle$ probability distributions of three sets of genes—all 7973 genes, 78 hub genes and 1370 housekeeping genes for four different network topologies: **a** Scale-free; **b** Non-scale-free; **c** Scale-free and complete-connected; **d** Non-scale-free and complete-connected. Note that $\langle s \rangle$ represents the average strength of all 7973 genes and this value remains the same in terms of hub or housekeeping genes (similarly in Fig. 3)

important genes behave differently with the usual hub nodes as they have higher strengths. In scale-free networks (1st and 3rd rows), the Zhang et al. and Onnela et al. rescaled clustering coefficients, which have been used in biological network before, follow the heavy-tailed distributions. Based on this characteristic, we can determine whether a network have scale-free property or not. Except in non-scale-free (2nd row), these two clustering coefficients are quite different with other coefficients previously used for social networks. The profiles of other four coefficients have sharp peaks in both scale-free and non-scale-free networks (1st and 2nd row). For complete-connected networks (3rd and 4th rows), the unweighted, Barrat et al. and Serrano et al. clustering coefficients take the value 1 according to their definitions, and the Lopez et al. clustering coefficient stay almost constant level for large-size networks. In addition, the Barrat et al. and Serrano et al. clustering coefficients follow very similar trend in scale-free and non-complete-connected networks (1st rows). We conjecture that this is due to the fact that these two definitions are similar that used only weights of links adjacent to node i (Table 1).

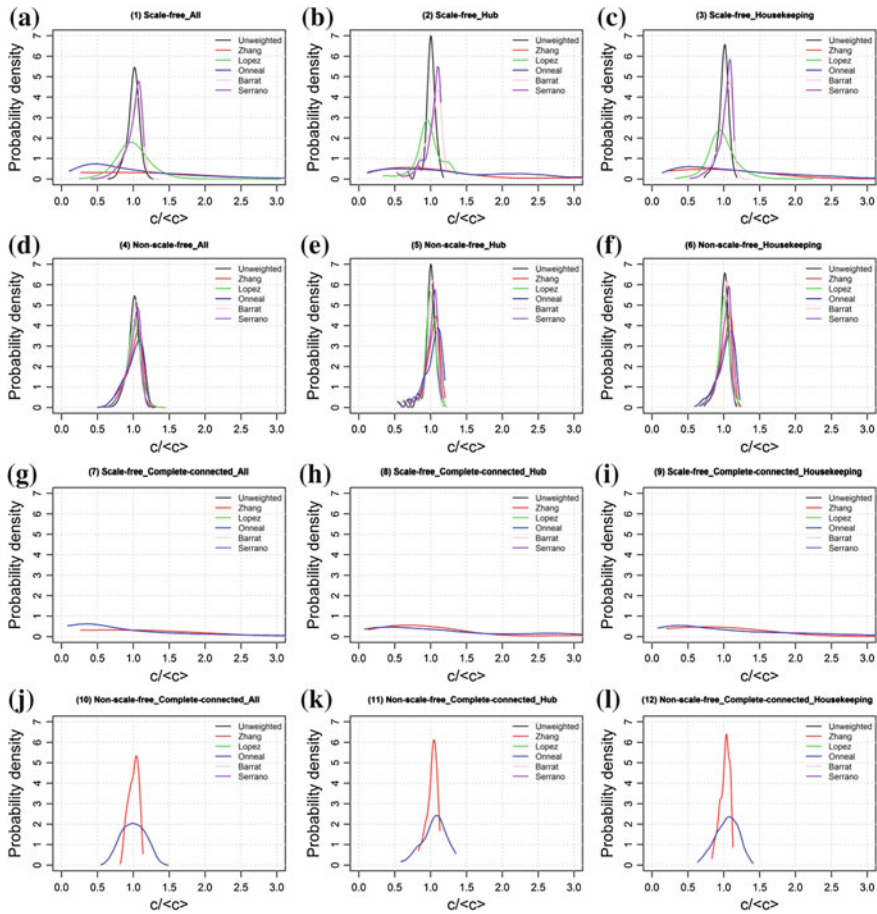


Fig. 3 Rescaled clustering coefficient ($c/\langle c \rangle$) probability distributions of unweighted and different weighted definitions for four different network topologies and three sets of genes: **a** Scale-free and all 7973 genes; **b** Scale-free and 78 hub genes; **c** Scale-free and 1370 housekeeping genes; **d** Non-scale-free and all 7973 genes; **e** Non-scale-free and 78 hub genes; **f** Non-scale-free and 1370 housekeeping genes; **g** Scale-free, complete-connected and all 7973 genes; **h** Scale-free, complete-connected and 78 hub genes; **i** Scale-free, complete-connected and 1370 housekeeping genes; **j** Non-scale-free, complete-connected and all 7973 genes; **k** Non-scale-free, complete-connected and 78 hub genes; **l** Non-scale-free, complete-connected and 1370 housekeeping genes. Note that in panel (g)–(l), the unweighted, Barrat and Serrano clustering coefficients were constant 1 and the Lopez et al. clustering coefficient stay almost constant level (not shown)

Table 1 Definition and application for six different clustering coefficients

Author and Reference	Definition ^a	Application
Watts et al. [23]	$C_i = \frac{\sum_{jk} a_{ij} a_{jk} a_{ki}}{k_i(k_i-1)}$	Friendship networks
Zhang et al. [20]	$C_{w,i}^z = \frac{\sum_{jk} w_{ij} w_{jk} w_{ki}}{\left(\sum_j w_{ij}\right)^2 - \sum_j w_{ij}^2}$	Gene correlation networks
Lopez et al. [24]	$C_{w,i}^L = \frac{\sum_{jk \in N(i)} w_{jk}}{k_i(k_i-1)}$	Social networks
Onnela et al. [25]	$C_{w,i}^O = \frac{\sum_{jk} (w_{ij} w_{jk} w_{ki})^{1/3}}{k_i(k_i-1)}$	Financial and metabolic networks
Barrat et al. [14]	$C_{w,i}^B = \frac{1}{s_i(k_i-1)} \sum_{jk} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{jk} a_{ki}$	Airport and scientific collaboration networks
Serrano et al. [26]	$C_{w,i}^S = \frac{1}{s_i^2(1-Y_i)} \sum_{jk} w_{ij} w_{ik} a_{jk}$	Airport, trade and scientific collaboration networks

^aThe subscript w indicates weighted; $N(i)$ is the set of neighbors of node i

4 Conclusions

In this paper we constructed a weighted gene correlation network for GBM using 117 samples on GPUs. The sets of housekeeping genes and hubs in PIN were introduced and we found that the distributions of strength of these two sets of important genes are different from the distribution of the entire network. We also compared the calculation results of six definitions of clustering coefficients in this weighted network. It was found that Zhang et al. and Onnela et al. coefficients have similar profiles in scale-free network and are different from other four coefficients. Moreover the distributions of clustering coefficients of housekeeping genes and PIN hubs are as similar as the distribution of the whole network, which disagrees with general understanding in unweighted biological networks. In conclusion, we investigated the different definitions of clustering coefficients in the weighted biological network and testified some sets of important gene have differences in the network indices, such as strength and clustering coefficient. As the weighted networks can reflect more characteristics of biological processes than the unweighted networks, more indices have to be well defined, where by more analyses (e.g., module identification or dynamics modeling) can be better performed. Besides the biology studies, weighted network is also useful for investigations of granular materials, automation and social problem. This study will be helpful for exploring how to construct the weighted network and implementing the biology computing algorithm on GPUs.

Acknowledgements This work is supported by the National Magnetic Confinement Fusion Science Program of China (Grant No. 2014GB104002), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA03030100) and CAS 125 Informatization Project (No. XXH12503-02-03-2). Thanks to Dr. Jun-Tu Sun for helpful advice on this work.

References

1. Elsen E et al (2007) N-body simulations on GPUs. arXiv preprint [arXiv:0706.3060](https://arxiv.org/abs/0706.3060)
2. Zhang Y et al (2005) Genome-scale computational approaches to memory-intensive applications in systems biology. In: Proceedings of the ACM/IEEE SC 2005 conference on supercomputing, 2005, IEEE
3. Dematte L, Prandi D (2010) GPU computing for systems biology. *Briefings Bioinform* 11 (3):323–333
4. Vigelius M, Lane A, Meyer B (2011) Accelerating reaction-diffusion simulations with general-purpose graphics processing units. *Bioinformatics* 27(2):288–290
5. Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8 (6):450–461
6. Zak DE, Aderem A (2009) Systems biology of innate immunity. *Immunol Rev* 227:264–282
7. Huang W et al (2012) Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator. *Science* 336(6077):75–79
8. Miller GA et al (2007) Clustering coefficients of protein-protein interaction networks. *Phys Rev E* 75(5):051910
9. Patil KR, Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* 102(8):2685–2689
10. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(1):559
11. Bernard A et al (2012) Transcriptional architecture of the primate neocortex. *Neuron* 73 (6):1083–1099
12. Vauleon E et al (2012) Immune genes are associated with human glioblastoma pathology and patient survival. *BMC Med Genomics* 5:41
13. Wang L et al (2009) Gene networks and microRNAs implicated in aggressive prostate cancer. *Cancer Res* 69(24):9490–9497
14. Barrat A et al (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci USA* 101(11):3747–3752
15. Kumpula JM et al (2007) Emergence of communities in weighted networks. *Phys Rev Lett* 99 (22):228701
16. Antoniou IE, Tsompa ET (2008) Statistical analysis of weighted networks. *Discrete Dyn Nat Soc*
17. Mirzasoleiman B et al (2011) Cascaded failures in weighted networks. *Phys Rev E* 84 (4):046114
18. Saramaki J et al (2007) Generalizations of the clustering coefficient to weighted complex networks. *Phys Rev E* 75(2):027105
19. Gautier L et al (2004) affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20(3):307–315
20. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4(1)
21. Rambaldi D et al (2008) Low duplicability and network fragility of cancer genes. *Trends Genet* 24(9):427–430
22. Chang C-W et al (2011) Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. *PLoS ONE* 6(7):e22859
23. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393 (6684):440–442
24. Lopez-Fernandez L, Robles G, Gonzalez-Barahona JM (2004) Applying social network analysis to the information in CVS repositories. In: International workshop on mining software repositories, IET
25. Onnela J-P et al (2005) Intensity and coherence of motifs in weighted complex networks. *Phys Rev E* 71(6):065103

26. Serrano MÁ, Boguñá M, Pastor-Satorras R (2006) Correlations in weighted networks. *Phys Rev E* 74(5):055101
27. Kalna G, Higham DJ (2006) Clustering coefficients for weighted networks. In: Symposium on network analysis in natural sciences and engineering
28. Albert R (2005) Scale-free networks in cell biology. *J Cell Sci* 118(21):4947–4957
29. Fu F, Liu LH, Wang L (2008) Empirical analysis of online social networks in the age of Web 2.0. *Phys A Stat Mech Appl* 387(2–3):675–684
30. Potapov AP et al (2005) Topology of mammalian transcription networks. *Genome Inform Ser* 16(2):270
31. Yellaboina S, Goyal K, Mande SC (2007) Inferring genome-wide functional linkages in *E-coli* by combining improved genome context methods: comparison with high-throughput experimental data. *Genome Res* 17(4):527–535

Design of Scalable Control Plane via Multiple Controllers

Wenbo Chen, Xining Tian and Zhihao Shang

Abstract Controllers is responsible for the entire network centralized control in SDN (soft-defined network). It attaches great importance to grasping the entire network resources view and improving the quality of network resources delivery. However, centralized controllers call for more responsibility in control apparatus, making the expansibility of the controller plane the key issues of the SDN. This article designs an extensible SDN controller layer. With the use of existing cluster management technology, it avoids single point failure of the controllers and gives full play to every controller. Every controller is able to effectively allocate network resources according to the entire network information, so that the entire network resources can be effectively scheduled.

Keywords SDN · Openflow · Control plane · Multiple controller

1 Introduction

The network has become one of the most important infrastructures in modern society development and technological advance. It deeply changes the mode of production and the way people live and study. However, the network remains stagnant due to its complexity. The increase of the network size and the network application make the traditional network framework difficult to satisfy the need of

W. Chen · X. Tian (✉) · Z. Shang (✉)
School of Information Science and Engineering, Lanzhou University,
Lanzhou 730000, China
e-mail: tianxn13@lzu.edu.cn

Z. Shang
e-mail: shangzh11@lzu.edu.cn

W. Chen
e-mail: chenweb@lzu.edu.cn

the enterprises, operators and the users. It is especially apparent in the age of cloud computing when the multi-tenant and the virtual machine migration of the data center call for high demand of network virtualization. On the other hand, it is hard for the network researchers to realize their innovative experiment of the existing network in the real network with the protocol and network equipment in hand. The 4D framework [1], Ethane's [2] network control and forwarding separation structure, etc. are put forward in succession to solve these problems.

The SDN technology, as represented by OpenFlow, is divided into data plane software and control plane software. The data plane software is responsible for data forwarding, while the control plane software is responsible for forwarding the corresponding strategy [3]. Such network control and forwarding separation structure makes the control layer get rid of the dependence on network devices by providing flexible and convenient programmability. By using centralized control, SDN is easy to achieve the pool of resources, dynamic on-demand scheduling apply and a better elastic expansion. In the meantime, its open API, i.e. the southbound and northbound interface, can give rise to the industrial chain and promote the rapid development of the industry.

The controller is of great importance in the SDN framework. It plays crucial part in mastering the entire network resources and improve the network resources delivery. However, the centralized control ability also implies that the security and performance of controller board also become a possible bottleneck. Once the controller has no guarantee on performance or security, the service ability of the whole network would be degraded, or even paralyzed.

The controller is of great importance in the SDN framework. It plays crucial part in mastering the entire network resources and improve the network link delivery. However, the centralized control ability also implies that the security and performance of controller board also become a possible bottleneck. With the transition from concept to practice of the SDN technology in the last two years, multi-controllers become one of the core issues of achieving its industrial deployment [4–6]. Only all controller share one completely consistent entire network topology can the logic decision's validity be in control. Second is expansibility. One of the reasons to bring about multi-controller is the need of central control towards expansion [7, 8]. Thus, expansibility is the prime goal. It should be guaranteed that a big enough scale can be supported, and that new controllers can be dynamically added while not influencing the existing network. Fourth is the mechanism of controller failure. When one controller fails, other controller can smoothly take over its switch. Last is to make every controller exert its maximum performance as possible while not creating bottlenecks.

This article puts forward a multi-controller framework, using Zookeeper to manage the cluster. Ryu is used as the controller, while OVS (OpenVswitch) works as the switch of OpenFlow. The controller can be dynamically added, forwarding the entire network topology information to Zookeeper for unified management. Distributed mutual exclusion election algorithms is achieved by using the distributed system lock mode, making controller failure coping mechanism come true.

2 Related Works

HyperFlow [9] is the first proposed OpenFlow multi-controller implementation in 2010. As one of NOX's applications, it is relatively easier to implement. HyperFlow is designed on the basis of distributed file system WheelFS. Network event in different controllers is implemented by file updating. Onix is put forward by Google, NEC and Nicira. It is a distributed SDN deployment solution for large scale network. Onix provides a network information base (NIB) to maintain the network global state. The design is to maintain NIB distribution mechanism, so that the consistency of the entire network state information is guaranteed. Literature [10] come up with a distributed control system, similar to Master/Slaves, to realize the scalability of data center network. Any of the existing controllers can be used, and JGroups are used for communication. In the beginning, a Master controller is selected to maintain global Controller-Switch map. The Master controller is monitored by other controllers. Once there exists any abnormal situation, another node is selected for replacement to avoid single point of failure of the Master. Such replacement cannot be detected in Switch's point of view. The ASIC, as is put forward in Literature [11], uses relatively mature existing technology. Therefore, it is relatively easier to deploy and implement. Issues such as consistency is implemented by MySQL, MemCached and so on and so forth software. ASIC uses technology such as load balancing, parallel computing, data sharing and cluster for implementation.

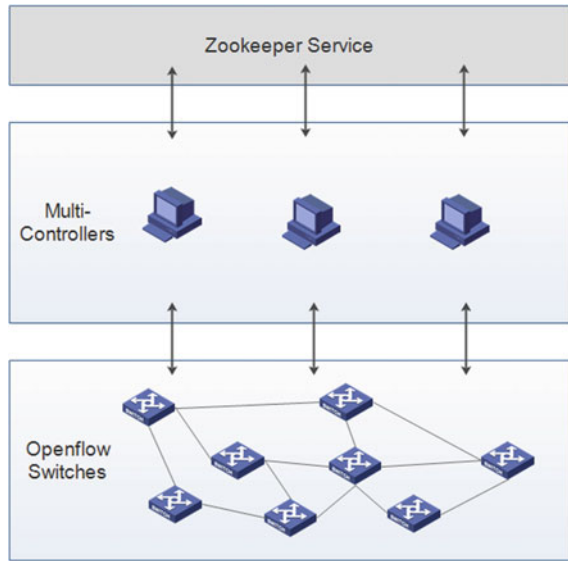
Kandoo is a multi-controller control plane design scheme based on hierarchical thought, proposed in Literature [12]. Such control scheme divides the controller into two types: root controller and local controller. The operation of the local controller only requires the application of local information, while the root controller needs the application of the entire network information. The local controller is more like adding a control agent on traditional single controller on the switch level. Root controller, on the other hand, plays a traditional controller role. Literature [13] put forward DISCO, which is specific to inter domain interaction design in wide area network. In DISCO, every controller controls one region. Controllers use AMQP protocol to communicate, and agent is used for the interaction of the aggregation routing information. On the contrary to other multi-controllers framework, DISCO strictly distinguishes the inter domain connection of heterogeneous. For instance, the MPLS tunnel and the SATCOM link of the low bandwidth high latency link.

3 Proposed Framework

3.1 Overview

Many sophisticated solution have been brought about on cluster management technology so far. This article uses Zookeeper to manage controller cluster, while

Fig. 1 The architecture of the framework



using Ryu as its controller. Zookeeper owns complete node monitoring mechanism and event trigger mechanism, and is able to detect node rapidly and perceive node change under distributed environment. The distributed coordination characteristics of Zookeeper is used to manage controller cluster in this article.

Zookeeper is able to provide data storage in a way which is similar to directory node tree of the file system. However, Zookeeper is not specifically designed for data storage. Its main function is to maintain and monitor the change of state of the stored data. The cluster management of data is achieved through monitoring the change of the stored data.

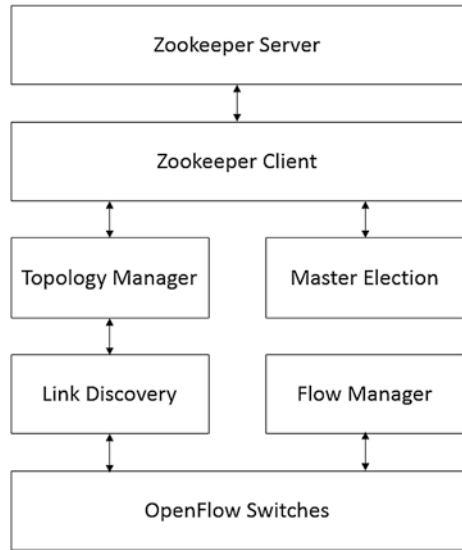
A single Zookeeper Server is able to provide the distributed management service. The Zookeeper cluster can also be used to improve the robustness. The demand of the operating procedure of Zookeeper Server on computing and storage resources is not high. A single Zookeeper Server is able to manage several nodes.

In the framework put forward in this article (see Fig. 1), every OpenFlow switch has multi-controllers, but only one controller can distribute flow table. This controller is the Master Controller of the switch. Other controllers that are connected to the switch can only query flow table. Information interaction and cluster management among multiple switches is done by Zookeeper. The controller obtain the entire network topology information by visiting the node file of the Zookeeper Server.

Figure 2 shows the four modules in the controller framework. Link Discovery module is responsible for link discovery and link information storage. Topology Manager converges link information and manage the entire topology information. The main controller of the switch is selected by Master Election.

Different from some of the existing multi-controller framework (The multi-controller shows its external performance as a whole. It uses master/backup mode on the inside, i.e. one Master Controller and others as its backups.), every controller is in a

Fig. 2 The modules in the framework



working state in the implementation of this article. That is to say, all controllers on the controller layer is equal. But for the switches, each switch will have to select one controller as its Master Controller among its many multi-controllers.

Every OpenFlow switch connects 3 controllers. The controllers and the switches can perceive the connection between each other. Conflict will occur if multiple controllers issue instruction simultaneously, so that an election system is required to make sure which controller will be the actual controller of the switch, i.e. the Master Controller of the Switch.

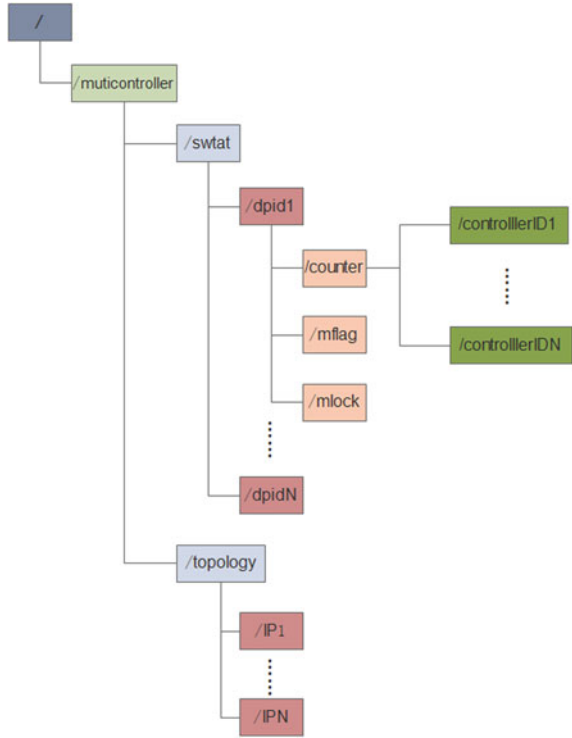
Controllers which connect to the same switch will determine which controller would be the Master Controller through election in the beginning. After the election, the controller would always be the Master Controller until its failure (e.g. The controller downs). Election coping with Master Controller failure: Other controllers connected to the Master Controller replace its place once the Master Controller of the switch fails. Election of the new Master Controller is conducted at once among the controllers. The heartbeat mechanism is used to screen failures. The winner of the election will be the Master Controller of the switch until it fails.

3.2 Implementation Details

3.2.1 Zookeeper File System

Znode *swtat* and its child node is used to complete every switch's controller election. Different switches use znode *dpid* as its unique identification, and uses znode *counter* and its child node *controllerID* as the identification of the different

Fig. 3 The Zookeeper file system



controllers connecting to the switch. Znode *mflag* is used to judge whether there exists the flag bit of the Master Controller. Znode *mlock* is used for multiple controller election (Fig. 3).

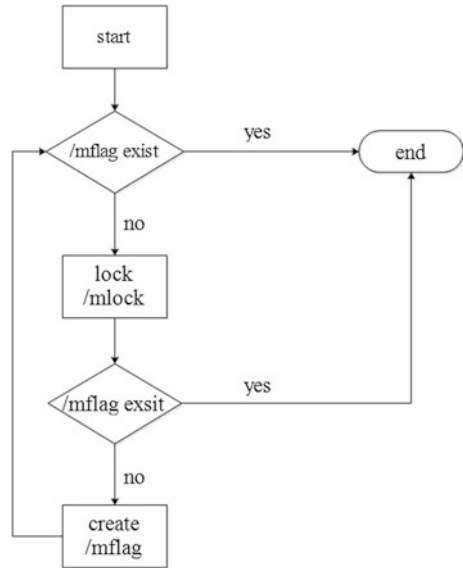
Znode *topology* is used to store the entire topology information. Its child node *IP* identifies different controllers. Znode *IP* stores equipment connection information in JSON format.

3.2.2 Election Mechanism

When a switch and its Master Controller disconnect, the node file of the *mflag* is deleted. All the controllers connects to it participate in the election. The election algorithm is shown in Fig. 4.

Every controller goes to lock znode *mlock*. If the *mflag* is found null after locking, node *mflag* will be created. If there exists *mflag*, the controller will exit the election. Otherwise, the controller will unlock node *mflag* and exit the election. In this way, the first controller to form *mflag* node will succeed in the election and become the Master Controller of the switch.

Fig. 4 The procedure of election



3.2.3 Link Discovery

The controller send packet-out message to the switch periodically, so that the switch packet-out its dpid (Datapath ID) to the entire network. The switch send packet-in message to the controller after receiving dpid from other switches. Then the controller store the link immediately. The controller store its link into Zookeeper periodically.

3.2.4 Entire Network Topology Information Acquisition

The controller can visit znode *topology* to get information of the entire network topology. IP address is used as the controller’s identification, so that different child node *IP* represent link information in different controllers. Take JSON format storage as an example: *{srcdpid: 1, dst: {dpid: 2, port: 3}, time: 1433071083}*. The controller can get the entire network topology information by visiting all the znode *IP*.

4 Analysis and Results

In this framework, the bottleneck of the system will appear in the interaction between the Zookeeper Server and the Ryu controller because the Zookeeper is used for cluster management. The interaction has 2 parts. One is controller election,

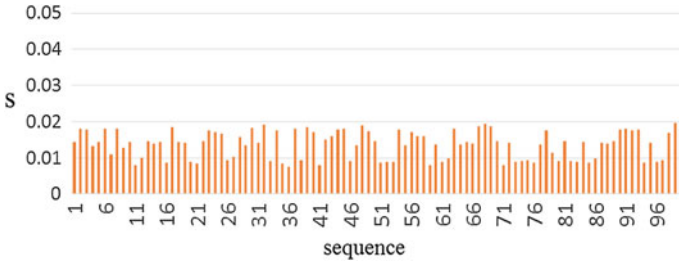


Fig. 5 Time of election

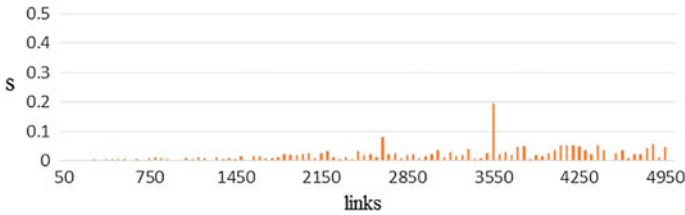


Fig. 6 Time of get topology

the other is entire network topology acquisition. We tested the time each interaction takes.

Figure 5 recorded how long the Master Controller Election will take in each 100 times. The abscissa is the number of test number, while the ordinate shows consumed time. It can be shown in the figure that all the delay is below 20 ms, and 30 % of them is below 10 ms.

Figure 6 records the time the controller consumed to acquire the entire network topology in a network which emulates different network link numbers. The abscissa is the network link number, and the ordinate is consumed time. It can be found that the consumed time for the controller to acquire topology increases when the network scale expands.

5 Conclusion

This article implements clustering cooperative work of multi-controllers, and makes full use of the advantage of Zookeeper to manage cluster controller. In the framework mentioned above, every switch connects many controllers. Under the circumstances when the Master Controllers fails, other controller can replace its role. In this way, single point failure caused by single controller is avoided. Meanwhile, compared to earlier single controller which controls the whole framework of the OpenFlow network, the cooperation of the multi-controller

effectively lightens the load of each controller. Due to the use of Zookeeper for cluster management, this framework has good expansibility, helping the cluster controller achieve dynamic smoothing add/delete operation. Distributed mutual exclusion election algorithms is achieved by using the distributed system lock mode, making controller failure coping mechanism come true. If the controller downs, other controllers can replace the failed controller and take over its switch. The entire network topology information is uniformly stored in Zookeeper. The controllers only has to go through znode to get the entire network topology. Lastly, the article analyzes the performance bottleneck of this framework, and measures the election time and the time needed for the controller to acquire the entire network topology information.

References

1. Greenberg A, Hjalmtysson G, Maltz DA et al (2005) A clean slate 4D approach to network control and management. *ACM SIGCOMM Comput Commun Rev* 35(5):41–54
2. Ethane: Taking control of the enterprise (2007)
3. Mckeown N, Anderson T, Balakrishnan H et al (2008) OpenFlow: enabling innovation in campus networks
4. Sezer S, Scott-Hayward S, Chouhan PK, Fraser B, Lake D, Finnegan J, ..., Rao N (2013) Are we ready for SDN? Implementation challenges for software-defined networks. *Communications Magazine, IEEE*, 51(7):36–43
5. Yeganeh SH, Tootoonchian A, Ganjali Y (2013) On scalability of software-defined networking. *Commun Mag IEEE* 51(2):136–141
6. Schmid S, Suomela J (2013) Exploiting locality in distributed SDN control. In: *Proceedings of the second ACM SIGCOMM workshop on hot topics in software defined networking*. ACM, pp 121–126
7. Banikazemi M, Olshefski D, Shaikh A, Tracey J, Wang G (2013) Meridian: an SDN platform for cloud network services. *Commun Mag IEEE* 51(2):120–127
8. Jain R, Paul S (2013) Network virtualization and software defined networking for cloud computing: a survey. *Commun Mag IEEE* 51(11):24–31
9. Tootoonchian A, Ganjali Y (2010) HyperFlow: a distributed control plane for OpenFlow. In: *Proceedings of the 2010 internet network management conference on research on enterprise networking*. USENIX Association, p 3
10. Yazici V, Sunay MO, Ercan AO (2014) Controlling a software-defined network via distributed controllers. arXiv preprint [arXiv:1401.7651](https://arxiv.org/abs/1401.7651)
11. Lin P, Bi J, Hu H (2012) Asic: an architecture for scalable intra-domain control in openflow. In: *Proceedings of the 7th international conference on future internet technologies*. ACM, pp 21–26
12. Hassas Yeganeh S, Ganjali Y (2012) Kandoo: a framework for efficient and scalable offloading of control applications. In: *Proceedings of the first workshop on hot topics in software defined networks*. ACM, pp. 19–24
13. Phemius K, Bouet M, Leguay J (2014) Disco: distributed multi-domain SDN controllers. In: *Network operations and management symposium (NOMS)*, IEEE, pp 1–4

Research on Learning Record Tracking System Based on Experience API

Xinghua Sun, Yongfei Ye, Li Hao, Zexin An and Xiaoyu Wang

Abstract It was an important thing of record and track learning behavior in the learning process. On account of the complexity of the structure and simplicity of the data transmission, the SCORM make it unable to obtain complete learning record. Under the analysis of the network learning model and relevant semantic elements, this paper presents learning record model and learning record tracking System architecture based on Experience API. The method of learning record be transferred from LMS to Learning Record System is described in detail. And this paper also gives the method of reading the learning record from the LRS which can achieve multidimensional analysis for cross environment learning record, and that will lead to a better support for mobile learning and individualized learning.

Keywords SCORM · Experience API · Learning record · LRS · TLA

1 Introduction

At present, it becomes a trend of learner behavior networking. Tracking learner behavior and recording learning data contribute to develop educational resources, to support teachers to organize and improve curriculum design, to make an effective assessment of learning resources and learners. Network learning behavior includes network resource browsing, online information retrieval, network information

X. Sun (✉) · Y. Ye · L. Hao · X. Wang
School of Information Science and Engineering, Hebei North University,
Zhangjiakou, Hebei, China
e-mail: sunxinghua08@gmail.com

Y. Ye
e-mail: yeyongfei005@126.com

Z. An
First Affiliated Hospital, Hebei North University, Zhangjiakou, Hebei, China

processing, network knowledge management, interactive network, network communication, web collaboration, knowledge generation network, network cooperation and self-reflection and monitoring, etc. Standardized learning behavior records tracking system has become hot research topic.

2 Overview of the Experience API

While the SCORM was successful in meeting the high-level requirements to solve the challenges within Web-based training systems, it was created prior to the widespread use of other learning environments and platforms such as mobile device, intelligent tutoring systems, virtual worlds, games, and other social networking tools that augment the performance of today's learner beyond formal training situations. Further, SCORM content was designed to be accessed and tracked via a learning management system (LMS). Today, Online learning provides more opportunities to capture more than just a learner's assessment score or course completion status. Learners, as well as education and training practitioners, expect new types of learning data to be captured and used within the aforementioned learning environments, in order to provide a personalized learning experience. They expect to learn informally and collaboratively, and to be able to use social networks as part of the learning experience. Users expect that their learning experiences will earn them credit, regardless of whether the learning activities are browser-based or not; thus there is a new requirement to free learners from being tied to an LMS. The Experience API gives learners, instructional developers, and instructors the opportunity to track and access data that far exceeds the current capabilities afforded by the SCORM.

The data from social networks such as Facebook or Twitter are delivered in the form of "streams" that can be widely applied and syndicated in many contexts. The Experience is an integrated approach to generate and capture learning stream data and then organizes the data into meaningful learning contexts. It is an interoperable way to encapsulate and exchange learning data through the use of a learning-based activity stream. These activity stream data include defined actors, verbs, and activities associated with the learning experience so that the data exchanged maintain contextual meaning.

A simple example of an activity stream that relates to traditional web-based learning is: "I (actor) completed (verb) the Information Assurance course (activity)." An example that reflects a modern informal learning scenario could be "Jill (actor) posted (verb) to the Project Management course student forum (activity)." It is expected that learning communities will develop standard options for each of the three elements to address their own domain-specific requirements (e.g., the medical community, government, higher education).

Another element of the Experience API is the Learning Record Store (LRS), which is the experience tracking and storage component of ADL's service-based approach and vision toward the Training and Learning Architecture (TLA).

The platform-independent LRS design allows flexibility in that it may be a stand-alone service or a complementary component of a traditional LMS. A goal of the TLA is to ensure that past investments in SCORM can be maintained, while offering the benefits of platform neutrality, intermittent or disconnected network scenarios, and the capability to move learning out of the desktop browser.

The LRS will leverage additional services for content brokering, user profiles, and competency networks to build a customized suite of TLA services. In addition, the LRS allows authorized systems to retrieve previously recorded activity stream statements, which enables the development of advanced third-party reporting and data analytics tools. The Experience API also moves beyond the single-learner approach, allowing for team-based exercises, collaboration, and direct instructor intervention. This enables group learning, informal learning, and social learning-on any device or platform.

3 Learning Record Model Based on Experience API

3.1 Learning Record Model Based on Experience API

The learning record model based on Experience API can be seen in Fig. 1.

The learners enter in the Internet and log in webpage, LMS or applications through user authentication. The E-learning resources generally include online courses, articles, web pages, serious games, etc. The learners browse the online

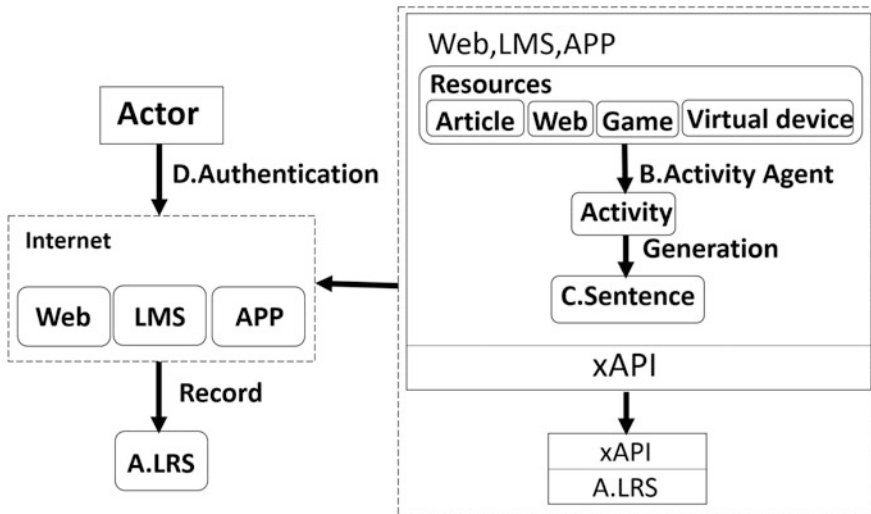


Fig. 1 Learning record model of experience API

learning resources for learning experience, which is transmitted to LRS through the API Experience protocol and specification. Specifically, the transmission process is: activity provider defines the learning activities engaged by learners and divides activities in different groups according to modules; activity generates statement, which is stored in LRS through activity generation statement API. Experience API consists of four interfaces, which are respectively Statement API, State API, Activity Profile API and Agent Profile API. Statement API is responsible for the storage and removal of statement in LRS; State API saves the activities in use to buffer cache; Activity Profile API can refer to the complete description of activities stored in LRS; Agent Profile API adds data related to the agent to the LRS.

3.2 Learning Record Information Interaction Process

The learners obtain learning experience through logging in webpage, LMS, applications or other learning terminals, and then the learning record information interacts with LRS to complete the information storage or extraction function. The specific process is: learners enter in the web pages, courseware in LMS or applications to learn for learning experience, which is then transferred to activity by the system, and the activity generates statement. Statement storages or extracts information through interaction with Statement API and LRS in Experience API. The learning record information between LRS and LMS is different from that between two LRS. In LMS, LRS only stores and acquires learning record, but the content packaging, release and output are all finished in LMS. The information data recorded in LRS can be sent through reporting tools between independent LRS or sent to LRS in LMS through the internal reporting tools of LMS.

4 Implementation of Learning Record Tracking System Based on Experience API

4.1 Learning Record Tracking System Architecture

With the combination of LMS and Experience API, the learning behaviors inside and outside (informal learning) can be recorded, so the integration of LRS into LMS can contribute to more complete functions of LMS. The single use of LMS can't achieve the tracking to learning records left by LMS outside learning. The addition of Experience API into LMS platform for architecture reconstruction can support the verb and activity semantic relations of Experience API, facilitate further data record analysis and Data mining based on Experience API and provide individualized learning experience for the learners. The LMS network system architecture based on Experience API is mainly of two types: firstly, LMS integrated model, i.e.

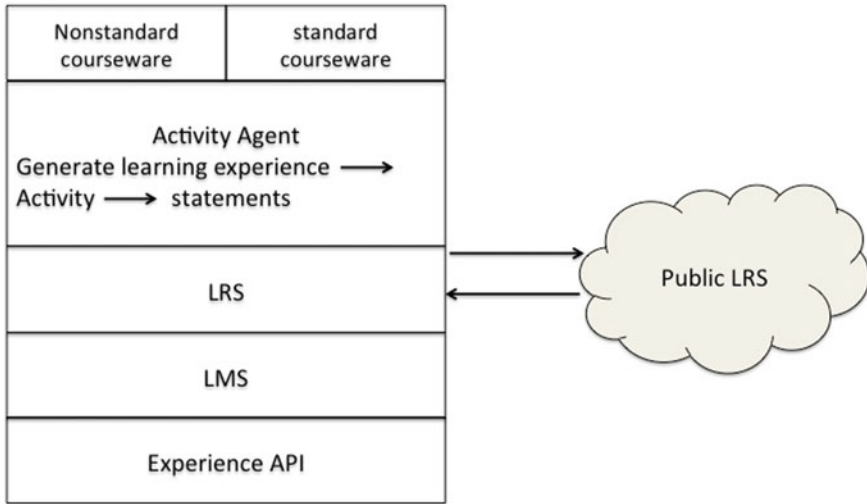


Fig. 2 Architecture of learning record tracking system

architecture reconstruction is made to the original LMS platform, including the reconstruction of resource integration model and that of platform integration model; secondly, plug-in model, i.e. architecture reconstruction is made to the carrying source based on webpage or application, to assist the collection of outside learning record by LMS platform. The system development of this research adopts the first architecture way—LMS platform integration reconstruction model. In this system, LRS, as the learning record storage system, is used for storing and querying learning records, but the content packaging, release and output are also completed inside the original LMS platform. The architecture reconstruction in the original LMS platform is to establish LRS learning record storage system and Experience API relevant mechanism inside the platform, and the LRS inside the system and Public LRS achieve synchronization of learning record. In this way, this system realized cloud storage of the learning record. The architecture figure is seen in Fig. 2.

4.2 How to Transfer Learning Record to LRS

After learners complete study of learning resources, the learning record tracking system will automatically generate statement and transfer it to LRS. Statement is the core of Experience API and has a very simple structure. It adopts the form of “Actor + Verb + Object” to describe a learning activity. All learning activities are described and stored in this form, for example “Xiaoming (Actor) completes (Verb) the stimulation exercise of CET-4 (Object)”. No matter what language is used for compilation, such description is stable and common. To ensure the distributed

feature of Experience API, statement is invariant in logic structure, but the activity content referred by the statement is changeable. Except of the structure of “Actor + Verb + Object” mentioned above, statement can also include situational information. This way ensures all functions of SCORM realized by Experience API and provides a more flexible structure. In one statement, the sequence of attributes is changeable, and the writing format of statement can be JSON or XML.

The procedures for the system to upload learning record to LRS are as follows:

1. Acquire the user name and password of LRS and get the permission;
2. LMS acquires the value of name, password and box(id).
3. Make get operation in Action and give response. Verb and Act can select language (en-US or en-UK, etc.) through LanguageMap; Verb can get verbs from <http://adlnet.gov/expapi/verbs/>.
4. Transfer objects into json nodes through to Json() method, and turn the string into json.
5. Transfer the contents of activity record into statement with proper format through SaveStatement() method and transfer it to LRS.

The core codes of the system uploading the learning record are as follows:

```
// actor
name = agent.getName();
box = agent.getMbox();
agent = new Agent();
agent.setName(name);
agent.setMbox(box);
// verb
verb = new Verb("http://adlnet.gov/expapi/verbs/" +
id);
idd = id;
verb.setDisplay(new LanguageMap());
verb.getDisplay().put("en-US", idd);
// object objectType:activity)
activity = new Activity();
actID = act;
activity.setDefinition(new ActivityDefinition());
activity.getDefinition().setType(new
URI("http://id.tincanapi.com/activitytype/unit-test"));
activity.getDefinition().setName(new LanguageMap());
// Language
activity.getDefinition().getName().put("en-US",
actID);
SaveStatement();
```

4.3 How to Read Learning Records from LRS

LRS allows authorized systems to retrieve previously recorded activity stream statements, which enables the development of advanced third-party reporting and data analytics tools. How we get learning record back from LRS. In true REST form, invoking the query call is a matter of issuing a GET on Experience API statements. The query API provides a number of parameters to filter the returned set of statements. This includes filtering on a statement's actor, verb, and object, but also includes ways to filter based on parts of the statement context, or limit the results with time boundaries. This allows for built in capability to ask things like "Who has completed 'Solo Hang Gliding' since yesterday?" or "What's all the activity happening in the context of 'Ping Pong' during last month?"

The core code of getting learning record from the LRS is shown below:

```
// retrieval statement;
StatementsResultLRSResponse lrsRes =
lrs.queryStatements(query);
    if (lrsRes.getSuccess()) {
        Sytem.out.println("success=====");
        // Take the subject
        lrsRes.getContent().getStatements().get(0).getActor().g
etName();
        // Take the predicate
        lrsRes.getContent().getStatements().get(0).getVerb().ge
tDisplay().get("en-US");
        // Take an object
        activity = new Activity();
        activity = (Activity)
lrsRes.getContent().getStatements().get(0).getObject();
        activity.getDefinition().getName().get("en-US");
        // Get the record content
        lrsRes.getContent();
```

Learning record display effect of Learning record is as shown in Fig. 3.



Fig. 3 Learning record display effect

5 Conclusion

Experience API is set of new technical specifications and its emergence will change the existing digital learning method. Firstly, it makes up the deficiencies of SCORM and relieves learning content from the browser. Secondly, it can track and record almost all types of learning activities.

This paper analyzes the learning record model based on Experience API. It proposes the learning record tracking system architecture based on Experience API and works out its implementation. This system integrates Experience API with LMS, reconstructs the architecture of original LMS, and displays the learning record to learners completely and accurately; the learners can track all learning records of their study inside or outside the platform with the reconstructed LMS, thus making up the deficiencies and shortcomings of SCORM standards. This system has now realized the writing and getting learning records in LRS of the TLA architecture, thus providing reference to the application of LMS platforms such as Moodle, etc. Next, it will carry out theoretical study and implementation to the learner profile, content agent and protocol as well as ability authentication, and gradually perfects the study of learning record tracking based on Experience API.

An EF6 Code-First Approach Using MVC Architecture Pattern for Watershed Data Download, Visualization and Analysis System Development Based on CUAHSI-HIS

Rui Gao, Yanyun Nian, Lu Chen and Qingguo Zhou

Abstract The main objective of this paper is to explore an information platform for sharing, managing, downloading, analyzing and visualizing of a diverse range of hydrologic observation data to support investigators, geotechnical experts do some research about watershed in Northwestern China. For this reason, we develop a Watershed Datacenter System (WDC) which adopts an Entity Framework 6 (EF6) approach based on Model-View-Controller (MVC) architecture pattern and several other useful technologies like cross-platform JavaScript libraries (jQuery, D3 and Dojo), ArcGIS API and Responsive web design. Besides, Observation Database Model (ODM), Web Services and Time-Series analysis tools are seamlessly integrated into our WDC with the help of open source HIS (Hydrologic Information System) by CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science, Inc.). The result shows that the WDC brings a lot of convenience for managing and analyzing of data onto watershed research.

Keywords Code-first · MVC · Watershed information system · Data sharing · CUAHSI-HIS

R. Gao · Q. Zhou (✉)
School of Information Science and Engineering, Lanzhou University,
Lanzhou 730000, Gansu, China
e-mail: zhouqg@lzu.edu.cn

R. Gao
e-mail: gaor13@lzu.edu.cn

Y. Nian · L. Chen
College of Earth and Environmental Sciences, Lanzhou University,
Lanzhou 730000, Gansu, China
e-mail: yynian@lzu.edu.cn

L. Chen
e-mail: chenlu14@lzu.edu.cn

1 Introduction

Water is the most essential resource for the sustainable development of human society and natural systems [1]. And watershed is a very important geographic unit, often plays a dominated role in the earth's environmental and resources management. Many researches are focusing on watershed which is the fundamental level of the earth system science. In the past 30 years, with the rapid development of the computer science and technology, remote sensing technology, Internet and information technology, watershed information has been developed rapidly. It's constantly pushing forward to the direction combing with digital and information technology.

However, there are some challenges in this research field when combing the hydrologic sciences and electronic information technology. The greatest challenge is how to store and classify a variety of data in the databases due to a wide range of watershed data, such as precipitation, evaporation, quality of water, soil moisture, data acquisition location information and so on. Another challenge is that how to share and manage observation data using automated system on the Internet. In the past, most observation data are usually encapsulated within files or databases which cannot easily be cataloged and shared. Now, many developed countries have begun to realize the importance of watershed information system, they establish some organizations to develop hydrologic information systems and have achieved good results. Such as USGS (the United States Geological Survey) releases water resource information for the USA on the Internet (<http://water.usgs.gov>) and provides rich and powerful hydrologic information service; The Canada government has also set up their own organization WSC to collect, interpret and disseminate and release standardized water resource data and information. Besides, the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) launches a hydrologic information system called CUAHSI-HIS, which provides a standard database schema (Observation Database Model, ODM) for use in the storage of point observations in a relational database. ODM is intended to allow for comprehensive analysis of information and to share data among investigators [2], which has become a very mature database schema through continuous improvement of HIS team has been widely applied in many parts of the world.

In China, the development of watershed information management and sharing mechanism is relatively slow and immature, which has not a perfect technical standard, suitable information sharing platform [3]. Most of these applications and systems for water resources are based on C/S structure, designed to work only on one computer and lack of multi-user support. The purpose of this paper is to explore a platform for sharing, managing, downloading, analyzing and visualizing of a diverse range of hydrologic observations data to provide research and analysis in Northwestern China. And then we describe an approach, there shows architecture, implementation of a web application of Watershed Datacenter System. Based on CUAHSI-HIS open source project, we adopt an Entity Framework Code-First approach using Model-View-Controller 5 which is a new architecture pattern

provided by Microsoft to implement data download and data management functionality. Besides, we create a data authorization functionality which is used to download and share data according to the user’s authority level considering the sensitivity of hydrologic data and China’s social circumstances. Finally, some cross-platform JavaScript libraries (jQuery, D3 and Dojo), Web GIS technology (ArcGIS API), HTML5 and HydroServerTSA tool are used to support visualization and analysis for hydrologic data.

2 Methodology and Design

2.1 ASP.NET MVC5

Web application development has come a long way since the beginning of the World Wide Web. Originally, front end and back end technologies come together to build web applications and developers need to a large of technologies to create just a simple web application [4]. In order to agile development, a new concept called smart user interface (Smart UI) was provided to helps developers construct a UI by dragging a set of server controls. As shown in Fig. 1a, this concept was to make Web development feel just the same as desktop development. It attempted to hide HTTP and HTML and developers didn’t need to work with a series of independent HTTP request and response [5]. However, reality proved more complicated. It goes back and forth with every request, leading to slower response times and increasing the bandwidth demands of the server. Besides, it often takes much trouble to page life cycle which can be extraordinarily complicated and delicate. The biggest drawback is that it is hard to maintain and extend. Mixing the domain model and

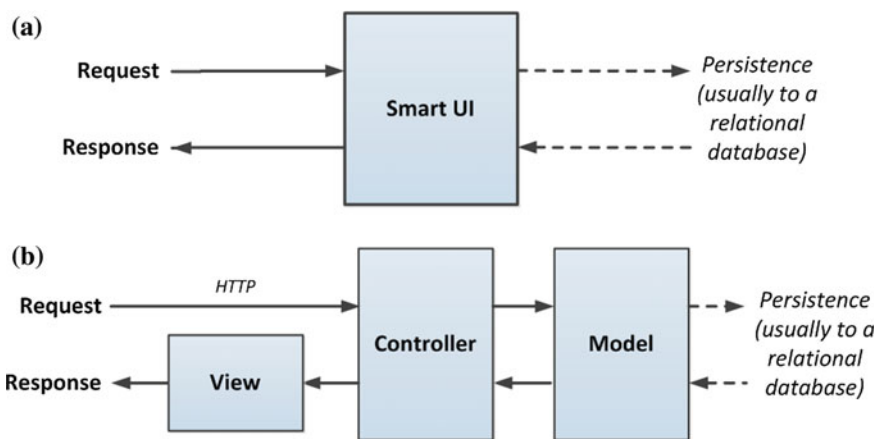


Fig. 1 a Smart UI architectural pattern and b Model-view-controller architectural pattern

business logic code with the UI leads to duplication, which brings up a terrible result that only adding any simple new feature are almost impossible to break the existing construct of code.

Model-View-Controller has been an important architectural pattern in computer science since 1970s and has gained enormous popularity today as an architectural pattern for Web applications. It is a powerful and elegant means of separating concerns, with the goal of creating applications that are easier to manage and maintain [5, 6]. ASP.NET MVC is a Web development framework from Microsoft, implements the effectiveness and tidiness of Model-View-Controller architecture pattern and provides greatly improved separation of concerns that is especially suitable for Web applications. As shown in Fig. 1b, it is separate the modeling of the domain, the presentation, and the actions based on users into three separate parts, as follows:

Model is responsible for managing the behavior and data of the application domain, responding to requests for information from view, and responding to instructions to controller.

View is display data and GIS information.

Controller is used for handling the interactions and updating the model to reflect a change in state of the Web application, and then passes information to the view.

In our Watershed Datacenter System (WDC), retrieving data from database and displaying, sharing for users are very important parts. It might be inclined to tie these pieces together to reduce the amount of coding and to improve application preformation. However, this seemingly natural but has some significant problems. On the one hand, the user interface tends to change much more frequently than the data storage system. On the other hand, coupling the data and user interface pieces is that business applications tend to incorporate business logic that goes far beyond data transmission [7]. For these reasons, in this paper we use ASP.NET Model-View-Controller 5 framework to develop which is the latest version of MVC supported by Microsoft.

2.2 *Entity Framework 6 Code-First Approach*

Due to the popularity of relational database management system (RDBMS) and the use of object-oriented concepts for software development, application objects inevitably need to be stored in object-oriented relational databases [8]. Unfortunately, it brings a set of technical difficulties when RDBMS is being used by object-oriented concept, which is often referred to as object-relational impedance mismatch [9].

For this reason, ADO.NET Entity Framework was developed by Microsoft, which was introduced out-of-the-box Object Relational Mapping that enabled .NET developers to work with relational data using domain-specific objects. It uses this conceptual model while querying from the database, creating objects from that data

and then persisting changes back to the database. Since Entity Framework 4.1, it introduced Code-First approach which is mainly useful in Domain Driven Design. With the Code-First approach, the developers can focus on the domain design and start creating classes as per your domain requirement rather than design your database first and then create the classes which match your database design [10].

As shown in Fig. 2, Database First, Model First and Code First are three ways of building an Entity Data Model that can be used with Entity Framework to perform data access. Database First is used to create the model by simple drag and drop from an existing database [11]. In Model First, some models are created by designer that ensure the generation of the database after a specified connection. Code-First approach provides an alternative to the Database First and Model First approaches to the Entity Data Model, which takes many benefits of defining the model with code, whether developers are working with an existing database or building one from scratch. In the Code-First approach, the developers don't use boxes and lines to create a .edmx model and they are focused on defining their domain model using POCO classes, which have no dependency on Entity Framework. It infers a lot of information about your model purely from the shape of your classes which is a cleaner, quicker, simpler way and has full control of this code and can easily modify.

CUAHSI-HIS are focused on bringing together hydrologic observations from multiple sources of the globe into a uniform, standards-based, service-oriented environment where heterogeneous data can be seamlessly integrated for advanced computer-intensive analysis and modeling [12]. They provide a Database schema (ODM) to store hydrologic observations data and optimize data retrieval for integrated analysis of information collected by investigators. In our WDC, we adopt

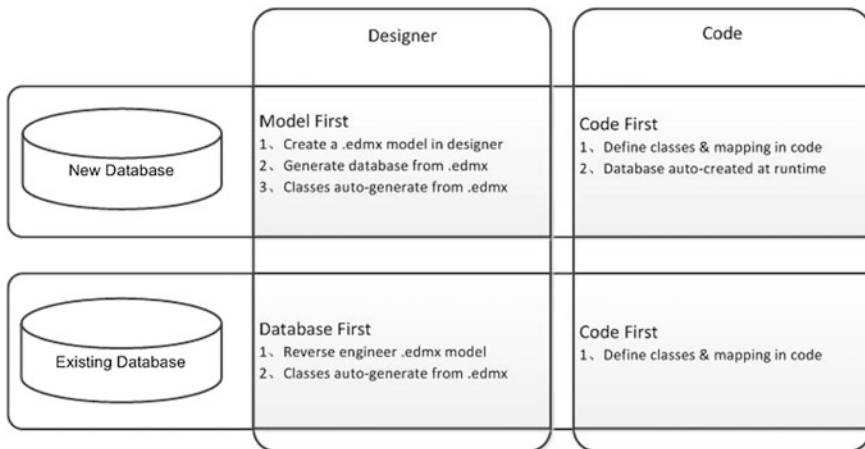


Fig. 2 Model first, database first and code-first approach [10]

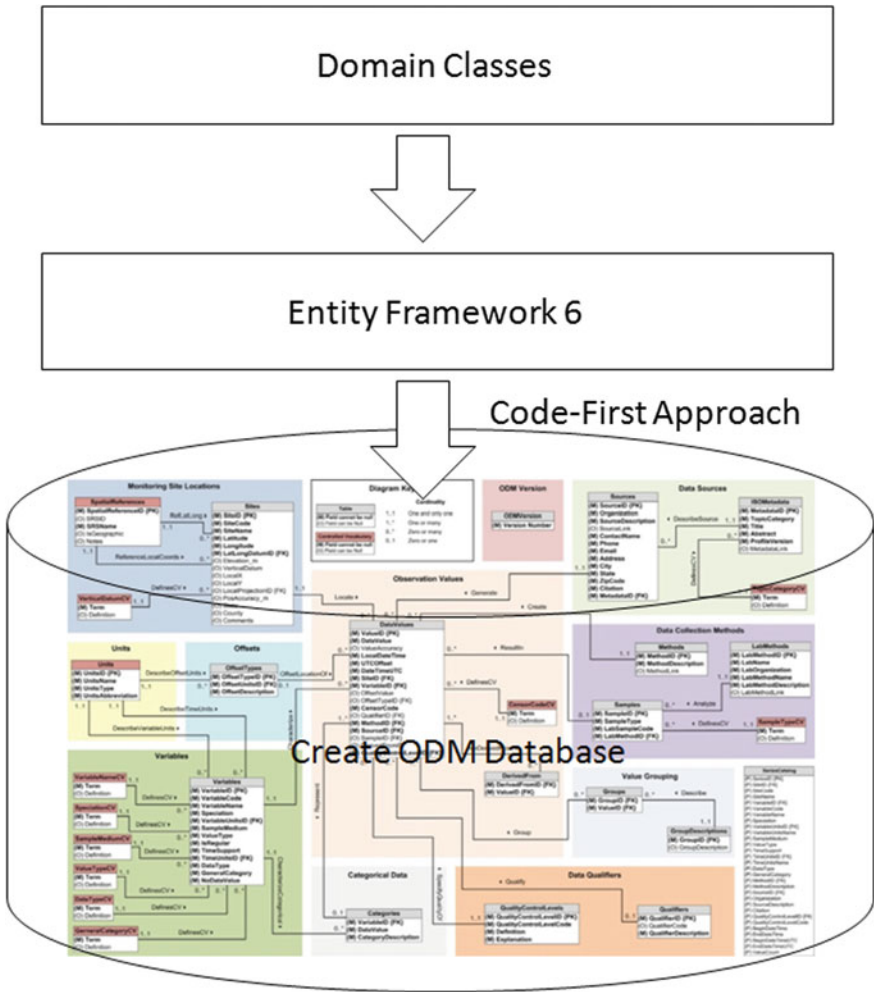


Fig. 3 Create an ODM database using code-first approach

Entity Framework Code-First approach based on MVC5 to implement the ODM database model, which is a cleaner, quicker, simpler development method (Fig. 3). On the one hand, ODM database model brings us a standard format to share data among investigators to facilitate the analysis of information which is a very mature database schema has been widely applied in many parts of the world. On the other hand, Entity Framework Code-First approach based on MVC5 can bring a better performance of website and change database structure by modifying Domain classes to suit WDC’s needs.

2.3 The Architecture of Watershed Datacenter System

System architecture is used to describe the overall system design and software system structure, which is great significance of our WDC. As shown in Fig. 4, 4-layers architecture are developed as follows, which aims to enhance their logical relationship and improve the system flexibility:

Data Storage Layer is responsible for storing all kinds of watershed data, which is the lowest layer in our 4-layers architecture.

Data Access Layer defines some database operations methods for querying, creating, deleting, updating and other operations with database through Entity Framework.

Business Logic Layer provides some services for Web application by calling the Data Access Layer.

Web Representation Layer is responsible for interacting with users, which includes View and Controller section of MVC. Besides, it contains two functional modules, Member and Manage.

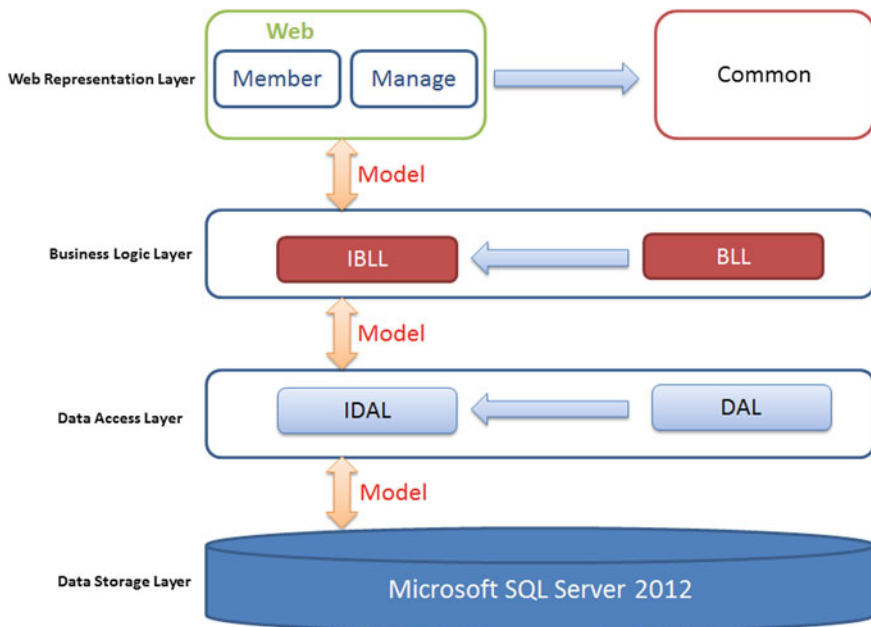


Fig. 4 The system architecture of watershed information datacenter

3 Implementation of the System

3.1 Hardware Condition

WDC was deployed on IBM xSeries 346 Server using 2 TB hard disk space, 2 GB of RAM and running on Windows Server 2008 as an operating system. And we use data for test purposes from middle reaches of the Heihe River Basin in northwest China, which is significant, complicated and highly impacted by agricultural irrigation. The objective was to demonstrate that WDC can download, manage data, and support real-time online analysis for multi-user activity.

3.2 Responsive Web Design

With the explosive growth of the Mobile Internet, more users will access the web through mobile devices than PC or other wireline methods. For this reason, we adopt an approach of Web UI design called Responsive Web Design (RWD) using HTML5, CSS framework, some jQuery plugin, which aim at crafting sites to provide an optimal view experience (easy reading and navigation with a minimum of resizing, panning, and scrolling) across a wide range of devices, such as PC, Mobile phone, Pad and other wireless devices (Fig. 5).

3.3 Functionality

As shown in Fig. 5, WDC mainly includes three modules: Data, Observation and Model.

The objective of the Data module is to put together hydrologic observation data and geographic data of the Watershed from multiple sources across China into a uniform, standard format. It's comprised of database (based on ODM model), ArcGIS Map servers, and Web services. Besides, some open source software components are seamlessly integrated into this module (such as HydroServer TSA). According to investigators' own needs, they can query particular observation data and then quickly analyze data through Time-Series tools in a certain research field. Moreover, they can download the interested hydrologic observation data.

In Observation module, most general functions of Web-GIS are equipped for our WDC, such as Zoom, Pan, Former View, Next View, Length, Area manipulation, Legend, Bookmark and so on. In addition, some advanced Web-GIS functions have also been developed combing ArcGIS API for JavaScript with jQuery in this module.

In Model module, we developed a visualization interface for analyzing the spatial distribution of outputs of MODFLOW model. Now, a spatial-distribution



Fig. 5 The responsive web design and three main modules of WDC

analysis of middle reaches of Heihe River basin has been provided. On the one hand, investigators can acquire the changing trend of groundwater on the observation well which has been selected and clicked during the custom time; On the other hand, the spatial distribution of groundwater level and contour in a custom year can be obtained by ArcGIS Geoprocess services in this module.

3.4 Data Visualization

There shows 4 types of charts in Fig. 6a(1–4), Time-Series chart, probability statistics chart, histogram chart and box whisker chart which provide investigators more in-depth analysis with the data. Besides, our WDC can also provide Spatial Analysis for our observation area. The users can choose their interested year and month to get the spatial-distribution analysis for groundwater level and drawdown. In Fig. 6b(1–2), it shows the spatial distribution of groundwater level and contour, from which the red color represents the higher elevation and the blue color represents the lower elevation, respectively. In Fig. 6b(3–4), the red color represents the reduced elevation and the blue color represents the increased elevation. The above has provided investigators a very intuitive user interface to visualize.

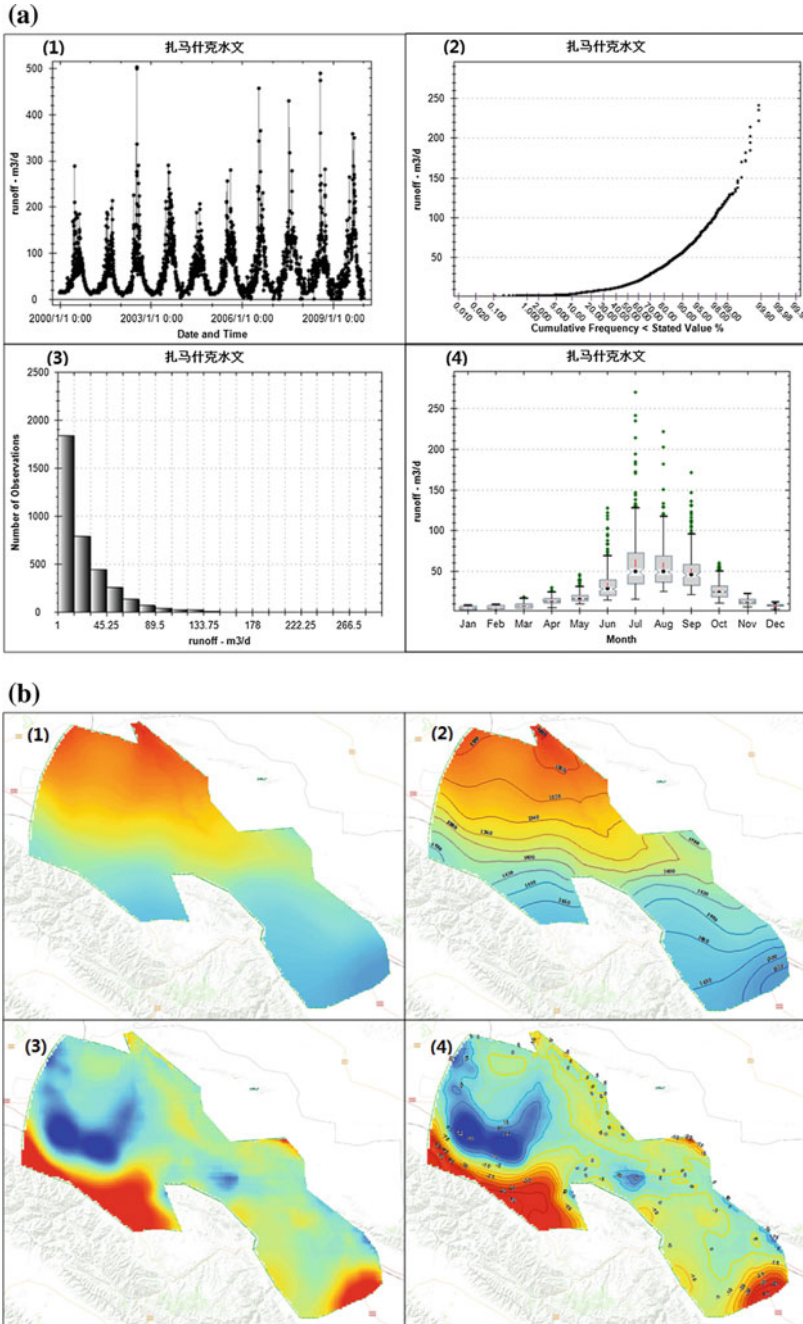


Fig. 6 Time-series analysis and spatial analysis. **a** Time-series analysis based on hydroserver TSA. **b** Spatial visualization

4 Discussion and Conclusion

The rapid development of Web technology has come a long way since the beginning of the World Wide Web. Developers combine HTML code with backend programming languages to create dynamic web applications will lead to unmaintainable and some trouble. At the same time, Microsoft embraces web technologies and provides a lot of frameworks, design pattern and developing approach that help developers to create web applications for mobile devices and other wireless devices. Especially since Code-First approach is introduced in EF6, it will take many benefits for developers in defining the model with code and architectural pattern. This paper establishes a watershed datacenter platform by using EF6 Code-First approach, Web GIS technology, Responsive Web Design and some cross-platform JavaScript libraries. Besides, the open source CUAHSI-HIS is seamlessly integrated to improve the analysis and visualization of data in WDC. Through the establishment of the watershed datacenter platform, this paper comes to the following conclusion.

One important advantage of this paper is that it establishes a watershed datacenter system for downloading, managing hydrologic data and other analysis and visualization functions to provide the convenience of investigators, geotechnical experts and public users. Three important modules in Watershed Datacenter platform, i.e., Observation module, Model module are provided for investigators, geotechnical experts, and public users; Besides, the Data module are controlled by data authorization function which can be used to download and share data according to the user's authority level considering the sensitivity of hydrologic data and China's social circumstances. The other important advantage is that the ODM database model and corresponding database were utilized in the system developed, which has rarely been applied in China. CUAHSI-HIS is a mature open source project which formulates its own standards for observation data and provides a perfect observation data model (ODM). Through this Watershed Datacenter System, ODM data format can be proved which can be efficiently used without paying much cost for many governmental and non-governmental projects in China. In terms of hydrologic observation data, it provides Time-Series analysis tools which have been seamlessly integrated into our WDC. All these enable investigators, geotechnical experts and public users to have a good grasp of the data structure of the platform.

Two problems which can be improved with further study: (1) Online MODFLOW service and real-time display should be implemented in the future. (2) Now, our WDC was developed based on ArcGIS API for JavaScript and ArcGIS Map Server, which are not open source. It will bring copyright issues when other organizations and developers want to apply in other places.

Acknowledgments This work was supported by National Natural Science Foundation of China under Grant No. 61402210 and 60973137, Program for New Century Excellent Talents in University under Grant No. NCET-12-0250, "Strategic Priority Research Program" of the Chinese Academy of Sciences with Grant No. XDA03030100, Gansu Sci.&Tech. Program under Grant

No. 1104GKCA049, 1204GKCA061 and 1304GKCA018. The Fundamental Research Funds for the Central Universities under Grant No. lzujbky-2014-49, lzujbky-2013-k05, lzujbky-2013-43, lzujbky-2013-44 and lzujbky-2012-44, Gansu Telecom Cuiying Research Fund under Grant No. lzudxcy-2013-4, Google Research Awards and Google Faculty Award, China.

References

1. Bao C, Fang CL (2007) Water resources constraint force on urbanization in water deficient regions: a case study of the Hexi Corridor, arid area of NW China. *Ecol Econ* 62(3):508–517
2. Horsburgh JS, Tarboton DG (2008) CUAHSI community observations data model (ODM) version 1.1. 1 Design specifications
3. Nian Y, Li X (2012) Design and implementation of hydrologic data sharing for the Heihe River Basin based on the open source hydrologic information system. In: 2012 international symposium on geomatics for integrated water resources management (GIWRM). IEEE
4. Pop DP, Altar A (2014) Designing an MVC model for rapid web application development. *Procedia Eng* 69:1172–1179
5. Freeman A, Sanderson S (2012) *Pro Asp. net Mvc 4*. Apress, New York
6. Heidke N, Morrison J, Morrison M (2008) Assessing the effectiveness of the model view controller architecture for creating web applications. In: Midwest instruction and computing symposium, Rapid City, SD
7. Model-View-Controller. <https://msdn.microsoft.com/en-us/library/ms978748.aspx>
8. Philippi S (2005) Model driven generation and testing of object-relational mappings. *J Syst Softw* 77(2):193–207
9. Smith KE, Zdonik SB (1987) Intermedia: a case study of the differences between relational and object-oriented database systems, vol 22, no 12. ACM
10. Lerman J, Miller R (2011) *Programming entity framework: code first*. O'Reilly Media, Inc.
11. Entity Framework Tutorial. <http://www.entityframeworktutorial.net/code-first/what-is-code-first.aspx>
12. Ames DP et al (2012) HydroDesktop: web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environ Model Softw* 37:146–156

Student-t Mixture Modelling for Image Segmentation with Markov Random Field

Taisong Xiong, Yuanyuan Huang and Xin Luo

Abstract In this paper, a Student-t mixture model is proposed for image segmentation based on Markov random field (MRF). For the clusters of pixels, their prior probabilities are regarded as a MRF. In the proposed model, at first, a factor of capture the spatial relationships between the pixels is given. Furthermore, student-t distribution is adopted to the component function of the distribution of pixels instead of the Gaussian distribution. To inference the parameters of the proposed model, gradient descent method is used during the inference process. Comprehensive experiments conducted on grayscale noisy image and real-world color images shows the effectiveness and robustness of the proposed model.

Keywords Student's t-distribution · Markov random field · Image segmentation · Gradient descent

1 Introduction

In the past ten years, finite mixture model (FMM) has been successfully applied to image segmentation [1]. The component function of FMM can be any probabilistic distribution. The FMM is referred to Gaussian mixture model if its component function is Gaussian distribution [2, 3]. To inference the parameter of the FMM, expectation-maximization (EM) algorithm [3] is in general adopted. Another

T. Xiong · X. Luo

College of Applied Mathematics, Chengdu University of Information Technology,
Chengdu 610054, People's Republic of China
e-mail: xiongtaisong@gmail.com

X. Luo

e-mail: luoxin@cuit.edu.cn

Y. Huang (✉)

Department of Network Engineering, Chengdu University of Information Technology,
Chengdu 610054, People's Republic of China
e-mail: iyyhuang@hotmail.com

probabilistic distribution, Student's t-distribution is a robust alternative to Gaussian distribution [4]. The definition of multivariate Student's t-distribution is given by

$$St(x|\Theta) = \frac{\Gamma(D/2 + v^2/2)}{\Gamma(v^2/2)} \frac{|A|^{1/2}}{(\pi v^2)^{D/2}} \left[1 + \frac{\Delta^2}{v^2} \right]^{-(v^2 + D)/2} \quad (1)$$

where D is the dimensionality of variable x , Δ^2 denotes the squared Mahalanobis distance and its definition is written as follows.

$$\Delta^2 = (x - \mu)^T A (x - \mu) \quad (2)$$

The Student's t-distribution mixture model (StMM) has also applied to signal process and image segmentation [5]. However, FMM cannot obtain satisfied segmentation results under the noisy condition [6]. The main reason is that FMM assumes that the pixels are independent of each other. In fact, the position information plays an important role in image segmentation [6]. The Markov random field (MRF) model which considered the spatial relationship between pixels is proposed in [7]. Some models based on MRF are proposed in [8–13] and applied to image segmentation. They all obtains better segmentation results.

Considering the aforementioned models, in this paper, we propose a Student's t-distribution mixture model which is based on the MRF. In the proposed model, the spatial relationship between the pixels depends on the context mixture parameter and posterior probability of pixel and its neighborhood. To inference the parameter of the proposed model, gradient descend algorithm is used. The segmented images includes synthetic grayscale image and real-world color images. The experimental results demonstrates the effectiveness and robustness of the proposed model.

2 Proposed Model

In this section, a smoothing prior $U(\Pi)$ is given. First, we define a factor F_{nk} to reflect the spatial relationship between pixels.

$$F_{nk} = \exp \left[\frac{\beta}{2N_n} \sum_{m \in \partial_n} (z_{nk} + (\pi_{nk} - \pi_{mk})^2) \right]. \quad (3)$$

where ∂_n is the neighborhood of pixel n , z_{nk} represents its posterior probability and β plays a smooth function. Throughout this paper, a second neighborhood system (3×3 window) is chosen for the proposed model and β is set to 40. The factor F_{nk}

is only affected by the context mixture coefficient and posterior probability. The definition of smoothing prior $U(\Pi)$ is given by

$$U(\Pi) = - \sum_{n=1}^N \sum_{k=1}^K F_{nk} \log \pi_{nk}. \quad (4)$$

The smoothing prior $U(\Pi)$ reflects the spatial relationship of pixels. According to the smoothing prior $U(\Pi)$, the MRF distribution can be written as follows

$$P(\Pi) = Z^{-1} \exp \left\{ \frac{1}{T} \sum_{n=1}^N \sum_{k=1}^K F_{nk} \log \pi_{nk} \right\}. \quad (5)$$

Furthermore, in the proposed model, the Student's-t distribution is chosen for the component function instead of the Gaussian distribution. Because the Student's-t distribution is a robust alternative to Gaussian distribution [5]. Combined with the Student's-t distribution, the log-likelihood function can be written as follows.

$$L(X|\Pi, \Theta) = \sum_{n=1}^N \log \left\{ \sum_{k=1}^K \pi_{nk} St(x_n|\theta_k) \right\} - \log Z + \frac{1}{T} \sum_{n=1}^N \sum_{k=1}^K F_{nk} \log \pi_{nk} \quad (6)$$

When all of data conditions are met, the log-likelihood function in (6) can be rewritten as follows [8].

$$L(X|\Pi, \Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \{ \log \pi_{nk} + \log St(x_n|\theta_k) \} - \log Z + \frac{1}{T} \sum_{n=1}^N \sum_{k=1}^K F_{nk} \log \pi_{nk} \quad (7)$$

When the all data is determined, the posterior probability can be obtained as the following equation.

$$z_{nk} = \frac{\pi_{nk} St(x_n|\theta_k)}{\sum_{j=1}^K \pi_{nj} St(x_n|\theta_j)} \quad (8)$$

Next, our objective is to inference the parameter set $\{\Pi, \Theta\}$. The parameter $\Pi = \pi_{nk}, n = 1, 2, \dots, N, k = 1, 2, \dots, K$ and $\Theta = \mu_k, \lambda_k, v_k, k = 1, 2, \dots, K$. To simplify the inference process, the parameter Z and T are both set to 1 which is similar to the method adopted in [9, 10]. Then the form (7) can be rewritten as follows.

$$L(X|\Pi, \Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \{ \log \pi_{nk} + \log St(x_n|\theta_k) \} + \sum_{n=1}^N \sum_{k=1}^K F_{nk} \log \pi_{nk} \quad (9)$$

Substitute the Student’s-t distribution in (1) into (9), the form of (9) is rewritten as follows.

$$L(X|\Pi, \Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \left\{ \log \pi_{nk} + \log \Gamma\left(\frac{D+v_k}{2}\right) - \log \Gamma\left(\frac{v_k}{2}\right) + \frac{1}{2} \log |A_k| \right. \\ \left. - \frac{D}{2} \log v_k - \frac{D+v_k}{2} \log\left(1 + \frac{(x_n - \mu_k)^T A_k (x_n - \mu_k)}{v_k}\right) \right\} + \sum_{n=1}^N \sum_{k=1}^K F_{nk} \log \pi_{nk}. \tag{10}$$

To maximize the log-likelihood function, the EM algorithm is adopted. In the E-Step, to inference the values of parameter Θ , gradient descend is used. The optimal value of Θ is obtained as follows.

$$\Theta^{(t+1)} = \Theta^{(t)} - \varphi \nabla L(\Theta^{(t)}) \tag{11}$$

where $\nabla L(\Phi^{(t)}) = [\partial L/\partial \mu_k, \partial L/\partial A_k, \partial L/\partial v_k]$, and t denotes the iteration number. φ is the learning rate. In general, its value is very little. In this paper, the value of φ is set to $1e-7$. The gradients of $\partial L/\partial \mu_k, \partial L/\partial A_k, \partial L/\partial v_k$ are obtained by the following equations, respectively.

$$\frac{\partial L(\Theta)}{\partial \mu_k} = \sum_{n=1}^N z_{nk} \left\{ \frac{v_k + D}{v_k + (x_n - \mu_k)^T A_k (x_n - \mu_k)} A_k (x_n - \mu_k) \right\} \tag{12}$$

$$\frac{\partial J(\Theta)}{\partial A_k} = \frac{1}{2} \sum_{n=1}^N z_{nk} \left\{ (A_k^{-1})^T - \frac{v_k + D}{v_k + (x_n - \mu_k)^T A_k (x_n - \mu_k)} (x_n - \mu_k)(x_n - \mu_k)^T \right\} \tag{13}$$

$$\frac{\partial J(\Theta)}{\partial v_k} = \frac{1}{2} \sum_{n=1}^N z_{nk} \left\{ \psi\left(\frac{D+v_k}{2}\right) - \psi\left(\frac{v_k}{2}\right) - \frac{D}{v_k} - \log\left(1 + \frac{(x_n - \mu_k)^T A_k (x_n - \mu_k)}{v_k}\right) \right. \\ \left. + \frac{(D+v_k)(x_n - \mu_k)^T A_k (x_n - \mu_k)}{v_k(v_k + (x_n - \mu_k)^T A_k (x_n - \mu_k))} \right\} \tag{14}$$

Because π_{nk} should satisfy the constraints. Therefore, the inference of π_{nk} cannot only use the gradient of $\partial L/\partial \pi_{nk}$. To make π_{nk} meet these constraints, the Lagranges multiplier φ_n is introduced.

$$\frac{\partial}{\partial \pi_{nk}} \left[L - \sum_{n=1}^N \varphi_n \left(\sum_{j=1}^K \pi_{nj} - 1 \right) \right] = 0. \tag{15}$$

From (15), we can obtain the following equation.

$$z_{nk} + F_{nk} - \varphi_n \pi_{nk} = 0. \quad (16)$$

To obtain the value of φ_n , according to $\sum_{k=1}^K z_{nk}$ and $\sum_{k=1}^K \pi_{nk} = 1$, we have

$$\sum_{k=1}^K z_{nk} - \sum_{k=1}^K F_{nk} - \varphi_n \sum_{k=1}^K \pi_{nk} = 0. \quad (17)$$

Then, we can obtain the value of φ_n .

$$\varphi_n = 1 + \sum_{k=1}^K F_{nk}. \quad (18)$$

According to Eqs. 16 and 18, the value of π_{nk} can be obtained

$$\pi_{nk} = \frac{z_{nk} + F_{nk}}{1 + \sum_{j=1}^K F_{nj}}. \quad (19)$$

When all of the parameters are determined, the label k of pixel n can be determined as follows according to max posterior probability.

$$z_{nk} \geq z_{nj}, \quad k, j = (1, 2, \dots, K). \quad (20)$$

The steps of Student's-t distribution mixture model based on MRF are summarized as in algorithm 1.

Algorithm 1: The Proposed model.

Initialize step:

The initial values of mean μ_k and variance A_k are determined by K -means. The value of v_k and π_{nk} are both set to 1.

Step 1: E-Step

Calculate the Student's t-distribution $St(x|\Theta)$ in (1).

Update the value of the posterior probability z_{nk} in (8).

Step 2: M-Step

Update the parameters $\Theta = (\mu_k, A_k, v_k)$ using (11).

Update the value of π_{nk} using (19).

Step 3:

If the convergent condition is not satisfied, then return to step 1.

Step 4:

To assign the class label of pixel using (20).

3 Experimental Results

In this section, the performance of the proposed model for image segmentation is evaluated visually and quantitatively. Some state-of-the-art models are adopted to compared with the proposed model. The experiments are conducted on synthetic grayscale image and natural color images. Two criteria are used to quantitatively evaluate the effect of the image segmentation results. The measure of correct classification ratio (CCR) is defined as follows [10].

$$CCR = \sum_{k=1}^K \frac{|GT_k \cap Seg_k|}{|GT|}. \quad (21)$$

where GT_k denotes the ground truth for the k th cluster and Seg_k is the k th cluster obtained by algorithm. The higher value of CCR denotes the better segmentation results. The measure CCR is used to evaluate the performance of algorithm for synthetic images. Another measure, the probabilistic rand (PR) index [14] is adopted to quantitatively measure the segmentation results of natural images.

3.1 Synthetic Images

First, a four-class ($K = 4$) synthetic image is used to test the performance of the proposed model, compared with the GMM [3], StMM [4, 5], SVFMM [9], MRFM [13], DPSMM [13]. The original image is shown in Fig. 1a. The noisy images is shown in Fig. 1b which is degraded by Gaussian noise with (mean 0, and 0.01 variance). It is very difficult to label the cluster of pixels from the noisy image. At the same time, the edge of cluster is blur because of the noise. The segmentation results obtained by these models are shown in Fig. 1c–h. GMM and StMM, which consider that the pixels are independent, obtain lower CCR than the other models. It shows that the spatial information plays an important role in image segmentation. Furthermore, the proposed model obtains best segmentation result because its CCR value is the highest among all the models. The experiment demonstrates the robustness and correctness of our proposed model. It also shows that our proposed model effectively captures the spatial relationships between the pixels.

3.2 Natural Image Segmentation

To further evaluate the performance of the proposed model, a natural color image segmentation dataset [15] is chosen for visual and quantitative comparison. The proposed model is compared with GMM, StMM, simulated field algorithm (SIMF) [11], SVFMM, MRFM, DPSMM. The color image shown in Fig. 2a is segmented

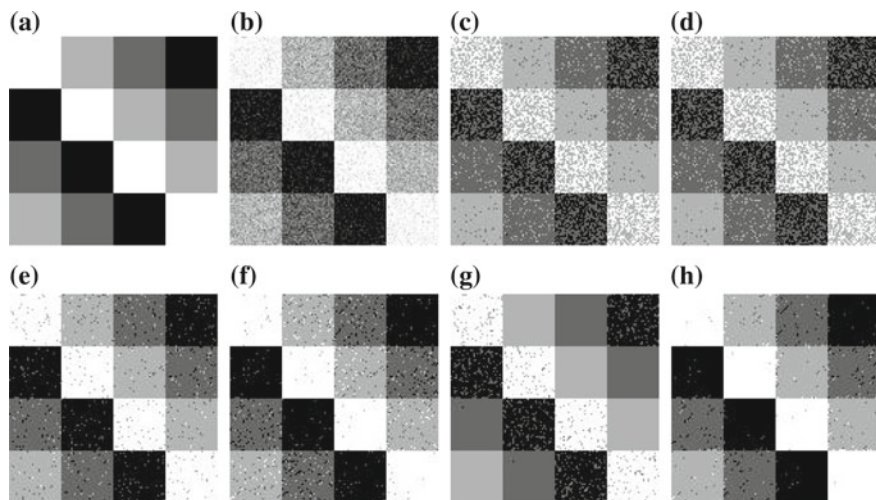


Fig. 1 Synthetic image segmentation (128×128 image). **a** Original image, **b** image corrupted by Gaussian noise (0 mean, 0.01 variance), **c** GMM (CCR = 72.45 %), **d** StMM (CCR = 72.24 %), **e** SVFMM (CCR = 95.55 %), **f** MRFM (CCR = 94.80 %), **g** DSMM (CCR = 93.90 %), **h** proposed model (CCR = 97.62 %)

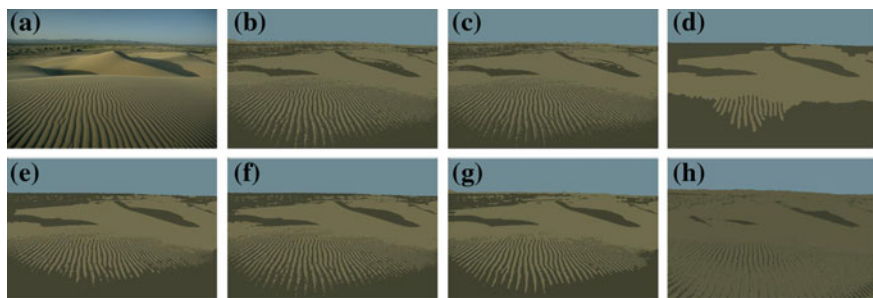


Fig. 2 Color image segmentation (178054). **a** The original image, **b** GMM (PR = 0.712), **c** StMM (PR = 0.713), **d** SIMF (PR = 0.608), **e** SVFMM (PR = 0.720), **f** MRFM (PR = 0.716), **g** DPSMM (PR = 0.713), **h** proposed method (PR = 0.736)

into three classes: sky, desert and shadow. The segmentation results obtained by all models are shown in Fig. 2b–h, respectively. The proposed model can segment the image very well compared with any other model. The edges of three clusters are more smooth in the segmentation result obtained by the proposed model than some other models.

Some natural color images [15] are used to evaluate the effectiveness of the proposed model against GMM, StMM, SIMF, SVFMM, MRFM and DPStMM. The PR values of image segmentation results obtained by all models are given in Table 1. From Table 1, the StMM obtains better segmentation results than GMM. It

Table 1 Comparison of image segmentation results based on Berkeley images: PR index

Image	K	GMM	StMM	SIMF	SVFMM	MRFM	DPSMM	Proposed model
108082	3	0.582	0.560	0.609	0.596	0.604	0.605	0.620
167062	3	0.853	0.834	0.822	0.863	0.808	0.878	0.983
306005	4	0.722	0.724	0.593	0.749	0.734	0.758	0.755
26031	2	0.481	0.481	0.506	0.490	0.497	0.480	0.655
374020	3	0.675	0.697	0.743	0.738	0.737	0.740	0.720
130034	2	0.511	0.514	0.531	0.502	0.502	0.505	0.571
100098	3	0.616	0.615	0.558	0.628	0.617	0.607	0.651
23080	3	0.733	0.748	0.748	0.740	0.763	0.739	0.743
351093	4	0.778	0.734	0.563	0.792	0.734	0.712	0.792
54005	3	0.493	0.501	0.540	0.481	0.480	0.480	0.593
104022	4	0.608	0.616	0.551	0.647	0.610	0.609	0.631
41025	3	0.593	0.596	0.563	0.589	0.574	0.608	0.610
15088	2	0.852	0.864	0.838	0.842	0.813	0.870	0.858
311068	3	0.635	0.629	0.514	0.612	0.613	0.609	0.655
293029	5	0.644	0.652	0.603	0.673	0.678	0.667	0.676
304074	3	0.659	0.663	0.684	0.684	0.683	0.683	0.676
78019	4	0.676	0.717	0.577	0.796	0.738	0.740	0.799
388016	5	0.665	0.686	0.560	0.695	0.705	0.683	0.693
178054	3	0.712	0.713	0.608	0.720	0.716	0.713	0.736
90076	5	0.689	0.715	0.602	0.725	0.725	0.711	0.740
Mean	–	0.659	0.663	0.679	0.678	0.667	0.670	0.708

proves that the StMM is more robust to noise than GMM for natural image segmentation. Furthermore, a higher mean PR value is obtained by the proposed model. These experimental results demonstrate that the proposed model owns more robustness and correctness for real world image segmentation. It proves that the representation of spatial relationship of the proposed model is more correct and effective than some other models. It also demonstrates that the proposed model owns better robustness and correctness.

4 Conclusions

In this paper, we propose a StMM which captures the spatial relationship between the pixels based on MRF. A factor is given and can effectively captures the spatial relationship between these pixels. Furthermore, GMM is replaced by StMM in the proposed model for the component function because of the robustness of StMM. Experimental results conducted on grayscale noisy image and real-world color images demonstrate the robustness and effectiveness of the proposed model.

Acknowledgments This work is partially supported by the National Natural Science Foundation of China (No. 61303126), and the Applied Basic Research Program of Sichuan Province (No. 2014JY0168) and the Scientific Research Foundation of the Education Department of Sichuan Province (No. 14ZA0178) and Foundation of Chengdu University of Information Technology (No. KYTZ201426) and (No. KYTZ201419).

References

1. Marco A, Luciano N, Donatella V (2008) A finite mixture model for image segmentation. *Stat Comput* 18:137–150
2. McLachlan G, Peel D (2000) *Finite mixture models*. Wiley, New York
3. Bishop C (2006) *Pattern recognition and machine learning*. Springer, Berkeley
4. Peel D, McLachlan GJ (2000) Robust mixture modelling using the t distribution. *Stat Comput* 10:335–344
5. Sfikas G, Nikou C, Galatsanos N (2007) Robust image segmentation with mixtures of student's t-distribution. In: *IEEE international conference on image processing*. pp 273–276
6. Thanh MN, Wu QM, Ahuja S (2010) An extension of the standard mixture model for image segmentation. *IEEE Trans Neural Netw* 21:1326–1338
7. Geman S, Geman D (1984) Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6:721–741
8. Sanjay GS, Hebert TJ (1998) Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. *IEEE Trans Image Process* 7:1014–1028
9. Blekas K, Likas A, Galatsanos N, Lagaris I (2005) A spatially constrained mixture model for image segmentation. *IEEE Trans Neural Netw*
10. Nikou C, Galatsanos N, Likas A (2007) A class-adaptive spatially variant mixture model for image segmentation. *IEEE Trans Image Process* 16(4):1121–1130
11. Celeux G, Forbes F, Peyrard N (2003) EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recogn* 36(1):131–144
12. Nguyen TM, Jonathan Wu QM (2013) Fast and robust spatially constrained gaussian mixture model for image segmentation. *IEEE Trans Circ Syst Video Technol* 23(4):621–635
13. Thanh MN, Wu QM (2013) A finite mixture model for detail-preserving image segmentation. *Sig Process* 93:3171–3181
14. Unnikrishnan R, Pantofaru C, Hebert M (2007) Toward objective evaluation of image segmentation algorithms. *IEEE Trans Pattern Anal Mach Intell* 29:929–944
15. Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. *IEEE Trans Pattern Anal Mach Intell* 33(5):898–916

Integrated Genetic Algorithm and Fuzzy Logic for Planning Path of Mobile Robots

Shixuan Yao, Xiangrong Wang and Baoliang Li

Abstract This paper presents an efficient control scheme of integrating mathematical model, fuzzy logic and genetic algorithm to solve the path planning problem of the CCD wheeled robots, which is based on the specific monocular structure and electric properties. Prior knowledge about the problem domain will cause deviation between the practical path and the ideal path when wheeled robot works in high-speed. This system integrates image information of CCD sensor, current position of the robot, current velocity, deflection angle and the battery capacity and so on, to plan the ideal path of autonomous mobile robot by means of establishing image model and path model. Meanwhile, the fuzzy logic controller based on genetic algorithm saves relevant feedback in the process of system running and updates data parameters to reconstruct the rule database, in order to control the robot move by the ideal path. The proposed control scheme could be useful for planning path of mobile robots. Experimental studies show that the scheme has good feasibility and it can achieve high control precision.

Keywords Image model · Path model · Fuzzy logic · Genetic algorithm · Database

S. Yao (✉)

School of EMU Application and Maintenance Engineering,
Computer Applications Technology,
Dalian Jiaotong University, Dalian 116024, Liaoning, China
e-mail: xiangr_wang@163.com

X. Wang

School of EMU Application and Maintenance Engineering,
Dalian Jiaotong University, Dalian 116024, Liaoning, China

B. Li

School of Mechanical Engineering, Dalian Jiaotong University,
Dalian 116024, Liaoning, China

1 Introduction

As the rapid development of computer technology [1], intelligent sensor [2] and intelligent control [3], autonomous mobile robots [4] become the development trend, and applications for the robots abound in mining, construction, forestry, planetary exploration and the military. This paper build mathematical model of image pixels and the object to plan feasible paths based on monocular CCD sensor [5]. When the mechanical structure and electric property of robots is specific, algorithm determines the stability and accuracy of the system. The system focuses on the mathematical model and control algorithm in order to make the robot run along the actual path.

This system integrates the acquired information to plan ideal path of autonomous mobile robot by means of establishing image model and path model [6–8]. Meanwhile, the fuzzy logic controller (FLC) [9] based on genetic algorithm (GA) [10, 11] saves relevant feedback in the process of system running and updates data parameters to reconstruct the rule database, in order to control the robot move by the ideal path.

2 The Design of the Control System

The system is mainly divided into three modules: perception, decision-making units and execution units. As shown in Fig. 1,

Image acquisition module uses CCD sensor to sample traffic information, then extracts the signal through process circuit. Speed acquisition module output the moving speed and acceleration. Power management module is used to provide the energy for modules of the system and detect the real-time battery capacity, in case that low power results the unstable or restarted system when it speeds up, and reminds it to recharge. Image analysis module deals the image signal with

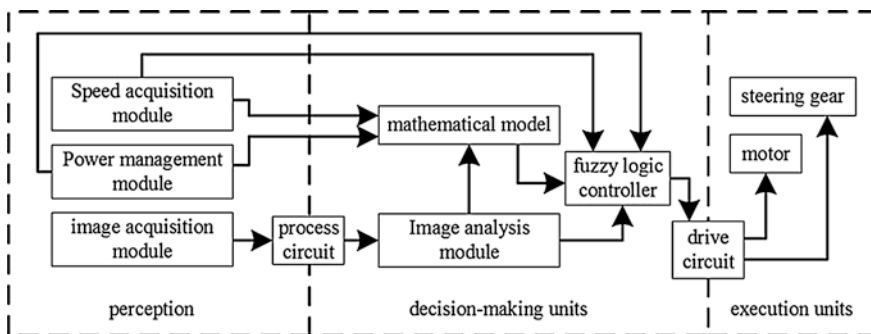


Fig. 1 The overall architecture of the system

binarization [12–14] and filtering [15], then stores it as pixel matrix and extracts characteristic parameters for subsequent modules. Integrating those information, planning a reasonable route according to the mathematical model, and then the FLC based on GA determine the best solution. Meanwhile, the FLC saves relevant feedback in the process of system running and updates data parameters to reconstruct the rules database.

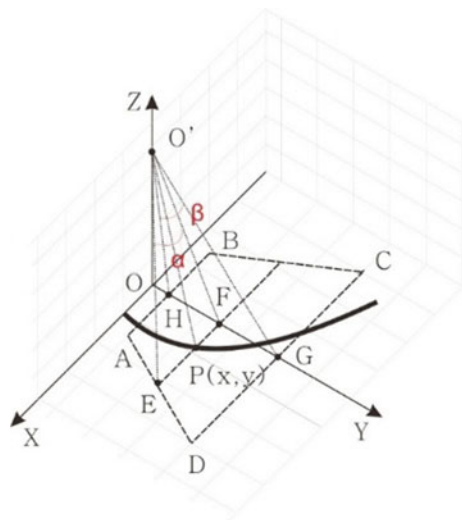
3 Image Model

The image is a set of pixels, and it is transformed into a matrix of m rows and n columns [16, 17]. As shown in Fig. 2, the model of CCD mounting position and calculation is established to restore the actual road information.

CCD is installed at the point O' , the road plane is formed by the x -axis and y -axis, O is origin of the x - y plane, $OO'H$ is blind area of the camera, image visual region is the area of $ABCD$ and the arc is the road boundary. Point P is on the arc in the visual area, and its projection angle on y -axis is $\angle OO'F$, its projection angle on x -axis is $\angle PO'F$. The perspective of visual longest segment HG on y -axis is β , the maximum perspective of point P on x -axis is α . Assuming the distance of HG is l_y , the distance of FE is l_x , the length of OH is l_b , the blind angel $\angle OO'H$ is θ , the height of OO' is h , the distance of point P on x -axis and y -axis is P_x and P_y , as shown in Eqs. (1) and (2), respectively.

$$P_x = \frac{h \times \tan(\angle PO'F)}{\cos(\theta + \angle HO'F)} \tag{1}$$

Fig. 2 Image model



$$P_y = h \times \tan(\theta + \angle HOF) \quad (2)$$

As shown in Fig. 2, the visualized line AB to CD correspond the data of matrix from the first line to the m -th line, while the line AD to BC correspond the data of matrix from the first column to the n -th column. The angle of β corresponding to the actual spacing is divided into m copies, while the angel α corresponding to the actual column spacing is divided into n copies.

Assuming the pixel data α_{ij} corresponds to the position of point P , so P_x, P_y can be rewritten, as shown in Eqs. (3) and (4), respectively.

$$P_x = \frac{h \times j \times \alpha}{\cos\left(\arctan\frac{l_b}{h} + \frac{i \times \beta}{m}\right)} \quad (3)$$

$$P_y = h \times \tan\left(\arctan\frac{l_b}{h} + \frac{i}{m}\beta\right) \quad (4)$$

4 Path Model

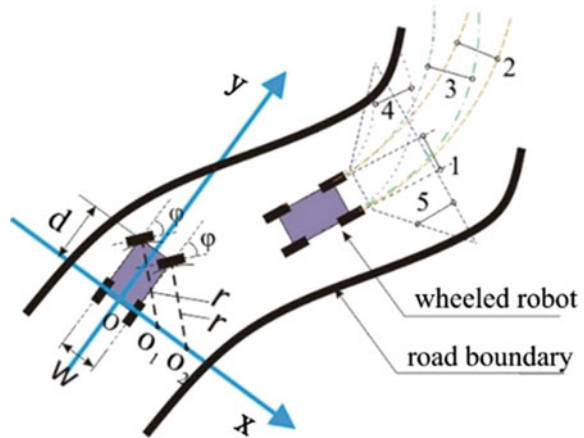
When wheeled robot turning, the two front wheels produce the same deflection of angle, and make circular motion around the center of the circle. To establish rectangular coordinate system, setting the axis of rear wheel as x -axis and setting the central axis of the robot as y -axis. As shown in Fig. 3, w presents the space between the two front wheels on x -axis, d presents the space between the two wheels on y -axis. When the deflection of angle is φ , two front wheels make circular motion around O_1 and O_2 with the radius of r respectively.

Selecting left front wheel as the research object, the relationship between the angle φ and the radius r is $r = d/\sin \varphi$. The origin O in Fig. 3 is the same point with the origin O in Fig. 2, so the coordinates of O_1 and O_2 are $(d \cot \varphi - (w/2), 0)$ and $(d \cot \varphi + (w/2), 0)$ respectively. Assuming α_{ij} is a pixel point of the CCD visible area. Combined with CCD image positioning and calculation model, the distance r between the point α_{ij} and the point O is:

$$r = \text{sqrt} \left[\left(\frac{j h \alpha}{n \times \cos\left(\arctan\frac{l_b}{h} + \frac{i}{m}\beta\right)} - d \cot \varphi - \frac{w}{2} \right)^2 + \left(h \times \tan\left(\arctan\frac{l_b}{h} + \frac{i}{m}\beta\right) \right)^2 \right] \quad (5)$$

where sqrt presents the square root. In the visual area of the CCD camera, the points that most accord with the r are obtained from the first line to the m -th line. A set of all the points that meet the condition constitute the final path of wheeled robot when the drift angle is φ .

Fig. 3 Path model



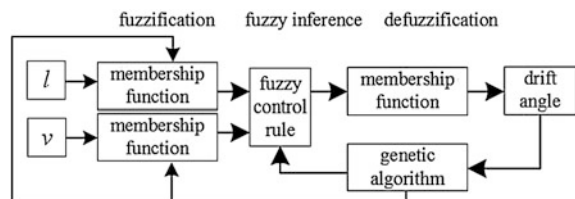
As shown in Fig. 3, steering gear has three types of drift angle. Path 1 presents the straight-line route which is the path of no deflection; Path 4 intersect the boundary which is the wrong path; Path 2 and 3 keep within the boundary which are the optional paths. It would be required that the wheeled robot working with the least drift angle of steering gear. So, path 2 and path 3 are optional paths, but the ideal path is determined by the current angle of the steering gear according to the principle of “the least drift angle”.

However, when wheeled robot works in high-speed, it will cause deviation between the practical path and the ideal path due to it will take some time when the steering gear running. So using the FLC based on GA to correct the practical path, and make the wheeled robot work along the ideal path.

5 The Design of FLC on GA

The essential design of FLC has strong subjectivity, so using GA to optimize the FLC in this paper. As shown in Fig. 4, it is the system block diagram of the FLC based on GA. The inputs of the FLC include the speed of v and the positional deviation of l which is the difference between current position and the ideal path; while the output is the drift angle ϕ of mobile robot. The membership functions and

Fig. 4 The system block diagram of the FLC



the fuzzy control rules, after optimizing by GA, are visited by FLC for the input fuzzification, fuzzy inference and defuzzification.

5.1 The Establishing of Rule Base for FLC

The inference rule of FLC is as follows: If l_1 is A_1^j and v_1 is A_2^j then φ_i is B^m . Where l_1 and v_1 are the input variables of the FLC; φ is the output variable; A_1^j and A_2^j are the fuzzy values of input variables, B^m is the fuzzy value of output variable; and $\mu_{A_1^j}(l_1)$, $\mu_{A_2^j}(v_1)$, $\mu_{B^m}(\varphi_i)$ are membership functions of fuzzy set A_1 , A_2 and B , respectively. Where $\mu_{B^m}(\varphi_i)$ uses the membership function of $\mu_{B^m}(\varphi_i) = 1$, and $\mu_{A_1^j}(l_1)$ and $\mu_{A_2^j}(v_1)$ choose symmetrical triangle-shaped membership function. Positional deviation l is divided into seven levels faintly and defining the left deviation is negative and defining the right deviation is positive, so the seven levels can be described as negative big (*NB*), negative middle (*NM*), negative small (*NS*), zero (*ZO*), positive small (*PS*), positive middle (*PM*), and positive big (*PB*), respectively. Meanwhile, the velocity v can be separated into three levels of low velocity (*N*), intermediate velocity (*Z*), and high velocity (*P*). And drift angle φ is also divided into seven levels and defining the left turning is negative and defining the right turning is positive, so the seven levels are the same with the levels of positional deviation l . The fuzzy basic function of the k -th rule is:

$$p_k(X) = \frac{(u_{A_1^j} u_{A_2^m})_k}{\sum_{i=1}^{15} (u_{A_1^j}(x_1) u_{A_2^m}(x_2))_i} \quad j = 1, 2, \dots, 7 \tag{6}$$

where m is the number of rules, and the output of the FLC can be described as:

$$\varphi(X) = \sum_{i=1}^m p_j(X) \varphi_j \tag{7}$$

5.2 The Optimizing of the FLC Via the GA

5.2.1 Parameter Coding

The triangle-shaped membership function which can be uniquely confirmed by three vertices. The seven integers of 0 to 6 are defined to present the seven language values, *NB*, *NM*, *NS*, *ZO*, *PS*, *PM*, and *PB*. According to the input and output fuzzy set, the joint encoding length of membership functions is $3 \times 3 = 9$, and the total

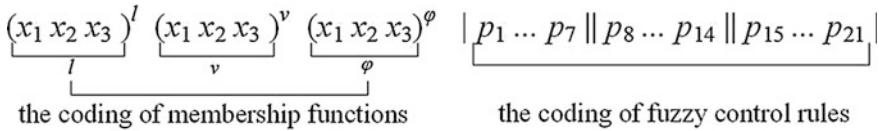


Fig. 5 Chromosome coding string

number of system rules is $3 \times 7 = 21$. Parameters can form a one-dimensional chromosome coding string, as shown in Fig. 5.

5.2.2 Confirming the Fitness Function

The good and bad of the fitness function may directly affect the evolution of the genetic algorithm. According to the characteristics of autonomous mobile robot, the method of adopting the discrete form of the objective function is used to evaluate the performance of the FLC, it is shown as Eq. (8).

$$J = \sum_{k=1}^{t_s} a_l |l(k)| + a_v |v(k)| + a_\varphi |\varphi(k)| \tag{8}$$

where t_s is the duration time of controller acting on the object; a_l , a_v and a_φ are the weighted coefficient of each item, and determine the proportion in the objective function, that is to say, greater value indicates bigger influence. Then do the calculation by the basic operations of GA include selection, cross and variation to optimize FLC.

6 Conclusions

In this study, the dimension of robot is 180 mm × 220 mm, the battery voltage is 7.2 V, the steering angle rate is 0.1 s/50°, the power of servo motor is 26.5 W. The road width is 450 mm, the radius of S-curve is 600 mm and set the initial position deviation is 30 mm. The optimal control rules is shown in Table 1 optimized by GA. The optimal membership functions of the input and the output are shown in

Table 1 The optimal control rules

φ (drift angle)		l (positional deviation)						
		<i>NB</i>	<i>NM</i>	<i>NS</i>	<i>ZO</i>	<i>PS</i>	<i>PM</i>	<i>PB</i>
v (velocity)	<i>N</i>	<i>PB</i>	<i>PM</i>	<i>PS</i>	<i>ZO</i>	<i>NS</i>	<i>NM</i>	<i>NB</i>
	<i>Z</i>	<i>PM</i>	<i>PS</i>	<i>PS</i>	<i>ZO</i>	<i>ZO</i>	<i>NS</i>	<i>NM</i>
	<i>P</i>	<i>PM</i>	<i>PS</i>	<i>ZO</i>	<i>ZO</i>	<i>ZO</i>	<i>ZO</i>	<i>NS</i>

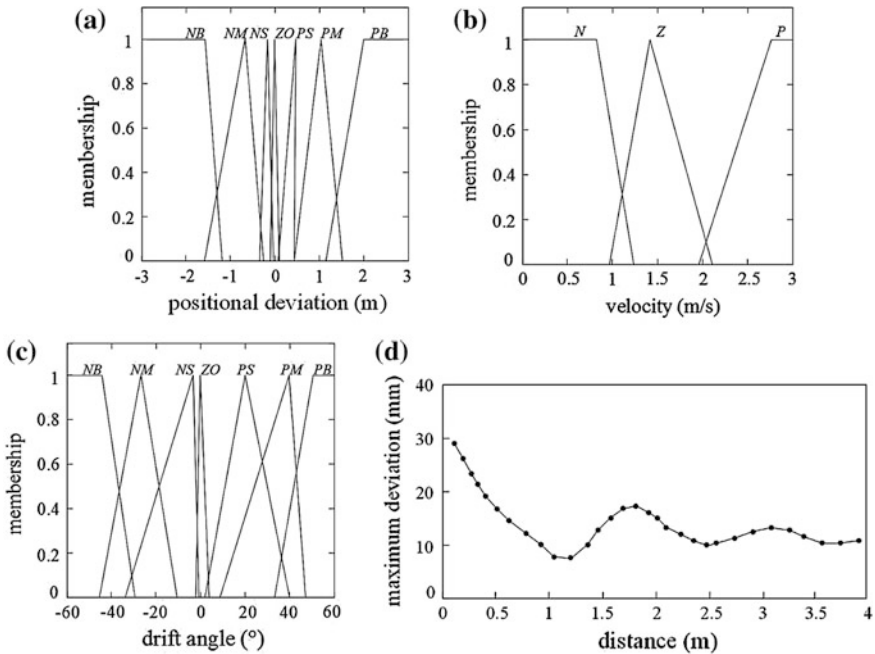


Fig. 6 Membership functions optimized by GA. **a** The function of position deviation, **b** the function of velocity, **c** the function of velocity, **d** the relationship between deviation and distance

Fig. 6a–c, respectively. The relationship between deviation and distance is shown in Fig. 6d.

In order to make the robot run along the ideal path, a review of the various control aspects has been conducted with particular emphasis upon intelligent decision-making capabilities. These capabilities include path planning model, decision-making module which is combined with FLC and GA. An account of the design of an intelligent controller for the autonomous mobile robot has been described.

To plan the ideal path, it is necessary to establish the model of CCD image transformation. The knowledge of planning path decisions is encapsulated in a fuzzy rule base. And then the FLC saves relevant feedback in the process of system running and updates data parameters to reconstruct the rule database, a compositional rule of inference is applied on the rule base to construct a set of applicable fuzzy decisions. The membership and the rule database of the FLC are optimized using method of GA. By the Fig. 6d, the robot will be close to the ideal path after trips of 1.5 m gradually, and the deviation can be kept in 10–20 mm. So the FLC based on GA can achieve high control precision.

Acknowledgement The authors wish to gratefully acknowledge the financial support provided for this study by National Key Technology Support Program (Grant No. 2015BAF20B02)

References

1. Bowditch RC et al (2012) Results of an Australian trial using SurePath liquid-based cervical cytology with Focalpoint computer-assisted screening technology. *Diagn Cytopathol* 40 (12):1093–1099
2. Pandharipande A, Caicedo D, Wang XY (2014) Sensor-driven wireless lighting control: system solutions and services for intelligent buildings. *IEEE Sens J* 14(12):4207–4215
3. Zhang M, Cheng WM, Guo P (2015) Intelligent recognition of mixture control chart pattern based on quadratic feature extraction and SVM with AMPPO. *J Coast Res* 73:304–309
4. Yan Z, Jouandeau N, Cherif AA (2013) A survey and analysis of multi-robot coordination. *Int J Adv Rob Syst* 10
5. Kang DW, Kang M, Hahn JW (2014) Measuring two-dimensional profiles of beam spots in a high-density spot array for a maskless lithography system. *Appl Opt* 53(36):8507–8513
6. Welch DA et al (2013) Simulating realistic imaging conditions for in situ liquid microscopy. *Ultramicroscopy* 135:36–42
7. Muksin U et al (2013) Three-dimensional upper crustal structure of the geothermal system in Tarutung (North Sumatra, Indonesia) revealed by seismic attenuation tomography. *Geophys J Int* 195(3):2037–2049
8. Kenkel NC (2013) Sample size requirements for fractal dimension estimation. *Commun Ecol* 14(2):144–152
9. Boukroune A, M'Saad M, Chekireb H (2010) Design of a fuzzy adaptive controller for MIMO nonlinear time-delay systems with unknown actuator nonlinearities and unknown control direction. *Inf Sci* 180(24):5041–5059
10. Kashki M, Abdel-Magid YL, Abido MA (2009) Application of novel reinforcement learning automata approach in power system regulation. *J Circ Syst Comput* 18(8):1609–1625
11. Agarwal PK, Chand S (2009) Fault-tolerant control of three-pole active magnetic bearing. *Expert Syst Appl* 36(10):12592–12604
12. Zhao Y et al (2015) A bistatic SAR image intensity model for the composite ship-ocean scene. *IEEE Trans Geosci Remote Sens* 53(8):4250–4258
13. Manik T, Holmedal B, Hopperstad OS (2015) Strain-path change induced transients in flow stress, work hardening and r-values in aluminum. *Int J Plast* 69:1–20
14. Castorena J, Creusere CD (2015) Sampling of time-resolved full-waveform LIDAR signals at sub-nyquist rates. *IEEE Trans Geosci Remote Sens* 53(7):3791–3802
15. Shao XX, Dai XJ, He XY (2015) Noise robustness and parallel computation of the inverse compositional Gauss-Newton algorithm in digital image correlation. *Opt Lasers Eng* 71:9–19
16. Reed KB, Okamura AM, Cowan NJ (2009) Modeling and control of needles with torsional friction. *IEEE Trans Biomed Eng* 56(12):2905–2916
17. Crotts APS, Hummels C (2009) Lunar outgassing, transient phenomena, and the return to the Moon. II. Predictions and tests for outgassing/regolith interactions. *Astrophys J* 707(2):1506–1523

Characterization of Noise Contaminations in Realistic Heart Sound Acquisition

Jun Huang, Booma Devi Sekar, Ran Guo, MingChui Dong
and XiangYang Hu

Abstract In practical clinical site, recording of heart auscultation signal is often challenged by the contamination of various non-cardiac noises. To address such key challenge, many researchers have developed various heart sound (HS) de-noising methods. Though, many of them on literature show promising results after adding Gaussian white noise to the ideal samples artificially and subsequently filter them out for performance evaluation. It has been proven that the noise which is recorded with the actual HS signal in clinical site does not pose the characteristics of Gaussian white noise. There is lack of study of true characteristics of site-sampled HS signal even it is fundamental and such important. As the first attempt, this paper investigates in depth the characteristics of several typical and common noise interferences that occur during HS clinical site acquisition. After summarizing the key features of such actual noises in time and frequency domains, a dynamic time warping based similarity algorithm is applied to indicate the destruction index of each noise type in contaminating the HS signal. The result show that lung sound and abdominal sound are the greatest disturbances in realistic HS acquisition, which should be brought to the forefront in designing the HS acquisition system as well as de-noising method.

Keywords Heart sound analysis · Noise characteristics · Coronary artery disease · Dynamic time warping

J. Huang (✉) · B.D. Sekar · R. Guo · M. Dong · X. Hu
Department of Electrical and Computer Engineering, University of Macau,
Avenida da Universidade, Taipa, Macau, China
e-mail: hjwongzeon@gmail.com

B.D. Sekar
e-mail: boomas@umac.mo

R. Guo
e-mail: grmust@gmail.com

M. Dong
e-mail: charley_dong@hotmail.com

X. Hu
e-mail: lovesea9999@hotmail.com

1 Introduction

With rapid development and popularization of wearable home health monitoring devices, the cardiovascular disease (CVD) diagnosis system based on biological signals such as sphygmogram (SPG), photoplethysmography (PPG), electrocardiogram (ECG), and heart sound (HS) signal has attracted many experts and scholars' attention. However, comparative study of these methods show that due to lack of qualified HS database, limited research has succeeded in developing heart health monitoring device based on HS signal analysis.

It has been observed that in practical clinical recording settings, HS acquisition is often challenged by various contaminations of non-cardiac noises including speech sound, lung sound, abdominal sound, ambient noise and electromagnetic noise etc. Recording HS signal with such noises not only results in the loss of pathological information but hinders further development of reliable HS analysis and diagnosis models. Thus to address such key challenges, many researchers have contributed towards the development of HS de-noising methods. It is noted that many such methods employ Gaussian white noise to contaminate ideal HS samples for simulation and performance evaluation. However, it has been proven by researchers [1] that the actual noise normally recorded with the HS signal in a clinical setting does not acquire any of the characteristics of Gaussian white noise. Literature also shows that there is a lack of research work relevant to the study of the characteristics of noises that actually contaminate HS signal. To support this further, it is worth noting that even an optimal wavelet de-noising method designed for eliminating Gaussian white noise has not achieved the desired results when tested in de-noising the real noise present in a HS signal [1, 2]. This indicates that the distinct characteristics of various noises could in fact make even an optimal de-noising method to achieve conflicting results.

In light of which, it is strongly believed that a deeper analysis of source and characteristics of practical noises is not only essential for sampling qualified HS signal, but also practically necessary for designing a more accurate CVD diagnosis system for home healthcare. Thus, this paper investigates the characteristics of some of the typical and common noise interferences that occur during clinical HS acquisition. After summarizing the key features of each noise type in time and frequency domain, additional experiments based on 500 HS samples acquired from hospital are conducted to reveal the distribution of each noise type. Subsequently, the Dynamic time warping (DTW) technique is employed to evaluate the individual destruction capability of each noise type.

2 Classification of Noise

Based on the analysis of nearly 500 HS samples acquired from patients in a hospital environment, we primarily classify the noise type present in HS signals as internal body noise and external body noise. Internal body noise includes speech sound (SS), lung sound (LS) and abdominal sound (AS), and whereas external body noises include ambient noise (AN) and electromagnetic noise (EN). The average spectrum profile of each noise interferences in frequency domain and their representative samples in time domain are shown in Fig. 1. The characterization of such noises will be theoretically described in this section.

2.1 Speech Interference

Speech interference refers to the speech sound (SS) that are specifically recorded from the patient during HS acquisition. As HS interfered by SS is rarely present in a

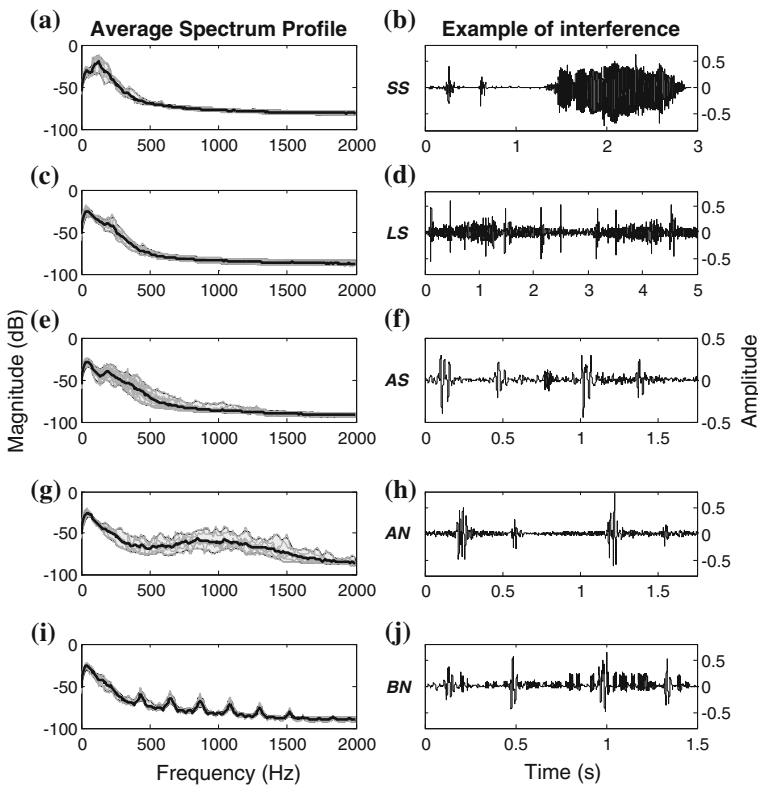


Fig. 1 Comparison of HS disturbed by different noise in frequency and time domain

well-controlled clinical recording settings, very few researches have taken efforts to study the characteristics of such interference. However study shows that HS sample are often disturbed by SS while examining the elderly patients or patients with delirium.

In fact SS acquired from the chest wall are filtered version of the ones originally generated by the vocal cords. Previous research [3] has proved that with the aid of air and parenchyma present in the lung, the chest could act as a low-pass filter to the SS. Furthermore, the air-born and structure-born mechanisms of sound transmission in lung are characterized by a fundamental resonance, whose frequency can be theoretically estimated as 110–150 Hz [4]. Experimental evaluation shows that, the fundamental frequency of SS is around 150 Hz for men and 230 Hz for women. This is actually further enhanced by the resonance as the overtone of SS decays exponentially below 700 Hz. The vibration of vocal cords is transmitted through the trachea and lung to the chest wall. In fact, despite of the low appearing probabilities of SS (3.6 % in this study), once if it does occur, its powerful resonance could severely disturb the HS component. This can be well evidenced in the time domain graph shown in Fig. 1b.

2.2 *Respiration Interference*

Lung sound (LS) is an unavoidable source of interference that overlaps with the frequency band of interest of the HS components. LS are generated by turbulent and vertical air flow within lung tracheal during both inspiration and expiration [5]. For patients suffering from respiratory disease, the LS is also accompanied by some adventitious sounds know as wheeze and crackle. These sounds would consequently disturb HS even more seriously.

As shown in Fig. 1d, LS normally has regular cycle with a longer duration which is approximately 3 times more than the cardiac cycle (0.6–1 s). In such case, when analyzing fewer HS cycle for diagnosis, the researchers quite often misdiagnose LS as heart murmurs. In frequency domain, the amplitude of LS exponentially decays with the frequency ranging from 75 to 2000 Hz [6]. Normal LS is continuous with dominant frequency ranging between 60 and 600 Hz. For abnormal LS, wheezes are continuous, longer than 250 ms, high pitched with dominant frequency of 400 Hz or more. Crackles are discontinuous, usually shorter than 20 ms with a wide spectrum of frequencies, between 200 and 2000 Hz [7].

The intensity of LS interference is related to the severity of subject's respiratory disease, as well as the position of HS acquisition. Experimental results in this study show that HS samples acquired form Pulmonary Area are most likely and whereas Mitral stenosis position are least likely to be affected by LS interference.

2.3 Abdominal Interference

Abdominal sound (AS) refers to the noises produced within the stomach and bowels. It is also interchangeably described as stomach rumble, bowel sound or peristaltic sound. AS generally is a rumbling, growling or gurgling noise produced by movement of the gastrointestinal tract, typically during digestion process, hunger or disease.

AS is a common interference (19.2 % in this study) that quite often contaminates the HS acquisition. It is much noticeable when recorded at the Mitral Area and the Tricuspid Area. According to the study conducted by researchers [8], the bowel sounds normally occur in the frequency range of 100–1200 Hz for duration of 5–200 ms with widely varying amplitudes. Some of the most common diseased AS generally occurs in the frequency range of 500–700 Hz with duration of 5–20 ms. The statistical result of 96 AS recordings in this study shown in Fig. 1e indicate that the distribution of AS appear in different frequency band. It also illustrates that the AS interference dominants at a frequency range of 100–700 Hz, which very clearly overlaps with the HS components.

2.4 Ambient Interference

Ambient noise (AN) refers to all kinds of sounds that are recorded from the HS acquisition environment. It typically refers to the noise which is recorded from other patients or health care personnel.

It is generally understood that the AN would travel through air and would be recorded during HS sampling. However, according to [9] when HS signal is recorded using the stethoscope, the chest acts as a collector of AN propagated through the air. This study demonstrates that when the AN is sampled via the chest path to the microphone, it is recorded with a slight amplification of about 6.8 %. Therefore, AN is one of the important interference one has to consider while developing a HS de-noising method.

2.5 Electromagnetic Interference

Electromagnetic noise (EN) is a type of interference which quite common occurs during HS sampling, but can be easily neglected with some simple techniques. The most typical EN is known as the ‘Bumblebee Noise’, which is caused by the changing electromagnetic field during transmission of Global System for Mobile Communications (GSM) mobile phone. This induces a varying current in the electronic components of the instrument used for recording HS signal [10].

Comparison study between HS with and without interference by an electromagnetic source (GSM phone call) placed at a 3 m distance is made in Fig. 1j. The intensity of EN is related to the distance from where the electromagnetic source is placed. It is noted that the HS sample is severely disturbed by an ordinary phone call even if it is received at a sufficient distance away from the HS sampling location. According to the Time-Division Multiple Access (TDMA) technique used in GSM, at every eight bursts period, a TDMA frame is formed with duration of $120/26 \approx 4.615$ ms [11]. Thus, in frequency domain the fundamental switching rate is between $1/4.615 \approx 217$ Hz. This coincides with GSM's working principle and as observed in Fig. 1i, the EN is formed by a fundamental wave of 217 Hz and its harmonics occurs at all integer multiples of 217 Hz.

Though, an ideal solution would be to turn off the electromagnetic interference sources while HS sampling, practically it is not always feasible. Therefore, researchers have developed algorithm designed based on prior knowledge of TDMA and Active Noise Control (ANC) [12, 13] at hardware level that cancels out the electromagnetic interferences. On the other hand, for electronic stethoscope with no access to the internal data sending structure of the GSM mobile, a notch filter at software level based on the frequency feature mentioned above is still the most straight-forward solution [14].

3 Methods

3.1 Heart Sound Dataset

The HS dataset were obtained from the 5th Affiliated Hospital of Sun Yat-Sen University, Zhuhai, China, using the ds32a+ digital stethoscope of ThinkLabs Inc. Totally 500 samples from 300 subjects were acquired and analyzed in the current study. All HS samples were recorded by following the same sampling criteria: adopt lying posture, record from Mitral stenosis position, stored as WAV format at 4000 Hz sampling rate with 16-bit resolution. The HS dataset were recorded under the same clinical environment at different time interval within 3 months period. As this basically satisfied the principle of random sampling, the acquired samples were interfered with various noises presented above.

3.2 Evaluation Method

Inspired by the similarity comparison technique widely applied in speech identification, the classical DTW algorithm is proposed to evaluate the distortion of HS signal due to different noises. Subsequently the destruction index (DI) is evaluated for each noise type.

The DTW algorithm is one of the most popular mathematical models used in the field of speech recognition and handwriting identification [15]. Its main concept is to find out the minimum characteristic values of the time calibration path between two temporal sequences. It is the non-linear alignment which produces a more intuitive similarity measure, allowing similar shapes to match even if the sequences are of different lengths and are out of phase in the time axis.

Suppose there are two sequences $A := (a_1, a_2, \dots, a_m, \dots, a_M)$ and $B := (b_1, b_2, \dots, b_n, \dots, b_N)$ of length M and N respectively. The distance $D_{(m,n)}$ between the two elements a_m and b_n can be expressed by Euclidean distance measure using Eq. 1

$$D_{(m,n)} = (a_m - b_n)^2 \tag{1}$$

Thus, the total distance $T_{(A,B)}$ between two segment samples can be computed using Eq. 2

$$T_{(A,B)} = \sum_{n=m=1}^N D_{(m,n)} = \sum_{n=m=1}^N (a_m - b_n)^2 \tag{2}$$

The aim of DTW algorithm is to find the best path function $m_i = f(n_i)$, which makes the total Euclidean distance to be minimum from N to M as in Eq. 3

$$DTW_{(A,B)} = \text{Min} \sum_{\substack{n_i = 1 \\ m_i = f(n_i)}}^N (a_m - b_n)^2 \tag{3}$$

Here, $DTW_{(A,B)}$ is the similarity result between two different sequences.

In this paper, each testing group includes two clean segments: C_1, C_2 and one disturbed segment D . Each testing sample contains two HS cycles, which are extracted from a HS sample disturbed by different noises. The basic principle is that different noises have varying influence on the HS signal, therefore a higher similarity between clear and disturbed segments reflect the lower destruction capability of the noise. In contrast, the lower similarity between the two segments would reflect the higher destruction capability of the noise. To reduce the error caused by individual difference, the destruction index (DI) is computed using Eq. 4

$$DI = \frac{DTW_{(D,C_2)}}{DTW_{(C_1,C_2)}} \tag{4}$$

Here the DI reflects the individual destruction capability of noise.

4 Result and Discussion

500 HS samples acquired from patients in a hospital in China were considered in this study. The statistical distribution of each interference category is 3.6 % (18/500) SS, 11.2 % (56/500) LS, 19.2 % (96/500) AS, 16.8 % (84/500) AN, 8.8 % (44/500) EN.

4.1 Noise Characteristics

As shown in Fig. 1, SS, LS, and AS are low frequency noises whose spectrum content predominantly concentrates in the frequency range below 600 Hz. Since the frequency and amplitude distribution of these noises overlap with normal HS components, the shape of HS signal are easily distorted in time domain. This results in poor performance of shape-based analysis methods such as morphological envelope extraction [16]. On the other hand, the frequency of AN and EN are mainly distributed above 500 Hz. According to the Equal loudness curves shown in Fig. 2, human ears are most sensitive at a higher frequency around 1000 Hz. Therefore, AN and EN would be heard clearly by human ears through a stethoscope and sometimes such noise could be even louder than the HS signal despite the intensity of AN and EN are generally lower than HS components, and the shape of HS doesn't seem to be seriously distorted. As the AN and EN could have great influence on auditory effect in the heart auscultation recording, they should be considered in designing the HS de-noising methodology.

Fig. 2 Equal loudness curves of human auditory system

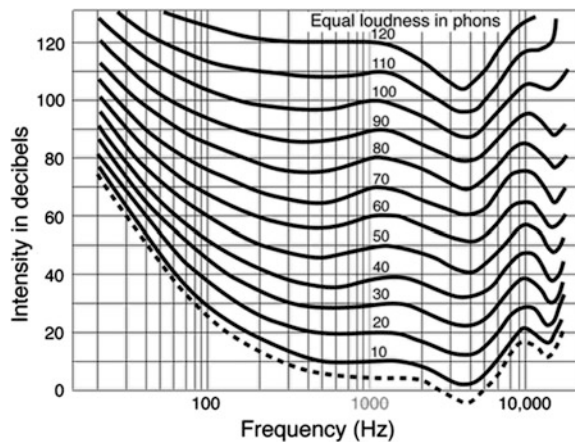


Table 1 Statistics of destruction index (DI) per sound group

	Sample_1	Sample_2	Sample_3	Sample_4		Avg.
SS	3.07 (6.03/1.96)	2.88 (34.74/12.04)	4.26 (8.36/1.96)	4.6 (17.91/3.9)	...	4.49
LS	3.91 (3.57/0.91)	2.12 (3.55/1.67)	3.21 (3.48/1.09)	2.47 (22.98/9.29)		3.18
AS	2.33 (2.18/0.94)	1.47 (5.99/4.07)	4.36 (8.13/1.87)	1.27 (10.08/7.95)		3.80
AN	2.49 (13.75/10.01)	1.43 (14.71/10.26)	3.34 (16/4.79)	1.84 (10.78/5.85)		2.84
EN	1.29 (5.8/4.49)	3.89 (28.65/7.37)	2.19 (9.47/4.32)	2.07 (12.05/5.81)		2.08

4.2 Noise Destruction Capability

The result of distortion evaluation is shown in Table 1. It can be observed that, although the denominator greatly varies between different samples, the DI in each category is relatively stable. This fact verifies the feasibility of proposed evaluation method. According to the statistics result shown in Table 1, the order of DI for each kind of noise is: SS > AS > LS > AN > EN, which reflects individual destruction capability of each noise. SS is the most harmful noise type which disturbs HS component, but its probability of occurrence is quite small. EN is least damaging to the HS, it also has a smaller probability of occurrence. LS and AS are internal bio-signals that cannot be avoided, both of them have high probability of occurrence and high DI value. Thus, LS and AS should be brought to the forefront of HS acquisition and de-noising. Considering that in this study, all the samples were acquired at a clinical setting, the value of DI is quite low. However, if the HS acquisition was to be conducted at outdoors, the DI of AN would become much higher, an efficient AN suppression method would then seem absolutely necessary.

5 Conclusion

HS is an auscultation signal from the heart which could reflect the health condition of the cardiac system. However, HS acquisition is highly vulnerable to interference due to internal and external body noises. This often results in loss of pathological information and false diagnosis in further developed diagnosis model based on HS signal. In this paper, some typical and common noise interference has been analyzed using Similarity algorithm based on DTW technique. Based on our experimental results, the characteristics of each noise contamination are summarized in Table 2. The results show that LS and AS should be brought to the forefront in designing the HS acquisition system as well as de-noising method.

It should be mentioned that the technical evaluation of the type of noise that actually corrupts HS recording presented in this paper could help researchers to consider the characteristics of different noise type in designing the HS de-noising method. This in-turn could further elevate the performance level of the diagnosis model based on HS signal. Also according to the experimental results presented, it

Table 2 Characterization of each noise contamination

Noise	Frequency range	Duration	Percentage of disturbance in HS (%)	Destruction index
Speech sound (SS)	110–700 Hz	Random	3.6	4.49
Lung sound (LS)	75–2000 Hz	0.6–1 s	11.2	3.18
Abdominal sound (AS)	100–1200 Hz	5–200 ms	19.2	3.80
Ambient noise (AN)	Whole range	Random	16.8	2.84
Electromagnetic noise (EN)	Integer multiples of 217 Hz	Random	8.8	2.08

is observed that the de-noising method would be more efficient if it could be designed to treat the HS signals in smaller segments instead of the whole signal.

Finally, it should be emphasized that the research study conducted in this paper has opened scope for our research team to design a wearable home health monitoring device that could provide feedback and warning messages to user about the type of noise present. This could greatly help user to avoid the interference source during HS acquisition.

Acknowledgement This research was supported by the Science and Technology Development Fund (FDCT) of Macau S.A.R with project ref. No. 016/2012/A1 and also by Research Committee of University of Macau under Grant No. MRG005/DMC/2015/FST.

References

1. Gradolewski D, Redlarski G (2014) Wavelet-based denoising method for real phonocardiography signal recorded by mobile devices in noisy environment. *Comput Biol Med* 52:119–129
2. Messer SR, Agzarian J, Abbott D (2001) Optimal wavelet denoising for phonocardiograms. *Microelectron J* 32(12):931–941
3. Pasterkamp H, Kraman SS, Wodicka GR (1997) Respiratory sounds: advances beyond the stethoscope. *Am J Respir Crit Care Med* 156(3):974–987
4. Korenbaum VI (1999) Protection of acoustic devices against near field own noise. Pacific Oceanological Institute, Vladivostok, p 32
5. Hardin JC, Patterson JL (1979) Monitoring the state of the human airways by analysis of respiratory sound. *Acta Astronaut* 6(9):1137–1151
6. Gavriely N, Nissan M, Rubin AH, Cugell DW (1995) Spectral characteristics of chest wall breath sounds in normal subjects. *Thorax* 50:1292–1300
7. Chang GC, Cheng YP (2008) Investigation of noise effect on lung sound recognition. In: *Machine learning and cybernetics, 2008 international conference on IEEE*, vol 3, pp 1298–1301
8. Bray D, Reilly RB, Haskin L (1997) Assessing motility through abdominal sound monitoring. *Engineering in Medicine and Biology Society*. In: *Proceedings of the 19th annual international conference of the IEEE*, vol 6, pp 2398–2400

9. Schmidt S, Zimmermann NH, Hansen J (2012) The chest is a significant collector of ambient noise in heart sound recordings. In: Annual computing in cardiology conference. CinC, pp 741–744
10. Harrison P (2007) GSM interference cancellation for forensic audio: a report on work in progress. *Int J Speech Lang Law* 8(2):9–23
11. Sempere JG (1997) An overview of the GSM system. In: IEEE Vehicular Technology Society, pp 1–33
12. Kuo SM, Morgan DR (1999) Active noise control: a tutorial review. *Proc IEEE* 87(6):943–973
13. Claesson I, Nilsson A (2003) GSM TDMA frame rate internal active noise cancellation. *Int J Acoust Vib* 8(3)
14. Claesson I, Rossholm A (2005) Notch filtering of humming GSM mobile telephone noise. In: Information, communications and signal processing. 2005 fifth international conference on IEEE, pp 1320–1323
15. Müller M (2007) Dynamic time warping. In: Information retrieval for music and motion, pp 69–84
16. Choi S, Jiang Z (2008) Comparison of envelope extraction algorithms for cardiac sound signal segmentation. *Expert Syst Appl* 34(2):1056–1069

Independent Component Analysis of Space-Time Patterns of Groundwater System

Chin Tsai Hsiao, Jui Pin Tsai and Yu Wen Chen

Abstract This study proposed a method based on Independent Component Analysis (ICA) to understand the mechanisms that cause regional groundwater head variations. To verify the capability of the proposed method, this method is applied to an ideal numerical groundwater model, which was developed by using MODFLOW. The unconfined aquifer parameters are set as homogeneous and isotropic. The values of the two groups of pumpages (sinks) and one rainfall recharge (sources) were time-variant, and the frequencies among the three sink/sources were different. The simulated heads were sampled from 64 selected observation wells within the model boundary with a daily time step for 5 years. The simulated heads of the 64 wells were inputted to ICA. The study results show that the ICA can successfully decompose the sampled heads into three independent components (ICs) resulted from the three sink/source. To identifying the physical meanings of the three ICs, the correlation coefficients between ICs and the three sinks/sources were computed, and their values are 0.9816, 0.888 and 0.684, respectively. The separating matrix of ICA was also used to identify the pumping well locations. The study results show that the proposed method provides a novel and efficient method to understand the spatiotemporal head variations of groundwater system and can be used to locate the pumping wells, which is crucial for the regional groundwater management.

C.T. Hsiao

Department of Information Management, Chung Chou University of Science and Technology, Changhua County, Taiwan, R.O.C.
e-mail: cthsiao.hsiao@gmail.com

J.P. Tsai (✉)

Department of Hydrology and Water Resources, The University of Arizona, 1133 E. James E. Rogers Way, 214, Harshbarger Bldg 11, Tucson, AZ 85721, USA
e-mail: skysky2cie@gmail.com

Y.W. Chen

Department of Civil Engineering, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C.
e-mail: bsjacky@gamil.com

Keywords Independent component analysis · Principal component analysis · Fourier transform · Groundwater system

1 Introduction

Head variations of a groundwater system are affected by many factors, such as rainfall, tide, river, artificial pumpage and so on. These factors make the head periodically change. Once the effect of each factor on the head variation can be extracted and qualified, we could propose better strategy for the regional groundwater management. There are a variety of methods to extract useful information from a messy signal. Longuevergne et al. [1] combines the Karhunen–Loève transform (KLT) with the kriging method to extract regional information from the head observations of 195 wells with a period of 17 years in the French and German area of the Rhine valley. Yu and Chu [2, 3] performed rotated empirical orthogonal function (REOF) analysis to analyze monthly head observations from 66 wells located in Taiwan’s Choshui River alluvial fan during 1997–2002. Results show that the first EOF contributes most to the spatiotemporal head changes of the shallow aquifer of the Choshui River alluvial fan. Likewise, the method of REOF is applied to decompose the space-time head variations and used to determine the potential recharge zones by investigating the correlation between the identified groundwater signals and the measured rainfall data [3]. Basically, both EOF and KLT belong to the category of PCA. These methods can convert the high-dimensional data to low dimensional principal components, and achieve the effect of the purpose of data compression.

PCA components are linear independence in terms of statistical properties. However, in statistic, ‘linear independence’ is totally different from ‘independence’. The independence denotes no correlation among the ICA components. Their joint probability (Pr) must equal to the product of their corresponding individual probability (i.e. For two ICA components, $\Pr(A \cap B) = \Pr(A)\Pr(B)$). However, the PCA components cannot satisfy the definition of independence. Therefore, PCA components are not independent from each other.

Independent component analysis (ICA) is a statistical technique for extracting a mixed signal to the independent signals. ICA can be seen as an extension to principal component analysis (PCA), and its implement process similar to data mining (DM), neural networks (NN), machine learning (ML) and blind source separation (BSS). However, ICA is a much more powerful technique than PCA in terms of signal extraction because ICA can extra the independent signals from a mixed signals. ICA has been widely used in various fields, including digital images, document databases, economic indicators, psychometric, etc. Yang [4] applied ICA to analyze 21 American hydrologic basins’ water quality data. The results show that the independence of the independent components is the functions of water quality variables and weights of independent components. Wang [5] Combined ICA with numerical groundwater model to identify the locations and amount of the local

$$\tilde{\mathbf{s}} = \mathbf{W}(\mathbf{x} - \boldsymbol{\mu}) \quad (1)$$

where \mathbf{W} is the weight matrix, $\boldsymbol{\mu} = \mathbf{E}\{\mathbf{x}\}$ are the expectation of \mathbf{x} , $\tilde{\mathbf{s}}$ are the principal components. According to Eq. 1, all the principal components are orthogonal to each other, and their variance equals to that of the original dataset. In practical application, the new variables with smaller variance can be seen as noise, which can be ignored. Therefore, the dimension of the original dataset can be reduced. When conducting PCA, the eigenvalue decomposition of the covariance matrix $\boldsymbol{\Sigma}_x$ of \mathbf{x} can be expressed as follows

$$\mathbf{E}\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\} = \boldsymbol{\Sigma}_x = \mathbf{E}\mathbf{D}\mathbf{E}^T \quad (2)$$

where \mathbf{E} is a matrix of eigenvector of the $\boldsymbol{\Sigma}_x$ with the property as

$$\mathbf{E}\mathbf{E}^T = \mathbf{E}^T\mathbf{E} = \mathbf{I} \quad (3)$$

\mathbf{I} is an identity matrix, \mathbf{D} is a matrix containing the descending eigenvalues of the $\boldsymbol{\Sigma}_x$ along the diagonal and zeros elsewhere. The weight matrix can be expressed as

$$\mathbf{W} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T \quad (4)$$

On a basis of Eq. 1, the expectation and covariance of the principal components $\tilde{\mathbf{s}}$ can be calculated

$$\mathbf{E}\{\tilde{\mathbf{s}}\} = \mathbf{E}\{\mathbf{W}(\mathbf{x} - \boldsymbol{\mu})\} = \mathbf{W}\mathbf{E}\{\mathbf{x} - \boldsymbol{\mu}\} = 0 \quad (5)$$

$$\begin{aligned} \mathbf{E}\{\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T\} &= \mathbf{E}\{\mathbf{W}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\mathbf{W}^T\} \\ &= \mathbf{W}\mathbf{E}\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\}\mathbf{W}^T = \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T = \mathbf{I} \end{aligned} \quad (6)$$

According to Eq. 6, the variables ($\tilde{\mathbf{s}}$) are uncorrelated and whitening.

2.2 Independent Component Analysis, ICA

ICA is a method for searching the underlying factors or components from multivariate (multidimensional) data. Figure 2 shows the concept model of ICA.

The observed data $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)^T$ are formulated as a linear combination of components $\mathbf{s} = (s_1, s_1, \dots, s_n)^T$ that are statistically independent:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (7)$$

\mathbf{A} is a mixing matrix. ICA is searching for a separating matrix \mathbf{W} so that

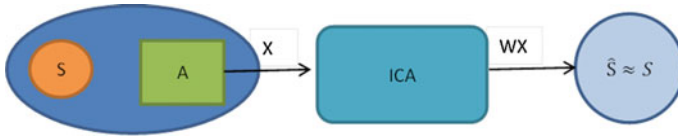


Fig. 2 Concept model of ICA

$$Wx = A^{-1}x = \hat{s} \tag{8}$$

\hat{s} denotes the estimates of the independent components. When conducting ICA, PCA can be seen as the preprocessor to reduce the dimension of Σ_x . Besides PCA, centering and whitening are also applied in the ICA. The difference between ICA and PCA is that ICA considers non-Gaussian data structure. Therefore, the higher-order statistical information can be utilized, and the ICs can be independent to each other, which is not possible by PCA. The analysis quality of ICA method depends on both of the objective function and the optimization algorithm. Ideally, these two classes of properties are independent in the sense that different optimization methods can be used to optimize a single objective function, and a single optimization method can be used to optimize different objective functions [7]. ICA is a popular method and has much well known software such as FastICA, JADE, etc. In this study, we used FastICA to develop our methodology. The algorithm of FastICA is developed based on a fixed-point iteration scheme by maximizing non-Gaussianity as a measure of statistical independence. Non-Gaussianity of the ICs is necessary for defining the identifiability of the Eq. (8). The classical measure of nongaussianity is kurtosis, defined as the fourth-order moment ($E\{s^4\}$) minus three times of the square of the second-order moment ($E\{s^2\}$), given by:

$$\text{kurt}(s) = \sum_{i=1}^n E\{s_i^4\} - 3(E\{s_i^2\})^2 \tag{9}$$

Kurtosis has disadvantages in practice when its value has to be estimated from a measured sample. The main problem is that it can be very sensitive to outliers [8]. Thus, kurtosis is not a robust measure of nongaussianity. The measure of non-gaussianity is given by negentropy J , which is defined as

$$J(y) = H(y_{gaussian}) - H(y) \tag{10}$$

where $y_{gaussian}$ is a Gaussian random vector of the same covariance matrix as y . $H(\cdot)$ is the entropy of random vector y . When the random vector y is in Gaussian distribution, the $H(y_{gaussian})$ will be maximum. While the y deviates more than from the Gaussian distribution, the smaller the $H(y)$ value. As a result, the negentropy J is always non-negative, and its value equals to zero if and only if y has a Gaussian

distribution. Thus, the maximum $J(y)$ can be obtained if the y deviates from the Gaussian distribution. Owing to the Eq. (10) is too difficult to calculate, Hyvärinen [9] proposed a simplified calculation formula:

$$J(y_i) \approx c[E\{G(y_i)\} - E\{G(v)\}]^2 \quad (11)$$

where G is a non-quadratic function, c is an irrelevant constant, and v is a Gaussian variable of zero mean and unit variance (i.e., standardized). Hyvärinen [9] proposed the following choices for the G function:

$$G_1(y) = \frac{1}{a_1} \log \cosh(a_1 y), \quad 1 \leq a_1 \leq 2 \quad (12)$$

$$G_2(y) = -\frac{1}{a_2} \exp\left(-\frac{a_2 y^2}{2}\right), \quad a_2 \approx 1 \quad (13)$$

$$G_3(y) = \frac{1}{4} y^4 \quad (14)$$

In this study, FastICA 2.5, which was proposed as a Matlab tool by Department of Information and Computer Science (Aalto University), are used to analyze the data. The reader with interest in fixed-point iteration scheme can see the papers [9, 10] for further recognition of FastICA.

3 Numerical Experiment

3.1 Development of the Ideal Case

To verify the decomposition ability of ICA, this study conducted an ideal groundwater simulation model for the numerical experiment. The ideal case with transient simulation is shown in Fig. 3. The aquifer is assumed to be homogeneous, isotropic and unconfined. The aquifer's hydraulic conductivity is 50 m/day and storage coefficient is 0.01. The designed simulation area is 20,000 m \times 20,000 m with a grid size of 1000 m \times 1000 m resolution. There were 64 observation wells and 2 groups of pumping wells within the study area. Boundary conditions in the west and east sides were set as constant-head condition with 19.5 and 0.5 m, respectively and in the north and south boundary are set as no-flow boundary conditions. The ground surface decreased from east side (20 m) to the west boundary (1 m). The thickness of the aquifer ranged from 101 to 120 m. The initial water levels were set as 1 meter below the ground surface in the entire simulated area.

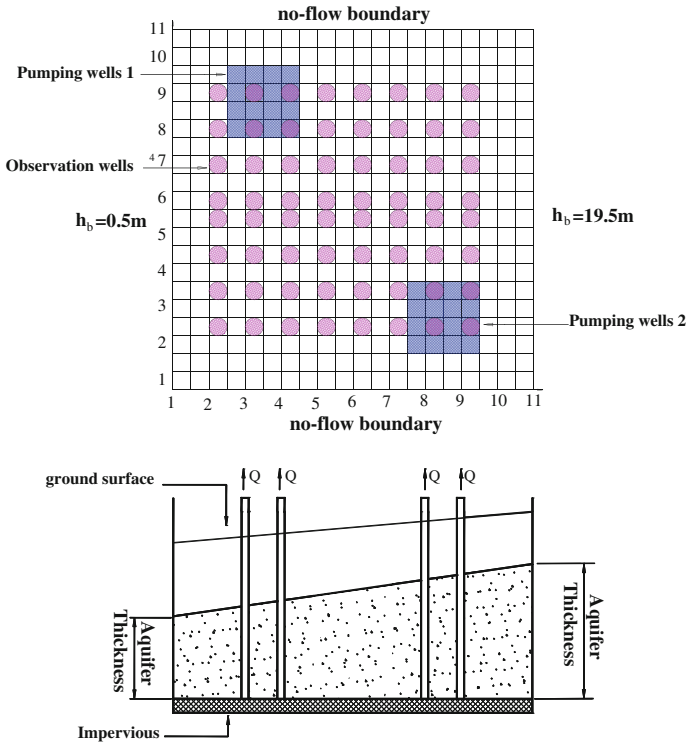


Fig. 3 Ideal case of groundwater simulation

3.2 Observation Data Created by Modflow

Firstly, this study created 1825 points input data, including a set of temporal rainfall data and 2 sets of temporal pumping data, for the MODFLOW simulation. To facilitate the case study, the frequency of the rainfall data was set to 1 with 1/year, the frequency of the pumping well group 1 was set to 120/year, and the frequency of the pumping well group 2 were set to 72/year. The input data are shown in Fig. 4.

When the input data were created and inputted into MODFLOW, the simulation heads were used as the head observation data. ICA decomposes the observation data at the selected 64 monitoring wells into its ICs. When ICA obtained the three ICs, this study compared the differences between the three input data and the ICs by using Fourier transform and correlation coefficient analysis. The separating matrix of ICA is used to identify the locations of the source signals. On a basis of the numerical experiment, this study verifies that ICA could be used to decompose the spatiotemporal signals of groundwater level fluctuations.

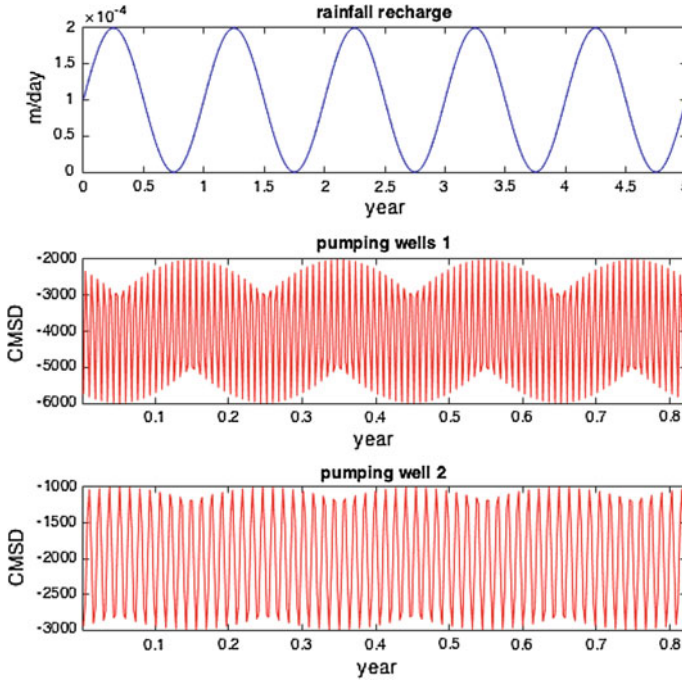


Fig. 4 The temporal pattern of the recharge and pumpage data of the ideal case

3.3 The Results of ICA

ICA has the following ambiguities. First, it cannot determine the energies of the ICs and leaves the ambiguity of the sign by multiplying the ICs with -1 without affecting the model. Second, the independent component can be estimated one by one but we cannot determine the order [10]. The ambiguities of ICA have been examined by applied FastICA 2.5 to decompose the spatiotemporal data observed at 64 observation wells in the ideal case. What we can do to overcome the ambiguities of ICA is exciting the model more than one times. Fortunately, the algorithm is efficient [9, 10] and the results can be identified easily after exciting FastICA multiple times. A key point to execute FastICA is to identify how many ICs are required and how many principal components retained at PCA. In this ideal case, the observation data obviously consisted of three ICs. Therefore, in our case study, 3 ICs were selected, and 4 principal components are retained from PCA based on trial and error method. The retained 4 principal components represent 99.99 % of the total amount of variance in the spatiotemporal head observations, and the 3 ICs are shown in Fig. 5.

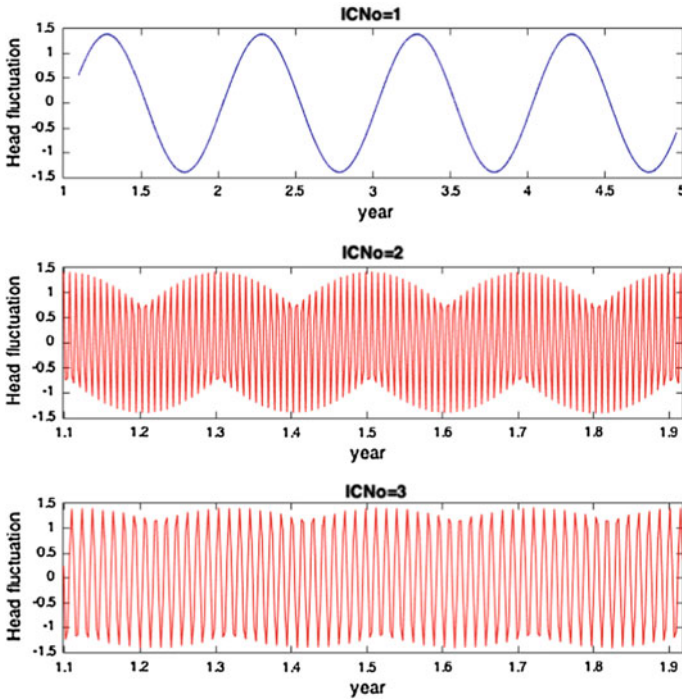


Fig. 5 The independent components obtained by ICA

In the ideal case, groundwater level fluctuations were resulted from rainfall and well pumping. According to the comparison of Figs. 4 and 5, we suppose that the IC 1 is consistent with the pattern of rainfall, and the IC 2 and 3 are consistent with the pattern of pumping well group 1 and 2. To further verify our supposition, we calculating the Fourier transform and correlation coefficient for the ICs and designed recharge and pumpages data. Figure 6 shows the frequency of the ICs. According to Fig. 6, the frequencies of the IC 1, 2, and 3 are 1/year, 120/year and 72/year, respectively. The frequencies of the ICs are consistent with that of recharge and pumpage, as shown in Fig. 4. Table 1 shows the correlation coefficient between ICs and three input data. From the data of Table 1, the correlation coefficients between the IC 1, 2, and 3 and their corresponding input recharge and pumpage data are 0.9816, 0.888 and 0.684, respectively. In addition, the correlation coefficients are very small between the ICs and the uncorrelated signal sources. By the analysis of correlation coefficient, it can be verified that ICA can successfully decompose the head observations into independent ICs.

From Eq. 8 ($\hat{s} = \mathbf{W}\mathbf{x}$), the separating matrix \mathbf{W} estimated by ICA was applied to transfer the observed signal \mathbf{x} (groundwater heads) back to the original signal \hat{s} (the head fluctuation made from recharge and pumpage). Notice that each IC represent a original signal source and has its physical meaning such that recharge and

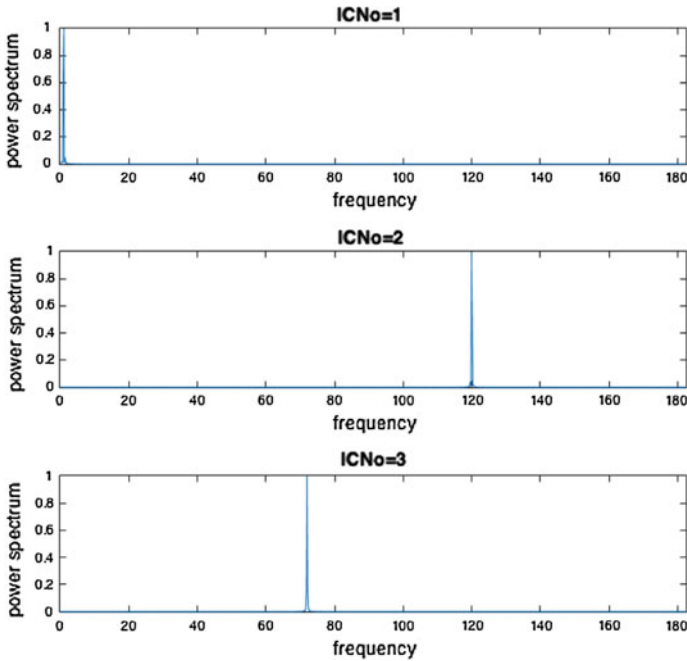


Fig. 6 The frequency of the independent components

Table 1 The correlation coefficient between ICs and three input data

	IC1	IC2	IC3
Rainfall	0.9816	0.0018	0.0003
Pumping wells 1	0.0028	0.8883	0.0048
Pumping wells 2	0.0018	0.0021	0.6841

pumpage. The physical meaning of each row of the separating matrix **W** represents the contribution levels to head fluctuation resulted from an original signal (for a specific IC) in space. Figures 7 and 8 show the contour of separating matrix of IC 2, 3 respectively. From Fig. 7, the spatial distribution of separating matrix of the IC 2 is distributed on the left side of the study area, which is the location of the pumping well group 1 (the source of IC 2). Similarly, from Fig. 8, the IC 2 is created by the pumping well group 2.

As a result, we can confirm that IC 1, 2, 3 represent the signals created by rainfall, pumping well group 1 and pumping well group 2, respectively. On a basis of the numerical experiment results, this study proves that ICA has high potential to be used in the field study to understand the spatiotemporal groundwater head fluctuations.

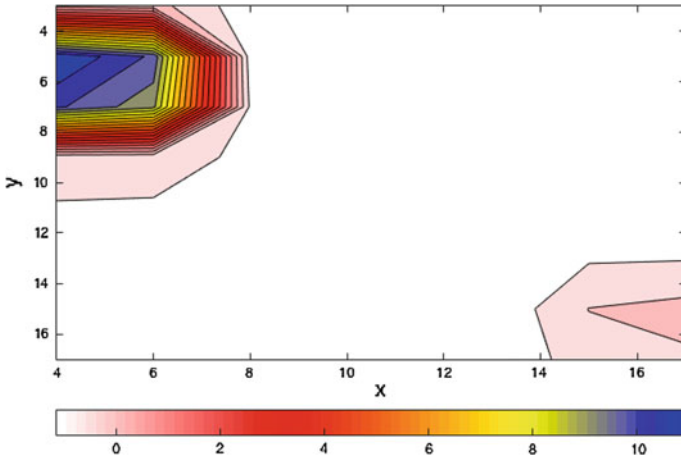


Fig. 7 The contour of the separating matrix of IC 2

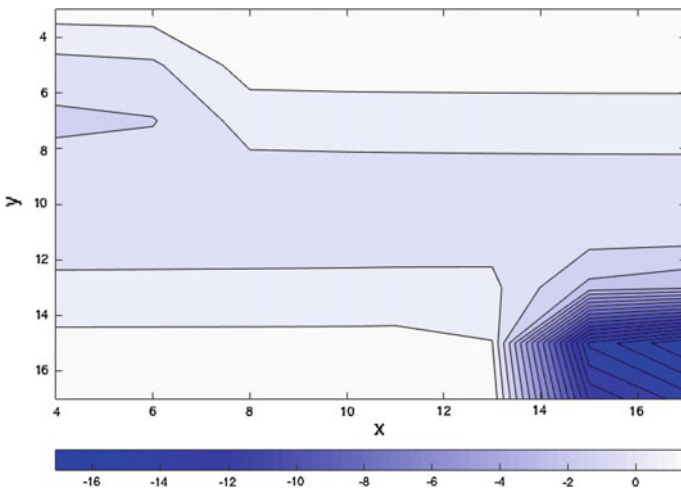


Fig. 8 The contour of the separating matrix of IC 3

4 Conclusion

This study successfully applied ICA to decompose the spatiotemporal groundwater head data. A numerical experiment is designed to demonstrate the capability of the proposed method. According to the comparison of ICs and their original signals by using Fourier transform and correlation coefficient analysis, we can confirm that the ICs are consistent to their original signals. The separating matrix of ICA identifies the spatial locations of the source signals. This study shows that one of the most

valuable features of ICA is its ability to clearly identify the locations of each signal source. These findings provide novel information, which can be widely used for groundwater management in the future.

References

1. Longuevergne L, Florsch N, Elsass P (2007) Extracting coherent regional information from local measurements with Karhunen-Loève transform: case study of an alluvial aquifer (Rhine valley, France and Germany). *Water Resour Res* 43:04430
2. Yu H-L, Chu H-J (2010) Understanding space–time patterns of groundwater system by empirical orthogonal functions: a case study in the Choshui River alluvial fan, Taiwan. *J Hydrol* 381(3):239–247
3. Yu H-L, Chu H-J (2012) Recharge signal identification based on groundwater level observations. *Environ Monit Assess* 184(10):5971–5982
4. Yang S-H (2001) Water river pollution sources separation using independent component analysis, in Department of Geography. National Taiwan University, Taipei, p 320
5. Wang R-F (2011) Local pumping source identification and pumping estimation of groundwater system, in Department of Civil Engineering. National Taiwan University, Taipei, p 86
6. Liu H-J, Hsu N-S (2015) Novel information for source identification of local pumping and recharging in a groundwater system. *Hydrol Sci J* 60(4):723–735
7. Hyvärinen A, Karhunen J, Oja E (2004) Independent component analysis. Wiley, Hoboken
8. Huber PJ (1985) Projection pursuit. *Ann Stat* 13(2):435–475
9. Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw* 10(3):626–634
10. Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. *Neural Netw* 13(4–5):411–430

Analysis of the Status Quo of MOOCs in China

Li Hao, Xinghua Sun, Chunlei Zhang and Xifeng Guo

Abstract MOOCs (Massive Open Online Courses), is a new network teaching mode which is currently popular in the world. Its specific features such as openness and massiveness are conducive to the reform and development of Chinese University Education. From the original theory of MOOCs, its connotation, features and types are explored, which helps to learn from the excellent practical experiences from foreign countries. The analysis of the advantages and limitations of the MOOCs can help us make the dialectical understanding of it, make full use of its positive effects on Higher Education in China, and try to avoid its negative effect.

Keywords Moocs · Special features · Merit · Limitations · Chinese higher education

(1) Major Research Projects of Hebei North University (ZD201303); (2) Teaching Reform Project of Hebei North University (JG201551).

L. Hao

Department of Political Science and Law, Hebei North University,
Zhangjiakou, Hebei, China
e-mail: haoli_03@163.com

X. Sun (✉) · X. Guo

School of Information Science and Engineering, Hebei North University,
Zhangjiakou, Hebei, China
e-mail: sunxinghua08@gmail.com

C. Zhang

Shuyuanxiang Primary School, Zhangjiakou, Hebei, China

1 Introduction

With the rapid development and popularization of Internet technology, almost all industries are affected to some degree. Education industry is no exception. The popular “Mu class” (MOOCs) is a notable case of traditional education influenced by Internet technology.

2 MOOCS and Its Special Features and Types

MOOCs is a outside vocabulary. It’s full name is Massive Open Online Courses. It means a large, open, online course. MOOCs were first proposed by American Bryan Alexander and Dave Cormier. Then Stephen Donis uses it in a large network course in 2008.

2.1 *The Definition of MOOCs*

As for the definition of MOOCs, the education circle has not formed a unified view at present. The author thinks that the following definition is reasonable. MOOCs is defined with free choose, open registration and open architecture of online course. It integrates social networks with the available network resources and is set up and pushed by experts from the research area.

2.2 *The Special Features of MOOCs*

From the literal analysis, we will find that MOOCS has three features: Massive, Online, Open.

1. Massive is namely large-scale. Students enrolled in the MOOCs are many. Even individual MOOCs courses are attracting tens of thousands of students. The traditional curriculum only accommodating dozens, hundreds of people is completely unable to compare with it. According to statistics, a course with the most registered students was learned by 160,000 people from 190 countries. Because of many students in class, when one student brings forward a problem, somebody answers in 15 min. This timely response is a kind of motivational motivation for students. Due to the large number of students, even though the proportion of people posting is not large, answers are sufficient to cover a variety of views in the MOOC platform. This is not only to promote the cooperation between peers, do more conducive to teaching goals.

2. Open is to be an opening to the outside world. MOOCs insists on the idea of opening, Implement the open registration, open learning, open learning methods, open teaching methods, open learning times, open learning evaluation. There are no access restrictions for all the Mu lesson platforms. MOOCs takes interest as guiding, advocates “making no social distinctions in teaching”. Anybody who wants to learn, regardless of gender, race, geography, can be registered to participate. This openness will break the wall of the University, so that people who really want to learn get quality-teaching resources easily, and promote education fair. According to a survey found by Stanford University, in most Mu courses the students are the people who already, graduate, work. They come to study admiring the purpose to update the knowledge, or to realize the wish to learn and not to learn. To teach those who really want to learn, the teacher will feel happy, will devote himself to the lecture, and thus enhance the teaching effect to the greatest degree. Because some students have related working experience, they can take place of assistant to answer questions proposed by other students. So, that will reduce the workload of teachers and assistants; improve the efficiency of the work of teachers.
3. Online is namely online. Mu class is online learning, whether teachers or students, do not have to visit the scene. You can then study any time from any place as long as you have access to a computer and the Internet. Without the limitation of space and time, professionals can always use their fragments of learning, which makes lifelong learning possible.

2.3 The Type of MOOCs

According to different teaching mode, the MOOCs can be divided into three types. XMOOC is a network of distance learning courses, based on behaviorism theory. The students learn by watching online video, they complete their task by online learning, online evaluation and other ways. CMOOC is the subject of the study, with the constructivist theory as the guide, social software as the platform. Students discuss one or several special subjects during a certain amount of time. They who participate in investigation and discussion share what one has learned to complete their construction of knowledge. TMOOC is a kind of learning style to complete task. The student compiles and composes work by various software tools, and submit works online. Among them, xMOOC is a dominant open course teaching mode.

The MOOCs education institution mainly consists of the consortium or university affiliated organizations of American University, including Coursera, Edx, Udacity, Khanacadem and Codeacademy et al. Coursera, Edx, Udacity are called troikas of MOOCs education. The network teaching platforms use the high quality teaching resources to create a variety of higher education courses, to create a new model of “fingertips”, to provide more opportunities for systematic study for all

kinds of students. Through these learning platform, students can not only get the learning resources that they are interested in, but also can help themselves to complete the study by designing personalized learning goals.

3 The Development of MOOCs in China

Those advantages of MOOCs, such as open of large-scale curriculum resources, high quality of the curriculum resources ensured by prestigious colleges and universities, individualization and independent selective of learning style, diversity of learning tools and learning resources provided by the network technology, make “MOOCs hot” spread the world quickly. According to incomplete statistics, at present, there are more than 13,000 MOOCs courses, including nearly 20 languages.

If 2012 was known as the first year of the world “MOOCs”, then 2013 is the first year of Chinese “MOOCs”. From 2013, the Universities of China also actively engaged in the wave of MOOCs. EDX announced the addition of 15 college online courses, including Tsinghua University and Peking University on May 21, 2013. On July 9th of the same year, the Fudan University, Shanghai Jiao Tong University signed a contract with Coursera. Within less than 2 months, frequent interaction of Chinese famous university and MOOC platform, indicates that the new mode of education of MOOCs has brought about tremendous impact on Chinese higher education since then, Chinese famous colleges and universities join MOOCs platform one after another.

Valuable is that the Chinese colleges and universities are actively creating a platform for local R&D and design of the MOOCs. Tsinghua University developed the online school, Shanghai Jiaotong University has developed a good university online, higher education press constructed excellent resource sharing platform for the course, Shanghai Jiaotong University and Peking University, Tsinghua University, Fudan University, Zhejiang University, Nanjing University, China University of science and technology, Harbin Industrial University, Xi'an Jiaotong University and Tongji University, Dalian University of technology, Chongqing University build Chinese “MOOCs” platform. In August and September 2013, led by the MOOCs center of East China Normal University, the C20 MOOCs of the basic education of Alliance (high school/middle school/primary school) had established. More than 20 national key high schools, more than 20 famous junior middle schools and more than 20 well-known primary schools joined the alliance. Anniversary of the establishment of a C20 MOOCs alliance, “Hua Shi MOOCs” network officially opened. All the web site will be open free to the national primary and middle school students. The president of Shanghai Jiao Tong University, academician of the Chinese Academy of Sciences Zhang Jie said “MOOCs is the largest education reform since the invention of printing, more importantly, it will reform university education and reshape the territory of Higher Education.”

At the same time, more and more domestic enterprises see the development prospects of MOOCs, but also actively respond to the promotion of Internet technology to the education industry. Netease Open Class, Guolairan open class, Sina, and shell network have launched their own open online courses. Youku reached exclusive official cooperation with Udacity to be the only Udacity platform to release courses in China. At present, Youku education channel has launched dozens of categories, nearly a thousand sets of latest Udacity online video courses translated into the Chinese.

In a period of time, the “MOOCs” becomes a common practice in our country, a hot word in the education sector. The words “flipped classroom”, “online and offline” are spread all kinds of educational forums and periodicals. In the background of the reform of the education system in Colleges and universities, joining the Mu class seems to have become the medicine to solve the ills of higher education, and become an imperative trend. Almost all the colleges and universities are concerned about the “MOOCs”, looking forward to join in this fashionable tide.

4 Introspection of the Heat of MOOCs

In the face of the craze that sweeps across the globe, we have to be calm down to “MOOCs” for dialectical thinking. MOOCs have positive significance. But if you think it as good medicine to cure all diseases, exaggerate its role, you are wrong. We need a comprehensive and objective understanding of the MOOCs. we should see its advantages, but also to see its limitations. We can neither turn a blind eye to the new education model, nor exaggerate its role. We require that the all-round, objective cognition admires class, now that needing to see it’s merit, will see it’s limitation again. This brand-new education pattern does not to wink at now that “admire class” unable face to face, can not exaggerate it’s effect excessively, comprehensive copy, the whole people all “admires class”.

4.1 *The Merit of MOOCs*

1. As an open online education model, Mu class makes quality education resources shared in a wider range. The openness of the MOOCs broke the University walls. It can not only promote high-quality education resources to exchange and share among colleges and universities, but also make the high-quality resources benefit more students not at school. For example, those who have already graduated participating in a job can register to learn through the MOOCs platform to renew their knowledge system. Those who failed to learn a course of their own interest in those days may acquire the opportunity to learn this course through the MOOCs platform. Even those who did not go to university, also can realize their college dream through this way.

2. MOOCs focus on the interactive cooperation online and offline, which will stimulate the reform of the traditional education to some extent. Lack of interaction is one of the drawbacks of traditional education. On tradition classroom, teachers as the main body, completely dominate the teaching process. Even though the teacher adopts the multimedia technology to play the teaching video, the students only accept the one-way information as the passive object and lack the feedback and interaction. But MOOCs forms such as flipped classroom, micro class, with the help of online and offline technology will compensate for this flaw. Online, outside the classroom, the students learn standardization courses through the video, master basic knowledge. Offline, in the classroom, the teacher guides the students to solve problem and share, discuss the related content. This not only realizes the effective interaction between the teachers and students, but also helps to promote the reform of the teaching mode, improve the teaching quality.
3. As a free online education model, MOOCs provide educational opportunities for more people who really want to study. All courses of MOOCs could register to learn for free, only need to pay only when you want to get the credit qualifications or certificate. Regardless of wealth, distinction, anyone wants to learn, as long as through a networked computer can enjoy national and even the world's most high-quality educational resources. This is of far-reaching significance to promote education equity.
4. The micro curriculum of the MOOCs platform, pays attention to the student's learning experience, and helps to improve the students' learning interest. According to the research, the attention of people on the Internet is not more than 15 min. The courses of the MOOCs are divided into 5–15 min micro curriculums, which is in accordance with the characteristics of human attention. There are a lot of questions to be answered in the course. The learning process is like each game to pass through. You can continue to listen only when you answer correctly. This advanced learning can't only make students master the learning progress; also firmly seize the attention of the students. At the same time, in the learning process once you encounter difficulties, you can directly put them on the platform. The teacher or other students will provide the relevant reference answer in 5 min to help you break through the pass successfully. Game style learns experiencing will meet the students studying interest, and improve learning quality.

In short, the positive significance of the MOOCs is that large-scale high-quality educational resources sharing promote fairness in education, innovation of the educational model and the teaching method, promotes the teaching reform of colleges and universities, thus contributes to the cultivation of innovative talents, promotes to significantly enhance the quality of higher education.

4.2 *The Limitations of Introspection*

1. MOOCs are online courses rather than online education. MOOCs is an online course, but not in the strict sense of online education. In general, online education has a very high rate of completion and excellent quality of Education. Institutions providing education are still the continuation of the combination of face-to-face teaching and online learning. These institutions not only teach students knowledge, but also through a variety of ways teach them how to do things. So, The educational ways of those institutions would expand the scope of education, improve the education effect. However, MOOCs only impart knowledge. It can not be considered a strict sense of education, can only be called online courses.
2. The advanced education technology of MOOCs is not equivalent to advanced education. The education progress would be driven by a series of factors such as spreading of advanced education ideas, the innovation of educational system, the promotion of the value of education, the progress of educational technology and so on. As one kind of factor, education technology can only change the education form to a certain extent. History has proved that, whether it is the radio, television or computer technology, their role of promoting education is far from the desired level. Advanced educational technology is the main content of MOOCs. If the MOOCs can not be combined with other factors, it will not be possible to promote the thorough change of education reality. Therefore, at the present stage, we should not expect too much of the MOOCs itself.
3. The outstanding quality of MOOCs curriculum is not equal to the outstanding quality of its training. MOOCs collect excellent courses national and even the world. Its courses are of high quality, but it causes a lot of criticism on training quality. Not all the courses are suitable for the form of MOOCs. For example, some liberal arts courses, completely using the form of MOOCs, can't achieve the ideal teaching effect. It is not all people are suitable for MOOCs study, only those who have learning motivation, enterprising, strong self-control, will achieve satisfactory results through MOOCs learning. Those without learning motivation, self management ability or learning ability, if they choose the learning model of MOOCs, no matter how high the quality of courses, they are likely to give up halfway. Some courses up to 80–95 % of the exit rate will be able to prove this point. Therefore, the excellent quality of the curriculum does not mean that the same quality of training.

5 Conclusion

As the product of the deep integration of Internet and education, the positive effect of MOOCs on higher education should not be ignored. It can effectively promote education fairness, improve education quality and effect, and renew education idea

quickly. So, we should use actively and develop the MOOCs, promote the reform of teaching, satisfy social demand much better. But, we want to see that the MOOCs are only an educational method and means, not the purpose of education. If we take all the courses in the form of the MOOCs, we won't get the effect we want. As a way to reconstruct the learning style, MOOCs is neither the only means of education, nor can it replace the traditional higher education.

Detection for Different Type Botnets Using Feature Subset Selection

Kuan-Cheng Lin, Wei-Chiang Li and Jason C. Hung

Abstract Information technology is developing rapidly today, which makes our life more convenient. Network is not only one of the important information technology products, but it also brings cybercrime, for example, Botnet infected and controlled computers which are usually established through a virus infection in many organizations, such as companies, schools or our home. Botnet do DDOS, phishing, sending spam and stealing of personal information. Every year the amount of infected victims is increasing. The botnet detection is more important day after day. However, Botnet often changed communication tools and transmission to hide, the detection has become difficult and botnet have multiple different implementations. We analyze three types of botnet traffic. There are IRC based, HTTP based and Peer to Peer based botnet. In this paper we construct simulation network to obtain different botnet traffic and extract flow data as some features. To find the important feature of botnet traffic, we use the Support Vector Machine as classifier with Swarm intelligence.

Keywords Botnet · Feature selection · Botnet detection

1 Introduction

The computer and networks are an indispensable technology now. No matter what the network technology brings us the comfortable life, it also bring with the disadvantage. Network technology is getting closely our life. We always save our information in the computer and internet. Is it safety?

K.-C. Lin (✉) · W.-C. Li

Department of Management Information Systems, National Chung Hsing University,
No. 250, Guoguang Rd, South Dist, Taichung City 402, Taiwan (R.O.C.)
e-mail: kclin@nchu.edu.tw

J.C. Hung

Department of Information Technology, Overseas Chinese University, 100, Chiao Kwang Rd, Taichung 40721, Taiwan (R.O.C.)

Botnet is a group. This group combines hacker and victims. The victims mean which computers infected a botnet virus and computers were remote controlled by controller. Botnet are mostly used for financial gain by controller. Botnet members attack websites or servers when the controller commands. These commands transport by standards based network protocols such as IRC, P2P or HTTP. Botnet members will infect more and more computers as his victims (bot) for gains, and it's why types of Botnet become more and more nowadays.

Botnet detection is a hard work. There are two categories of botnet detection, signature-based and behavior-based. The basis for the signature-based detection collected in the past with the current sample characteristics on ways to detect. Signature-based detection usually has a higher accuracy rate, but the disadvantage is needed to be update constantly and maintain the database of information attacks. Signature-based detections lack for ability to compare the unknown virus prevention [1]. Wang et al. [2] are using behavior-based detection to find new unknown botnet and malware, but there is a higher false positive rate than signature-based detection. Behavior-based detection includes feature detection and abnormal behavior detection. The difference between feature detection and abnormal behavior detection is that whether using machine learning to detect or not. Choi et al. [3] use feature-based method proposes a botGAD framework implement detection system that can instantly detect large-scale malicious network activity and abnormal behavior. Chen et al. [4] use abnormal behavior detection with neural network to construct a DDos defense system and contains k-means clustering algorithms. There is a high accuracy in this system.

We construct a virtual architecture to simulate botnet and we capture the network traffic to analysis with feature selection. Feature selection can enhance the accuracy of classifier, and it often used in intrusion detection system. In this paper we want to analysis the features which the computers just infected a virus. We have prepared three different viruses such as are ZEUS (HTTP) [5], NZM (IRC) [5] and Peacomm (P2P) [6]. We adopt Support Vector Machine as our classifier with a biological algorithm, MAFSA. MAFSA combine SVM provided a high accuracy for botnet traffic analysis [7].

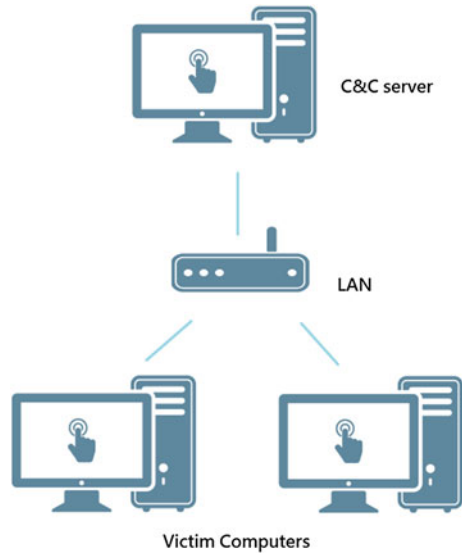
There are different protocols with different botnet types. We want to know the traffic changing when the computer get infect virus with different botnet viruses.

2 Method

2.1 Data Collection

We use a combination of three computers to network simulation. One of the computer play the role of control and command server, the others are victims which to infect botnet virus. The simulated environment shown as Fig. 1.

Fig. 1 Simulated environment



In order to obtain data, we use Wireshark to capture the traffic data. We let the victims getting infected once a botnet virus. And we collected victim computers' first hour traffic data. The purpose is to find that different of traffic feature in the botnet early infection. The next step is formatting all computers as new one. And set up the next botnet environment and repeat. So that we have three botnets' traffic each of 2 h data. On the other hand, we capture 1 h of normal computer traffic within common acting such like web browsing, telnet, sending and receiving E-mail, etc. We mixed the normal traffic and botnet traffic to simulate a real situation of botnet infection.

2.2 A Modified Artificial Fish Swarm Algorithm

The main idea of Artificial Fish Swarm Algorithm is simulate the behaviors of fish swarm and its swarm intelligence to solve optimization problems. There are three search steps in this algorithm. The three steps are swarm, follow and prey. Chen et al. [7] proposes a method of Modified Artificial Fish Swarm algorithms. MAFSA modified two changes dynamic vision and search best fish swarm to improve searching ability.

Table 1 Features for experiment

Feature no.	Feature name	Description
F1	Total_count	Total packets within the same connection
F2	Source_count	Different source IP counts
F3	Port_count	Number of using transmission port
F4	Low_port	The min port number
F5	High_port	The max port number
F6	Proto	Protocol (i.e. ICMP = 1, TCP = 6, UDP = 17)
F7	Total_volume	Size of total packets
F8	Min_packet	Size of smallest packet
F9	Max_packet	Size of largest packet
F10	Mean_packet	Mean size of packets
F11	Std_packet	Standard deviation from packets
F12	Time_regularity	Packet transmission time

2.3 Data Transformation

We must to transfer the raw data into train data for feature selection so we need to extraction the feature from the traffic which we capture. We set up 12 features for our experiment shows as Table 1.

These features made from traffic data. The train data transport from the raw data base on destination IP address. The F12 is defined by Liao et al. [1]. We define γ as a fixed time interval array that contains $n - 1$ counters, i.e., $\gamma = \{\gamma_2, \gamma_3, \dots, \gamma_n\}$, α as a frequently array, β as an infrequently array, and t as a constant value between 0 and 1, the default value is set to 0.5 (e.g., see Eq. 1).

$$\begin{aligned}
 \gamma_i > \frac{2t \sum \gamma_i}{n}, \quad \text{then} \quad \alpha_j &= \gamma_i \\
 \gamma_i \leq \frac{2t \sum \gamma_i}{n}, \quad \text{then} \quad \beta_k &= \gamma_i \\
 \text{Time_Regularity} &= \text{avg}(\alpha) * (\text{avg}(\alpha) - \text{avg}(\beta))
 \end{aligned} \tag{1}$$

3 Experiment Results

In order to classify the traffics which are malicious or normal traffics, we use a model combine Modified Artificial Fish Swarm Algorithm (MAFSA) and SVM. On the other hand, we used the five-fold cross-validation to make the experiment more reliable. The purpose of this experiment is to find a better subset to detect the botnet.

Tables 2, 3 and 4 shown as the results of experiment.

Table 2 Result of NZM (IRC)

Experiment							
Features	SVM*	SVM + MAFSA					
		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Selected count
F1—Total_count				V	V		2
F2—Source_count		V		V		V	3
F3—Port_count		V					1
F4—Low_port			V		V	V	3
F5—High_port			V	V	V	V	4
F6—Proto		V				V	2
F7—Total_volume		V	V	V	V	V	5
F8—Min_packet						V	1
F9—Max_packet		V			V	V	3
F10—Mean_packet							0
F11—Std_packet				V		V	2
F12—Time_regularity		V	V				2
No. of selected features	N/A	6	4	5	5	8	5.6
Average accuracy rate (%)	0.12	1	1	1	0.978	0.978	0.991

Bold value indicates result of simulation demonstrates the feasibility of the proposed approach
 *SVM (without feature selection)

Table 3 Result of ZEUS (HTTP)

Experiment							
Features	SVM*	SVM + MAFSA					
		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Selected count
F1—Total_count				V	V		2
F2—Source_count				V		V	2
F3—Port_count							0
F4—Low_port		V	V	V	V		4
F5—High_port				V			1
F6—Proto			V	V	V		3
F7—Total_volume		V		V		V	3
F8—Min_packet							0
F9—Max_packet			V		V	V	3
F10—Mean_packet		V			V	V	3
F11—Std_packet		V	V	V	V	V	5
F12—Time_regularity			V	V		V	3
No. of selected features	N/A	4	5	8	6	6	5.8
Average accuracy rate (%)	12.676	1	1	1	1	1	1

Bold value indicates result of simulation demonstrates the feasibility of the proposed approach
 *SVM (without feature selection)

Table 4 Result of Peacomm (P2P)

Experiment							
Features	SVM*	SVM + MAFSA					
		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Selected count
F1—Total_count		V	V		V		3
F2—Source_count			V	V			2
F3—Port_count							0
F4—Low_port			V		V	V	3
F5—High_port					V	V	2
F6—Proto			V	V			2
F7—Total_volume		V	V	V		V	4
F8—Min_packet			V	V			2
F9—Max_packet		V			V	V	3
F10—Mean_packet		V	V		V	V	4
F11—Std_packet		V	V	V	V		4
F12—Time_regularity		V		V		V	3
No. of selected features	N/A	6	8	6	6	6	6.4
Average accuracy rate (%)	13.216	1	1	1	1	0.978	0.996

Bold value indicates result of simulation demonstrates the feasibility of the proposed approach

*SVM (without feature selection)

It's important for botnet detection to find the critical feature in the early botnet infection. And mixing the same normal data flow through three different botnet traffic to classify with feature selection. Although the three different botnet traffic separately mixed with a normal data flow is used the five-fold cross-validation but the selected feature subsets by this validation are different. It means that each kind of botnet commutation doesn't have the same behavior. If we want to detect a new botnet, we must have a trained data. However, in these experiments, the features F4, F7 and F11 which are low_port, total_volume and std_packet are selected over 10 times totally. According to this result, when we have a trained data, we can find out the botnet usually has the fixed amount malicious traffic and standard deviation over a period of time and this characteristic is regardless of botnet commutation type.

4 Conclusion and Feature Works

All of experiments have high classification accuracy which means that the number of botnet networks to collect data over time lead to fewer items of raw data. The limitation of the research is that we have to figure out the botnet's behavior in early infection so that we captured just in the first hour. More victims and data collected

time may be advantage for the result of experiment. In addition, we have a good detection model, MAFSA.

A Modified Artificial Fish Swarm Algorithm (MAFSA) [7] to find the best solutions and combines Support Vector Machine (SVM) for feature selection. We reach the high accuracy by using SVM with MAFSA in this experiment. And we discover that there are the same features, such as fixed amount malicious traffic and traffic standard deviation, in the different botnet networks.

This experiment would be effective with more victim computers to improve the result, making it more practical. Another opinion is to hybrid the different botnet virus to infect a victim, and this condition happen in our environment.

Acknowledgments The work was supported by the Plan-103B1220 of Ministry of Science and Technology of Taiwan. We thank the Ministry of Science and Technology for funding this study.

References

1. Liao JL, Lin KC Feature selection for botnet detection using back-propagation network. National Chung Hsing University unpublished master
2. Wang K, Huang CY, Lin SJ, Lin YD (2011) A fuzzy pattern-based filtering algorithm for botnet detection. *Int J Comput Telecommun Networking NY, USA* 55:3275–3286
3. Choi H, Lee H (2012) Identifying botnets by capturing group activities in DNS traffic. *The Int J Comput Telecommun Networking NY, USA* 56(1):20–33
4. Chen J-H, Zhong M, Chen F-J, Zhang A-D (2012) DDos defense system with turing test and neural network. *IEEE international conference on granular computing, Handzhou, China*, pp 38–43
5. Malware Data Base Underc0de [Online]. Available: <http://malwares.underc0de.org/?dir=Botnets>
6. Open Malware Community Malicious Code Research and Analysis [Online]. Available: <http://www.offensivecomputing.net/>
7. Lin KC, Chen SY, Hung JC (2014). Botnet detection using support vector machines with artificial fish swarm algorithm. *J Appl Math*

Rotation Invariant Feature Extracting of Seal Images Based on PCNN

Naidi Liu, Yongfei Ye, Xinghua Sun, Junhua Liang and Peng Sun

Abstract In order to acquire a kind of stable and efficient feature sequences to identify different shape of seal images in different angles. Pulse Coupled Neural Networks (PCNN) are adopted to extract the energy logarithmic sequences of seal images, the input image is a binary image, different shape of seal images used as input data of PCNN network to acquire their energy logarithmic sequence as the standard sequence. Then the same flows are used to match the logarithmic sequences of images to be recognized with the standard sequences. In addition, angle rotated seal images also be recognized as identified images. Statistical results analyzed are based on Pearson correlation coefficient. The experimental results of different shapes stamp statistics show that using Pearson correlation coefficient and for statistical experiments compared the sequence that obtained more desirable results. Through many seals experiments proved that the result of Pearson correlation coefficient can reach more than 0.99. The energy logarithmic sequence of different shape of seal images can be used as the feature sequences, which is not impact by the seal's chop angles, and the feature has a certain stability.

Keywords Computer image processing · Seal image · Rotation invariant · Pulse coupled neural network · Energy logarithmic sequence · Pearson correlation

(1) Major Research Projects of Hebei North University (ZD201303). (2) Hebei Province Population Health Information Engineering Technology Research Center.

N. Liu (✉) · Y. Ye · X. Sun · J. Liang · P. Sun
School of Information Science and Engineering, Hebei North University,
Zhangjiakou, Hebei Province, China
e-mail: lnd1987@126.com

Y. Ye
e-mail: yeyongfei005@126.com

1 Introduction

With the development of science and technology, the circulation of seals become more extensive and convenient, at the same time, the counterfeit seals become simulate, In order to maintain the stability and normal of our social, avoid the unnecessary economic losses, recognition of the seal becomes particularly important.

Since the early 80s, Many experts and scholars focus on the experimental study of seal identification problems, Different image processing technology in the feature extraction of seal images also obtain well apply [1], In view of the different during the feature extraction research objects, The seal identification methods can be roughly classified into three categories: First, regard the partly one or more feature points as research object, match the feature points after coherent processing. Such as: Ho et al. [2] who raised the view that based on Edge Seal automatic identification, which use SIFT feature matching point to match the test seal with the reservation seals. Zou [3] who raised the view which based on the using of greedy algorithm segmentation. The arithmetic triangle-mesh the minutiae through triangle mesh. Hongwei Dai al [4] proposed a study based on the rotation invariant Seal Identification method, this viewpoint select the invariant feature point firstly to get stable characteristic vector. The second category is regard the whole seal as the research object of feature extraction and use the whole feature into recognition. such as: Hongyan et al. [5] proposed a credential authentication system based on embedded system architecture, regard texture spectrum features as characteristic to identify the seals. Such as: Ueda and Matsuo [6] proposed a use of stamp topical, global feature matching method of identification.

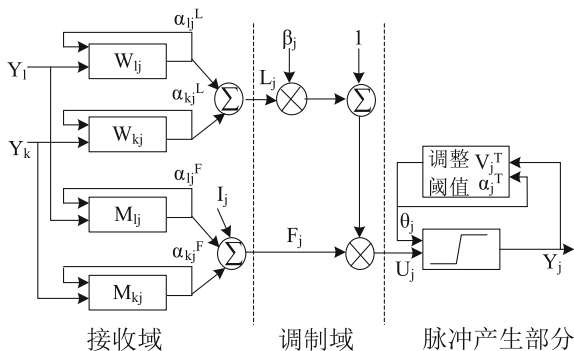
In this experiment, the whole seal are deemed as the object of feature extraction study subjects, in order to achieve the classification of different shapes stamp shapes, and easier to seal further identification. Shape feature extraction is seemed as the central of our experiment, shape description is use the special sequences to describe the spatial information of shapes. Shape description is the foundation of shape recognition, shape matching, shape searching [7, 8], experiments using pulse coupled neural networks (PCNN) to extract the seals' shape feature sequence.

2 PCNN and Feature Extraction

1989, Eckhorn studied the issued by the sync pulse phenomenon of cat cerebral cortex neurons, then he raised pulse coupled neural network model [9]. It is mostly similar between the PCNN and the biological neural model. So it is used in the image object recognition, the image fusion, the image texture processing, the image feature extraction and other fields. The model of single neuron is shown in Fig. 1 [10].

The model is a simulation of real neurons, a number of objective and subjective factors which affect the operation of real neurons, such as the impact of age, ambient temperature, were not taken into account. Model consists of receiving domain, modulation domain, pulse generating part add up to three parts.

Fig. 1 Basic PCNN neuron



$$L_j[n] = \exp(-\alpha_{kj}^L \cdot t) \cdot L_j[n - 1] + V_L \cdot \sum_k W_{kj} Y_k[n - 1] \tag{1}$$

$$F_j[n] = \exp(-\alpha_{kj}^F \cdot t) \cdot F_j[n - 1] + I_j + V_F \cdot \sum_k M_{kj} Y_k[n - 1] \tag{2}$$

$$U_j = F_j(1 + \beta_j L_j) \tag{3}$$

$$\frac{d\theta}{dt} = -\alpha_j^T \theta_j + V_j^T Y_j(t) \tag{4}$$

$$Y_j = \text{Step}(U_j - \theta_j) \tag{5}$$

After receiving domain receives the input signals from other neurons or externalities, the signal are transited by the via channel F or L-channel transmission, the signal which transited by F via as the feedback path to input modulated manner via and the signal transmission by path L connected to the coupling modulated.

The working mechanism of modulating portion can be reflected by the formulas (1), (2). W_{kj} , M_{kj} are two right of synaptic connection, these two parameters determine the adjacent neurons to convey the strength of the information center neuron ability to impact on central neurons. α_k^L and α_k^F are two decay time constant, V_L , V_F , respectively express the feedback amplification coefficient and connection amplification coefficient, I_j said input constants. Equation (3) is neurons internal activity expression. The L_j signal transited by L-channel and then plus a positive offset, multiply modulation the F_j which transited by F-channel. The offset of this model is seemed as 1. β_j is coefficient of coupling. And have the function to adjust the surrounding neurons. It also have the function of influence of central neurons firing cycle. β_j is a constant between [0, 1] in normal circumstances. Since the change in the signal F_j slower than the signal L_j , obtained by multiplying the modulation signal U_j is approximately a rapidly changing signal is superimposed on an approximately constant signal.

Pulse generating section constituted by comparator and the leakage integrator, dynamic threshold is determined by amplification coefficient and time constant. Neurons internal action items determine whether ignition by compared with the dynamic threshold. Equation (4), V_j^T and the amplitude coefficient α_j^T denote the threshold value and the time constant, Eq. (5) is output items.

When PCNN used in the feature extraction, translation, rotation, scaling, distortion, etc. invariance [11]. In this experiment, the number of sequences of PCNN energy instead of the usual sequence as a stamp image entropy feature extraction.

The experiment use the different shapes of seals, and gived their energy to the number of sequence features, PCNN parameters chosen to $\alpha_k^L = 1$, $\alpha_{kj}^F = 0.1$, $\alpha_j^T = 1$, $V_L = 0.2$, $VF = 0.5$, $V_j^T = 27$, $\beta_j = 0.1$, 40 iterations [12, 13].

The selected seals are circulating widely and under the seal of the relevant provisions. Figure 2 shows the seal image used in the experiments, using a low-pass digital filter to stamp collection image denoising filter [14], in the simple background case, the red seal experimental extraction through RGB model [15] on seal image binarization With Top-hat morphological transformation [16–18]. The image processing, that is, after the expansion of the image to make the first opening operation of corrosion, and then open the original image and the resulting operation Images obtained by subtracting the ideal seal binary image.

The 6 pictures in Fig. 2 are output energy logarithmic sequence after image preprocessing. Energy logarithmic are shown in Fig. 3a–f.

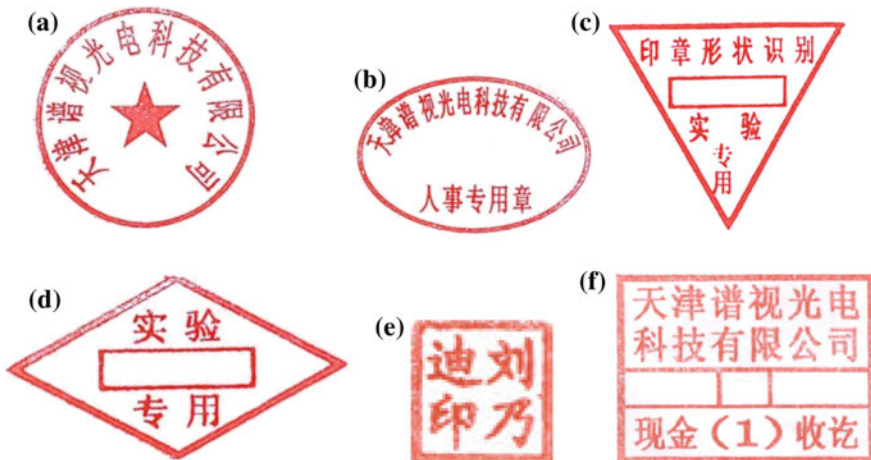


Fig. 2 Scanned seal images. **a** Round sample seal, **b** Oval samples stamp, **c** Triangle sample seal, **d** Diamond sample seal, **e** Square sample seal, **f** Rectangular sample seal

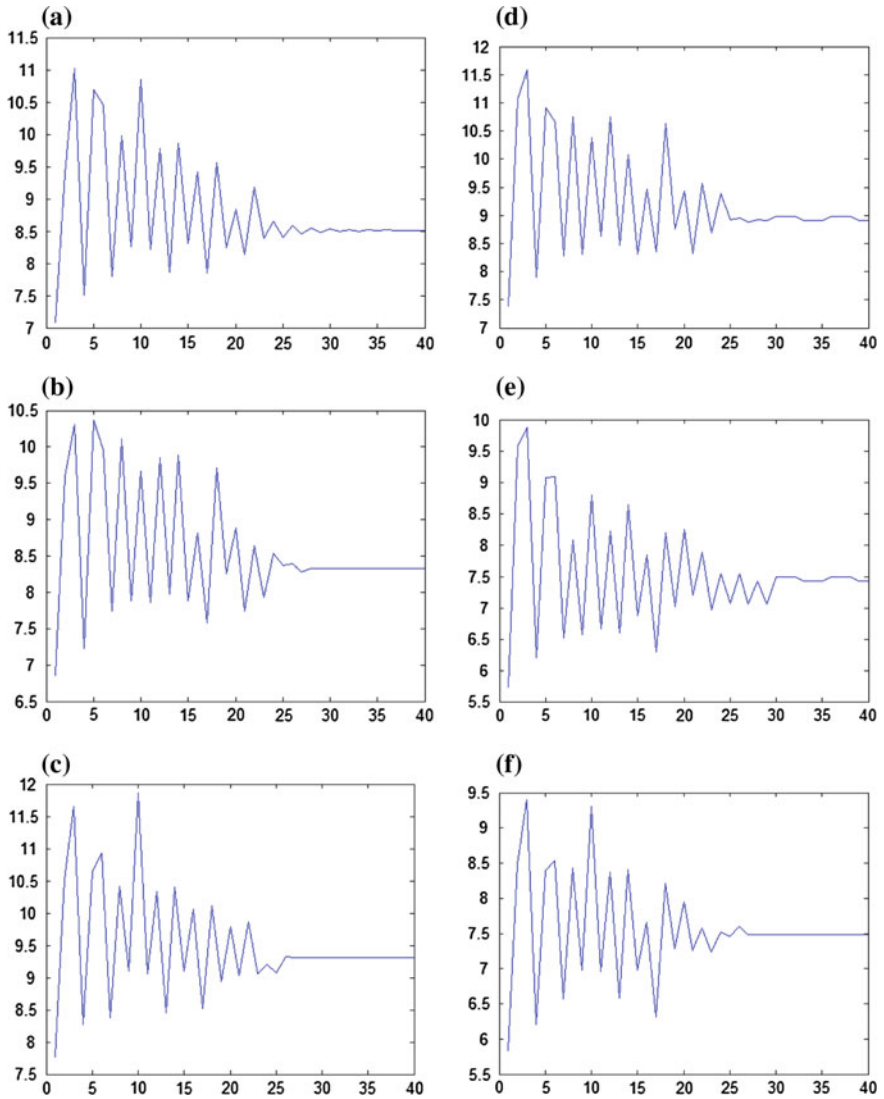


Fig. 3 Energy logarithmic sequences of different seals which processed by PCNN. **a** Energy to the circular seal sample number sequence, **b** Energy to the oval seal sample number sequence, **c** Along with the energy of a triangular sample number sequence, **d** Along with the energy of a diamond sample number sequence, **e** The seal of the energy of a square sample number sequence, **f** Along with the energy of a rectangular sample number sequence

3 Pearson Correlation Coefficient

In order to observe the relationship between the sequence of each group, we need some kinds of methods to analysis the difference and relation between different sequence. The observed values of the variables are pairs, for experiment 40 times the ignition, can be considered as 40 pairs of pairs of sequences, and between each sequence are independent, so you can consider using the Pearson (Pearson) correlation coefficient [19] to compare the image trend, whereby each sequence of visual analysis.

Suppose two variables X, Y, then the Pearson correlation coefficient between two variables can be calculated by the following Eq. (6).

$$P_{x,y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \tag{6}$$

In the above formula, N represents the number of variable values, based on the value of each sequence can be obtained N Pearson correlation coefficient between the two sequences, in this experiment, N = 40, Fig. 2 Energy for each seal rooms Pearson correlation coefficient P is shown in Table 1.

The analysis table can be found for the Pearson correlation coefficients between the sequences of different shapes PCNN seal in (0.9, 1), Pearson correlation coefficients in Table 1 is a diagonal position.

The results show that there is some correlation between sequence, but the combination of the variance can easily find where its distinctive shape. For shape recognition seal, the seal can be more energy as the sample sequence for number sequences stored energy will be detected on the seal number and the sample sequence matching sequence. Figure 4 shows the two shapes to be recognized stamp images.

PCNN energy logarithmic image of Fig. 4 are corresponding to the number of images in Fig. 5a, b.

Table 1 Pearson correlation coefficients between the different shape of seals

P	Figure 2a	Figure 2b	Figure 2c	Figure 2d	Figure 2e	Figure 2f
Figure (a)	1.000	0.957	0.977	0.939	0.917	0.943
Figure (b)	0.957	1.000	0.921	0.976	0.924	0.931
Figure (c)	0.977	0.921	1.000	0.920	0.932	0.970
Figure (d)	0.939	0.976	0.920	1.000	0.941	0.924
Figure (e)	0.917	0.924	0.932	0.941	1.000	0.944
Figure (f)	0.943	0.931	0.970	0.924	0.944	1.000

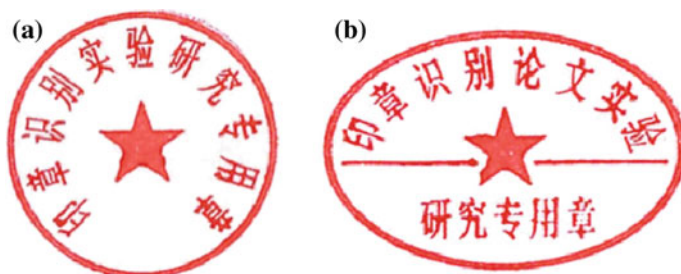


Fig. 4 Seals to be identified. **a** Be recognized seal 1, **b** To be recognized seal 2

Fig. 5 Seal's energy logarithmic sequences of seals in Fig. 4. **a** Be recognized seal of energy logarithmic sequence 1, **b** To be recognized seal of energy 2 logarithmic sequence

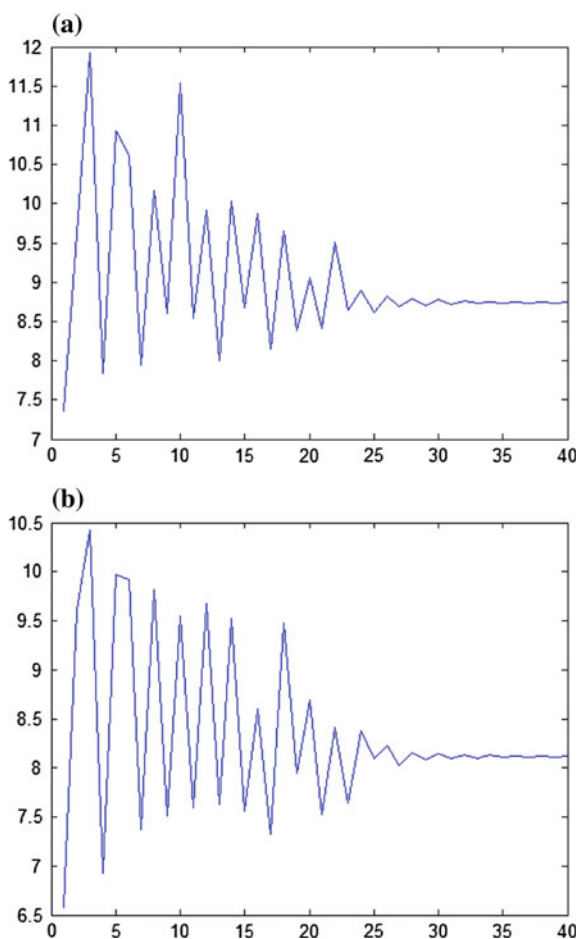


Table 2 Pearson correlation coefficients between the sample sequences and the energy logarithmic sequences of seals in Fig. 4

P	Figure 2a	Figure 2b	Figure 2c	Figure 2d	Figure 2e	Figure 2f
Figure 4a	0.991	0.918	0.981	0.910	0.900	0.930
Figure 4b	0.960	0.995	0.941	0.987	0.948	0.948

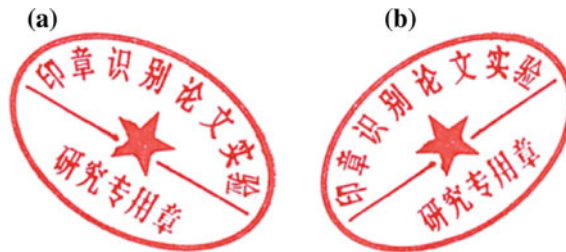


Fig. 6 Seals with different stamped angles to be identified. **a** Be recognized seal 3, **b** To be recognized seal 4

Identified seals's PCNN sequence Pearson correlation coefficients of each sequence of samples and is shown in Table 2.

Along many different shapes experiments show that if the seal in the same shape, its Pearson coefficient is 0.99 or more, the conclusion can be obtained from the table better reflected.

To further explore PCNN respect to the seal of the angle is not the same recognition effect, this will be recognized in different angles stamp seal paper pick 6 two around the seal image as an example to elaborate.

Figure 6's PCNN energy logarithmic sequences are shown in Fig. 7.

Energy test images at different angles stamped seal on the Pearson correlation coefficient of the sample sequence number sequence shown in Table 3.

From Table 3, we can see that, after the energy stamp image rotated through the PCNN treating the resulting sequence of samples of several sequences match can still accurately determine the shape of the seal by energy logarithmic sequence. And the characteristic shape of the seal extraction has good general applicability.

Fig. 7 Energy logarithmic sequences of each identified seals in Fig. 6. **a** Be recognized seal of energy logarithmic sequence 3, **b** To be recognized seal of energy logarithmic sequence 4

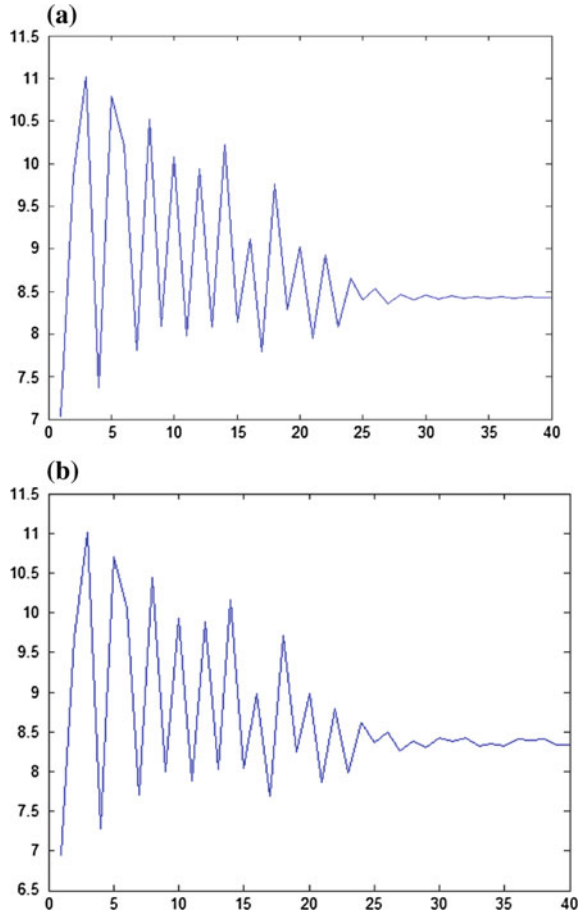


Table 3 Pearson correlation coefficients between the sample sequences and the energy logarithmic sequences of seals in Fig. 6

P	Figure 2a	Figure 2b	Figure 2c	Figure 2d	Figure 2e	Figure 2f
Figure 6a	0.972	0.993	0.941	0.973	0.928	0.928
Figure 6b	0.968	0.993	0.936	0.972	0.926	0.928

4 Conclusion

The seal shape recognition is based on PCNN, which based on pixels corresponding to neurons. And make a full use of the rotation invariant, scale-invariant features. During the experiment, the energy logarithmic of different shapes of common seal pretreated after PCNN, obtained as a characteristic sequence, and the unidentified

seal' PCNN energy logarithmic are matched with the characteristic sequence. If they are the same shape, the Pearson correlation coefficient above 0.99, with better shape discrimination. In addition, the energy of the number sequence is a seal shape recognition feature that is not to be stamped seal of the rotation angle recognize the impact has better applicability, to provide a protection for the follow-up work.

Acknowledgment Major Research Projects of Hebei North University (ZD201303); 2.

References

1. Liu H, Lu Y, Wu Q et al (2007) Automatic seal image retrieval method by using shape features of Chinese characters. In: Proceedings of IEEE international conference on systems, man and cybernetics, 2007, pp 2871–2876
2. Ho J, Tie R, Zhou Y, Zhang H (2010) Based automatic identification credential edge difference. *Instrument* 31(1):85–91
3. Zou, Jin Y (2007) Triangulation analysis based on greedy algorithm seal matching. *Comput Eng Des* 28(5):1199–1201
4. Wei X, Xue Y (2012) Research on automatic identification method for circular seal. *Radio Eng* 42(10):51–54
5. Sun H (2010) Tian tree learning, Zhang Xuedong credential recognition based on spectral measure HSI space and texture. *Nat Sci* 28(2):221–224 (Shenyang Normal University)
6. Ueda K, Matsuo K (2005) Automatic seal imprint verification system for bank-check processing. In: Proceedings of third international conference on information technology and applications, 2005, pp 768–771
7. Cai H (2013) Target shape feature extraction. *Comput Mod* (4):107–124
8. Wang B (2012) Based on a constant area shape descriptors Fourier transform. *Acta Electron* 1:84–88
9. Huifei, Zhao X (2010) Mold target feature extraction based on pulse coupled neural network. *J Jilin Univ* 28(5):474–478
10. Gu X, Yu D (2001) PCNN principles and applications. *Circ Syst* 3(6):45–50
11. Johnson JL, Padgett ML (1999) PCNN models and applications. *IEEE Trans Neural Netw* 10 (3):480–498
12. Deng X, Ma Y (2012) Improve PCNN adaptive parameter setting and model. *Acta Electron* (5):955–964
13. ANS adaptive setting and model improvement (2013) PCNN parameters based on visual information. *Comput Sci* 40 (6):291–294
14. Lixia M, Qiuping Z (2004) Application of image processing technology in the seal of recognition pretreatment. *Wuhan Univ Inf Sci* 29(8):691–693
15. Ye P, Ting L (2012) Seal based on RGB color image pre-processing features. *Autom Inf Eng* 5:18–21
16. Tao Z, Li Z (2004) Pretreatment of image, Wang Jian, Chen Yun, Wang Lin seal recognition. *J Sci Instrum* 25(4):401–403
17. He J, Liu T, Zhang Z (2008) An adaptive morphological algorithm to segment Chinese square seal in bank check image. In: Proceedings of SPIE, 2008 (7156): 71560Y1-12
18. Li B, Liu X, Guo X, et al. Top-hat morphological filtering in image pre-processing application and FPGA implementation. *Photoelectr Control* 18(10):76–81
19. Winter R (2012) Conflict over new evidence Pearson coefficient method based synthesis. *Telecommun Eng* 52(4):466–471

The Taguchi System-Two Steps Optimal Algorithm Based Neural Network for Dynamic Sensor Product Design

**Ching-Lien Huang, Yung-Hui Chen, Chun-Hsiung Tseng,
Tian-Long John Wan, Lung-Cheng Wang and Chang-Lin Yang**

Abstract The key successful factor of the new product design (NPD) of sensor product industry is the selections of the best parameter level. General speaking, decreasing the error rate of product parameter selection by increasing the number of experiments is the commodity trend in NPD goals. For above reasons, previous studies focus on structured approach for the replacement and management of selection of the parameter level in product design with the purpose of increasing efficiency and effectiveness, but rarely on a dynamic environment. Consequently,

C.-L. Huang

Department of Industrial Management, Lunghwa University of Science and Technology, 300, Sec. 1, Wanshou Rd, Guishan, Taoyuan 33306, Taiwan
e-mail: lynne.line@msa.hinet.net

Y.-H. Chen

Department of Computer Information and Network Engineering,
LungHwa University of Science and Technology, Taoyuan, Taiwan
e-mail: cyh@mail.lhu.edu.tw

C.-H. Tseng (✉)

Department of Information Management, Nanhua University,
Chia-Yi 62249, Taiwan
e-mail: lendle_tseng@seed.net.tw

T.-L.J. Wan

Department of Information Management, Lunghwa University of Science and Technology, Taoyuan, Taiwan
e-mail: johnwan.mr@gmail.com

L.-C. Wang

Department of Component BG, Lunghwa University of Science and Technology,
Taoyuan, Taiwan
e-mail: steven.lc.wang@hotmail.com

C.-L. Yang

Department of Business Administration, Fu Jen Catholic University,
New Taipei, Taiwan
e-mail: 051125@mail.fju.edu.tw

this work presents a novel algorithm, the Taguchi System-two steps optimal algorithm, which combines the Taguchi System (TS) with two steps optimal (TSO) method, which is shown how product adjusted under a dynamic environment in product design. The utility of the parameter level are selected. The two step optimal (TSO) method links the decisions for selections of parameter level in two different times and can be used to focus on dynamic sensor product design system (DSPDS). From the results, the proposed method might possibly be useful for our problem by selecting the parameter level size and adjusting the parameters by TSO and neural network (NN) in the DSPDS is observed in this study.

Keywords Taguchi system (TS) · Dynamic product design system (DPDS) · Dynamic sensor product design system (DSPDS) · Two steps optimal algorithm (TSO) · Neural network (NN) · New product design (NPD)

1 Introduction

Currently, the selections of the best parameter level in NPD are the key successful factors for the sensors manufacturing industry. Generally speaking, decreasing the error rates of product parameter selections by increasing the number of experiments is the NPD trend. Therefore, the selection of best parameter level sometimes leads to more cost increasing and jobs reworking. From previous papers, the TS method has been successfully combination of various kinds of other's method and tools by adjusting the parameters and parameter levels [1–12].

Moreover, the TSO has been successfully adopted in dynamic environments and the NN has also been used in dynamic environment in the past works [13, 14].

Consequently, this work presents a novel algorithm, the Taguchi System-Two Steps Optimal algorithm, which combines the Taguchi System (TS) with the two steps optimal (TSO) method, which is shown how product adjusted under the dynamic environment in product design. The benefits of parameter level are selected; then, the TSO links the decisions for the selections of parameter level in two different times and can be used to focus on DSPDS. The remainder of this paper is organized as follows. Section 2 describes the TS algorithmic process for selecting of parameter level and presents the TSO algorithm, which changing parameters adjusting system in a dynamic environment. Section 3 illustrates the algorithm's effectiveness and shows the analysis. Conclusions are finally drawn in Sect. 4, along with recommendations for future research.

2 The Taguchi System-Two Steps Optimal Algorithm

The process of TS-TSO algorithm will be generated in this section. Moreover, the Verify and Validate are also applied to estimate the efficient of the proposed methodology which will be discussed in the later part of this section. And, the proposed algorithm starts with the following. First, to establish the TS model, one set of data is chosen from a system. The parameters of the original data are selected to calculate the signal-to-noise (SN) ratios. The most important task is to determine which parameter levels are selected. Second step, the TSO algorithm is applied. The purpose of TSO algorithm is to establish a DSPDS, which is completed using the next two detail steps. First, whether β is significant is determined. Next, the formula $Y_i = \beta M_i$ is applied to a DSPDS for adjusting product parameters to obtain a new β value which is expected to be 1.

In sum, the algorithmic procedure has the following three steps.

Step 1. Establish the Taguchi system

In the beginning, the levels of the important parameter of the product are chosen based on the SN ratios, which is shown in Eq. (1)

$$\eta = 10 \log_{10} \frac{\bar{Y}^2}{S^2} \quad (1)$$

The process is as follows.

1. **Assess the parameters of product:**
Estimate the parameters.
2. **Decide parameter level number of product:**
Set and select the parameter level number of product from the data set as control parameters.
3. **Compute the SN ratios of product:**
Compute the SN ratios of parameters for product.
4. **Verify the new system:**
Use the CI to verify the new system.

Step 2. Establish the dynamic sensors product design system

This section has two steps. First, the SN ratios are maximized and selected from the data set. Secondary, if the SN ratio is insignificant, then adjusts the value of β . The process procedure is summarized as follows.

5. **Construct the dynamic sensor product design system:**
The TSO and NN algorithm is utilized to construct and verify the DSPDS. Figure 1 presents the TS-TSO algorithm.

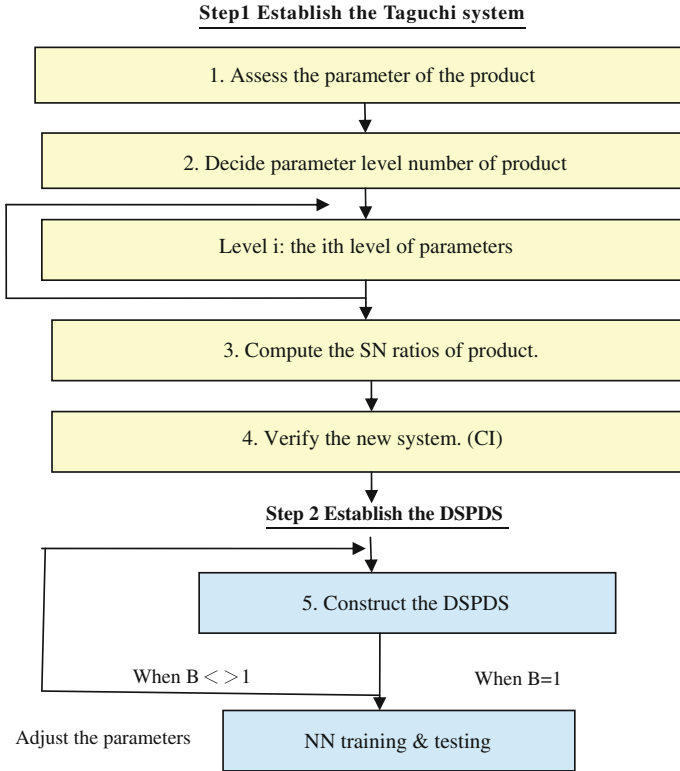


Fig. 1 The TSO-NN algorithm

3 Verification

3.1 Establish the Taguchi System

The proposed case is a company that produces speed sensor of winding machine. It is the important part in many kinds of machine. Therefore, the proposed algorithm is applied in NPD. For establishing the TS model, the parameters, the level numbers of parameter, and the observation values are collected and coded from v_1 to v_4 , L_1 to L_3 and y_1 to y_3 .

In total, 27 data are selected from the data set. The TS-TSO algorithm is applied as follows.

At first, the original data and the SN ratios of these parameters are calculated (Table 1).

Consequently, the levels of the parameters are selected form the data set. The expected value of the SN ratios is 1461.89, which indicates an optimal state (the target value is 1500).

Table 1 The values of the original data (parameters)

v ₁	v ₂	v ₃	v ₄	y ₁	y ₂	y ₃
45	0.05	7350	15	1773	1880	1793
45	0.055	7500	20	1603	1598	1594
45	0.06	7650	25	1473	1481	1485
50	0.05	7500	25	1791	1815	1808
50	0.055	7650	15	1647	1650	1647
50	0.06	7350	20	1386	1383	1387
60	0.05	7650	20	1857	1865	1852
60	0.055	7350	25	1561	1540	1565
60	0.06	7500	15	1441	1447	1444

Next, the confidence interval (CI), which confirms product design system reliability—is derived using Eq. (2).

$$CI_1 = \sqrt{F_{\alpha,1,df2} \times Ve \times \left(\frac{1}{n_{eff}}\right)} \tag{2}$$

where $F_{\alpha,1,df}$ is the α value of F, α is an obvious level, $1 - \alpha$ is a CI, df is the degree of freedom for an item, v_e is the combination error item of variance, and n_{eff} is the number of experiments, which is shown as follows.

The CI of sample experiments is ± 5.7 . For testing, the second data set is extracted and the experiment is performed at 83 %. CI must be within this range, the structure of the new product design (NPD) system is valid.

3.2 Construct the Dynamic Sensor Product Design System

The process of TSO and NN will be generated and discussed in this section. Moreover, the Verify and Validate are also applied to estimate the efficient and effective of the proposed methodology will be discussed in the later part of this section.

3.2.1 Establish the Dynamic Sensor Product Design System

The TSO algorithm is applied to determine whether the DSPDS is good. The processes are as follows.

Step 1. Maximize the SN ratios

The SN ratios of the levels of parameters are selected from the data set. Tables 2, 3, 4, 5, 6 and 7 present the levels values of parameters and the experiment results.

Table 2 The SN ratios of experiment results

Level	v ₁	v ₂	v ₃	v ₄
1	42.92	40.81	42.61	47.77
2	53.11	50.61	49.31	52.17
3	47.99	52.6	52.1	44.09
Delta	10.19	11.79	9.49	8.08
Rank	2	1	3	4

Table 3 Means values of Y for experiment results

Level	v ₁	v ₂	v ₃	v ₄
1	1631	1826	1585	1636
2	1613	1601	1616	1614
3	1619	1436	1662	1613
Delta	18	390	77	23
Rank	4	1	2	3

Table 4 The comparison of the variable, SN ratios, and means

v ₁	v ₂	v ₃	v ₄	SN	Mean
<i>The standard values of variables, the SN, and mean</i>					
50	0.055	7500	20	61	1579
<i>The actual values of the SN ratios at the beginning</i>					
60	0.06	7500	15	54	1444
<i>The actual values of variables, the SN, and mean at the second time</i>					
50	0.06	7650	20	65.956	1461.89

Table 5 Expected values of Y

Experiment No.	1	2	3	4	5	6	7	8	9	10
Expected values of Y	1492	1503	1494	1497	1502	1508	1502	1499	1492	1500

Table 6 β values at the second time

β	v ₁	v ₂	v ₃	v ₄
Level 1	1.087	1.217	1.057	1.091
Level 2	1.075	1.067	1.077	1.076
Level 3	1.079	0.957	1.108	1.075
Delta	0.012	0.260	0.051	0.015

Table 7 The combination of the parameters

The combination of parameters	Mean	Std
Original combine experience (Predicted) A2 B2 C2 D2	1579	-12.6
Primary combine experience (True) A3 B3 C2 D1	1444	3
Predicted combine experience (True) A2 B3 C3 D2	1498	5.7

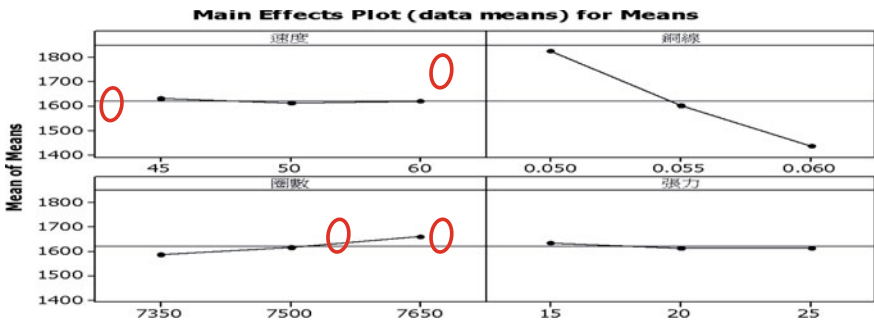
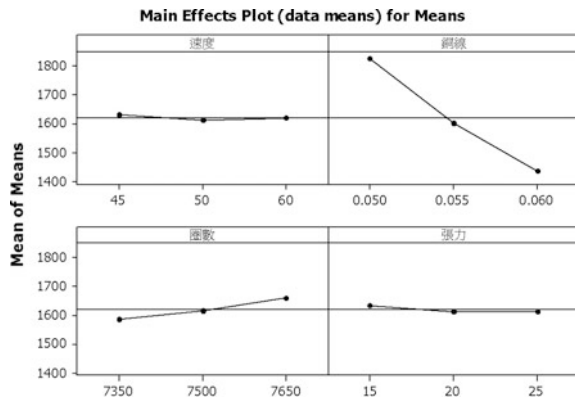


Fig. 2 The SN ratios of the experiment results

Fig. 3 The means values of the experiment results



And Figs. 2, 3, 4 and 5 shows the comparison of the parameters and the experiment results.

To maximize the SN ratios of these data, levels of the parameters are chosen based on the largest SN ratio, which are shown in Tables 3 and 4.

The SN ratios and mean values of parameters are computed and shown in Figs. 3 and 4.

The expected values of Y are computed and shown in the following.

When the value of β is not equals to 1, the value has to be adjusted in the DSPDS. The processes are shown in the next step.

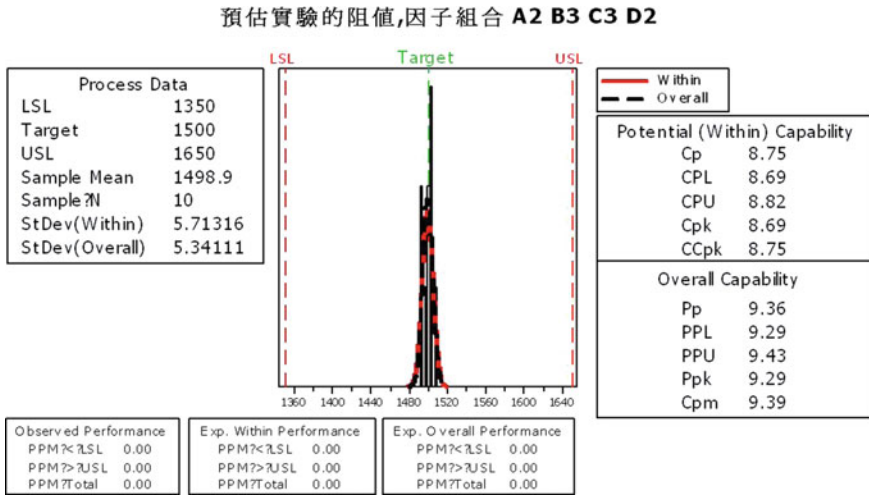


Fig. 4 The target value of experiment results

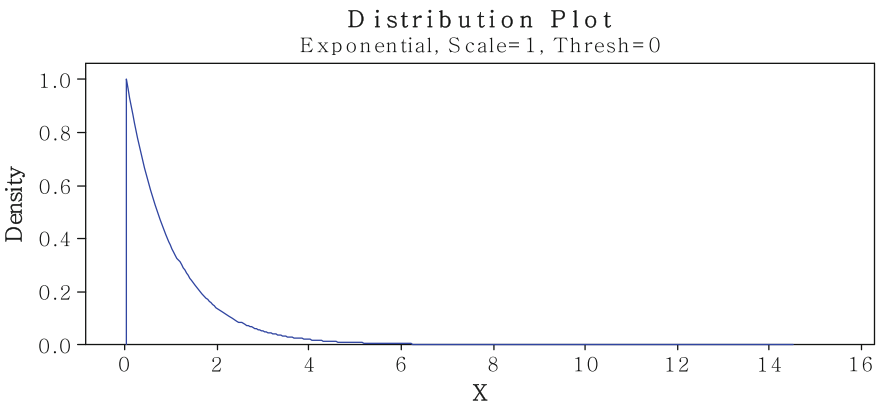


Fig. 5 The RMSE values of testing for 4-4-4 structure

Step 2. Adjust the β values

For the second time analysis, β is computed using the following formula for a DSPDS. Therefore, the estimated value of the SN ratios, $\hat{\eta}$, is 65.956.

$$\hat{\eta} = \sum_{i=1}^4 v_i - n\bar{\eta} \tag{3}$$

From Eq. (3), the value of $\hat{\eta}$ is obtained as $\hat{\eta} = 65.956$.

The β value is 1.053, which can be adjusted to 1 by Eq. (4). is about equal to 1.

$$\hat{\beta} = \sum_{i=1}^4 \beta_i - n\bar{\beta} \quad (4)$$

From Eq. (4), the combination value of the β equals $\hat{\beta} = 0.999$.

This step adjusts the value of β , which is almost close to 1.

Then, parameters are adjusted. After that, the SN ratios and the means were adjusted at 65.956 and 1461.89 (Table 4).

3.2.2 Verify the Methodology

This section, the NN algorithm is applied to model the DSPDS, and determine whether it is good. The process is shown in following.

Modeling

The relationship model between parameters and responses is developed using a BP NN, in which 18 inspection data are used for training and 9 lots are used for testing. The model structure is selected using 4-4-4 (input-hidden-output) (Table 7).

Testing the NN Model

As the average standard deviation estimate (RMSE) is the convergence criterion employed in network training and testing. Then, to determine the options of the NN structure, the architecture 4-4-4 is chosen to get the convergent performance. Restated, the RMSE of training error is 0.01, and the testing error is 0.01. And, these two RMSE values for training and testing are convergent (Fig. 5).

According to data for the confirmation set, the formula for desirable functions is $Y_{ij} = F(v_1, v_2, v_3, v_4)$, which is applied for a DSPDS can map in 4-4-4 (input-hidden-output) NN structure successfully.

4 Conclusion

The discussing of the proposed algorithm is confirmed by changing the value of parameters and the number level of experiment. Therefore, the above analysis could show that the condition toward in the target value. Restated, the optimal pattern chosen has the following variables levels. They are A_2 , B_3 , C_3 , and D_2 . (Fig. 2). And, the mean value of Y has adjusted and is shown in Fig. 3. Thus, the combination mean value is adjusted at 1498, std is 5.7, therefore the combination is small, and closely in target value in 1500 Ω resistance value (Table 7). And, the original values of the parameters are adjusted at 50, 0.06, 7650, and 20 (Table 4).

Besides this, in modeling a DSPDS aspect, the NN algorithm shows us the 4-4-4 structure is the optimal selection from other's architecture, which shows RMSE is convergence in 0.01.

It is concluded that the propose algorithm can be applied successfully to dynamic environments for solving the PD problems.

References

1. Ozelcik B (2011) Optimization of injection parameters for mechanical properties of specimens with weld line of polypropylene using Taguchi method. *Int Commun Heat Mass Transfer* 38:1067–1072
2. Lin CH, Shih SJ, Lu AT, Hung SS, Chiu CC (2012) The quality improvement of PI coating process of TFT-LCD panels with Taguchi methods. *Optik* 123:703–710
3. Liao CN, Kao HP (2010) Supplier selection model using Taguchi loss function, analytical hierarchy process and multi-choice goal programming. *Comput Ind Eng* 58:571–577
4. Su CT, Yeh CJ (2011) Optimization of the Cu wire bonding process for IC assembly using Taguchi methods. *Microelectron Reliab* 51:53–59
5. Yiamsawas D, Boonpavanitchakul K, Kangwansupamonkon W (2011) Optimization of experimental parameters based on the Taguchi robust design for the formation of zinc oxide nanocrystals by solvothermal method. *Mater Res Bull* 46:639–642
6. Cheah ELC, Heng PWS, Chan LW (2010) Optimization of supercritical fluid extraction and pressurized liquid extraction of active principles from *Magnolia officinalis* using the Taguchi design. *Sep Purif Technol* 71:293–301
7. Boothroyd G, Dewhurst P, Knight WA (2002) *Product design for manufacture and assembly* Marcel Dekker, New York
8. Solehati N, Bae J, Sasmito AP (2012) Optimization of operating parameters for liquid-cooled PEM fuel cell stacks using Taguchi method. *J Ind Eng* 18:1039–1050
9. Rouzbeigi R, Edrissi M (2011) Modification and optimization of nano-crystalline Al_2O_3 combustion synthesis using Taguchi L16 array. *Mater Res Bull* 46:1615–1624
10. Sasmal S, Goud VV, Mohanty K (2011) Optimisation of the acid catalysed pretreatment of areca nut husk fiber using the Taguchi design method. *Biosyst Eng* 110:465–472
11. Coşkun S, Motorcu AR, Yamankaradeniz N, Pulat E (2011) Evaluation of control parameters' effects on system performance with Taguchi method in waste heat recovery application using mechanical heat pump. *Int J Refrig* 35:795–809
12. Liu WL, Chien WT, Jiang MH, Chen WJ (2010) Study of Nd: YAG laser annealing of electroless Ni-P film on spiegel-iron plate by Taguchi method and grey system theory. *J Alloy Compd* 495:97–103
13. Huang CL, Hsu TS, Liu CM (2010) Modeling a dynamic design system using the Mahalanobis Taguchi System—two-step optimal based neural network. *J Stat Manage Syst* 13 (3):675–688
14. Huang CL, Lin CI, Tai SH (2012) The component search—two-steps optimal algorithm for data-mining in dynamic environments. *J Stat Manage Syst* 15(2, 3):249–260

Accurate Analysis of a Movie Recommendation Service with Linked Data on Hadoop and Mahout

Meng-Yen Hsieh, Gui-Lin Li, Ming-Hong Liao, Wen-Kuang Chou
and Kuan-Ching Li

Abstract This paper proposes a movie recommendation service with linked film-related data as dataset, and the recommender supplies recommendation services with a number of present item-based collaborative filtering algorithms. In order to effectively predict user movie preference, mechanisms developed in MapReduce are adopted for rapid individual recommendation computation, whereas the recommender is combined by three types of movie resources, MovieLens, Douban, and movie Trailer to be the dataset of recommendation service. NoSQL database is used to support for the movie dataset maintenance. The proposed prototype collects user feedback, the metrics, precision rate, recall rate, and F-score, and then applied to similarity mechanisms from Mahout.

Keywords Recommendation · Movie · Mapreduce tasks · Big Data

M.-Y. Hsieh (✉) · K.-C. Li
Department of Computer Science and Information Engineering,
Providence University, Taichung, Taiwan
e-mail: mengyen@pu.edu.tw

K.-C. Li
e-mail: kuancli@pu.edu.tw

G.-L. Li · M.-H. Liao
Software School, Xiamen University, Xiamen, China
e-mail: glli@xmu.edu.cn

M.-H. Liao
e-mail: liao@xmu.edu.cn

W.-K. Chou
Department of Computer Science and Information Management,
Providence University, Taichung, Taiwan
e-mail: wkchou@pu.edu.tw

1 Introduction

Due to advances of Internet technology, application service and digital data are found in explosive growth, and users always search through Internet application service to look for information they desire. However, they often spend large amount of time in querying with keywords, but still meet difficulties to find data with accuracy, even searching a single service domain. It is significant and important to satisfy immediate user real desire by supplying right information on application services [1, 2]. Consequently, to develop service recommendation is a useful alternative to user searching mechanism in mostly service domains.

Recommendation systems [3–6] assist users locate items that they could be interested in or might not found by themselves. These systems not only depend on data filtering algorithms, but also predict user's interest and likes. For example, the Amazon.com's recommendation system using collaborative filtering and provides general and personalized recommendation services, since Amazon website records and analyzes user online behavior of business browsing, shopping, rating, and others.

Most recommendation systems recommend users "objects" according to user's historical behavior, user similarity, and user's rating data. They are divided into two categories; the former one is content-based recommender system, and the latter is collaborative filtering recommender system. Content-based recommendation systems are composed by recommendation services for users by supplying similar products according to their historical preferences. When gaining more rich past records for user preference, the systems can provide more accurate recommendation services to users. Nevertheless, a content-based recommender system could has some disadvantages: (1) it is not easy to automatically analyze multimedia items that users chose before, (2) the system has no chances to recommend the products that user never bought before; (3) the product content relative to user purchase records is difficult to be interrupted precisely corresponding to user interest, while one item is represented with various semantics in products.

After a general instruction about recommendation service, the advantages and challenges depicted in this section, possible recommendation systems with collaborative filtering mechanisms related works are presented in Sect. 2. Section 3 describes the proposed recommendation architecture, while movie data is the dataset for recommendation, linking from few movie-based network resources. In Sect. 4, not only the experience of development environment for the architecture is shown, but also performance analysis and accuracy measurement between three legacy item-based CF recommendation services are proposed for the movie recommendation.

2 Related Work

Basic recommendation systems always recommend users and predict their preferences on business goods by purchase track, product similarity, and rating score. In [7], it is proposed a recommendation system to recommend users products from the

Amazon site according to their data in Twitter. The system retrieves meaningful terms for users, after interpreting and analyzing text semantics of data that users left in Twitter possibly including messages, articles and multimedia. Each of the terms gains an appropriate weight value as the term's significance according to what frequency they appeared in Twitter. With results obtained from MapReduce computations, the system would recommend the users with enough weights the relevant products from Amazon.

In [8], authors compared various similarity measurement algorithms relative to recommendation service. Accurate metrics on Hadoop and non-Hadoop environments are adopted to analyze the performance of these algorithms only with Movielens dataset. In [9], it is developed a content-based recommender system called Cinemappy, with a recommender engine leveraging graph information within DBpedia. The Cinemappy, as a mobile APP, provides movie information and movie theaters to users based on their profiles and current locations. Besides, the free available Web sources, *Google Place* or *Trovacinema* are presented as geographic criteria in the APP. The engine includes the three modules, denoted as Contextual Pre-filtering, Content-based Recommender, and Contextual Post-filtering. Pre-filtering scheme analyzes user profile. Then, recommender is responsible for similarity calculation between movies according to user interest. Post-filtering scheme provides movie theaters relative to the results of the previous modules.

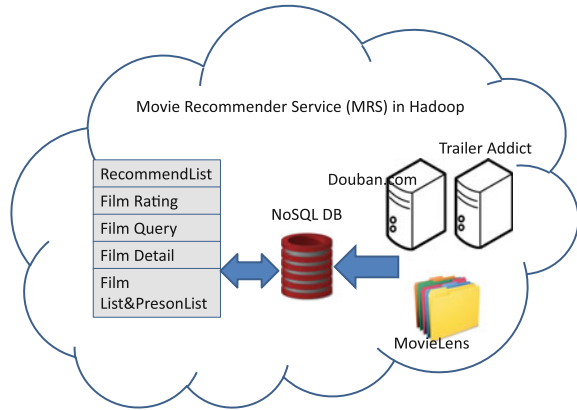
Ontology representation in a quickstep recommendation system to retrieve user preference is proposed [10]. The system creates user interest profile by extracting webpage topics that users browsed before, and user explicit feedback. Besides, recommendation calculation for users is scheduled every day according to user interest profiles and classified topics.

Content-based recommender systems cannot suggest users accurately because of ambiguous semantic problems of each term from user browsing or tracking records. At present, collaborative filtering (CF) better than content-based (CB) approaches are actively applied for personalized recommendation. More implicitly, user preference can be built by CF-oriented recommender or a mix of CF and CB approaches. Additionally, Big Data environment is required to calculate user preference while using CF approaches. Consequently, to build a recommendation service with linked data is an immutable trend, whilst large amount of available resources appropriate for users are free and public, then can be linked with additional data to originate a novel service useful to them.

3 Movie Recommendation Architecture

It is depicted in Fig. 1 the movie recommender architecture with necessary components operating in Hadoop.

Fig. 1 A movie recommendation architecture



3.1 Movie Linked Data

The main dataset adopted by this paper is from MovieLens. Douban movie descriptions and Movie Trailers are represented as the free resources of the linked data to support the proposed movie architecture capturing user feedback.

The dataset from MovieLens that the proposed architecture adopt includes four files, links.csv, movies.csv, ratings.csv, and tags.csv. Unfortunately, the architecture operates only with last two files. User rating data in ratings.csv is organized in such a way that records the rating score each user attributed for each of film in the dataset, and each rating score is on a scale from one to five. The information of film identity and title is set up in the movies.csv.

The architecture collects linked data based on the film data of MovieLens to query other two free available movie resources, and manages movie recommendation data in a NoSQL database.

One of the Douban’s released APIs involves the utilization of MovieLens to search film details by film name and release date. Each of the query results from the API includes twenty kinds of information formatted with JSON style related to the searching movie; for example, the Douban identity, the original and Chinese name of the movie, movie type, published year, among others. Ambiguous movies are always eliminated, while the movie name and its published years are matched for searching. In this research paper, only seven parts of the information are selected to the NoSQL database.

Table 1 illustrates the details of one movie in the database, integrated partial data from MovieLens and Douban. Data of Field id and name is belonging to MovieLens, and The others are from Douban. The two fields of name and published_year are the conditions for unambiguous movie searching.

Table 1 Illustration of a movie record in NoSQL database

Id	Name	Chinese_name	Published_year	Doubanid
1	Toy Story	玩具總動員	1995	1291575
Type		Adventure Animation Children Comedy Fantasy		
Image		http://img6.douban.com/view/movie_poster_cover/lpst/public/p1283671408.jpg		

Trailer Addict [11] is the other free available resource for capturing movie details about movie trailer, director, writer, stunt, among other information. One query with a unique movie can capture the information of about eight video trailers, including video name, linking address, and size. Each linking video has a basic screen resolution with 720 pixels.

3.2 Recommendation Service

The system adopts three similarity measurements for item-based collaborative filtering in Mahout. The first is Euclidean distance similarity to calculate preference vectors of each item. The shorter distance between user preference vectors in the dataset, the more similar vectors are. For instance, two users A and B give a rating score to items x and y , individually. Each user has a vector (x, y) in the space, and the Euclidean distance is the straight distance between A and B vectors. Since the distance similarity value could be greater than one, Mahout adjusts the value scaling from zero to one by the following equation:

$$sim(A, B) = \frac{1}{1 + d_{A,B}}, \quad \text{where } d_{A,B} = \sqrt{(B.x - A.x)^2 + (B.y - A.y)^2} \quad (1)$$

If case any two of users have rating scores to at least one same item, the similarity value will be not zero. In general, no similarity is possible without common items between user ratings.

Two vectors can have a Cosine similarity, while measuring the angular value instead of the cosine value of the angle between them. During similarity calculation, the vector direction is more important than the vector value. For example, users A and B have individual vectors for item ratings. The smaller the angle degree is, the similarity is greater they have, and vice versa. The Cosine similarity is different to the Euclidean distance similarity. Any of the two vectors is changed trending to the central point, they still have the same angle, but gain the different Euclidean distance.

Tanimoto coefficient similarity, called Jaccard similarity coefficient, calculates the relationship between items without the necessity of rating scores. For example,

two items, X and Y, have ratings from common users. Then, the similarity is defined as the ratio of intersection of ratings to difference between the sum of the ratings and intersection of the ratings, shown as in (2).

$$T(X, Y) = \frac{X \cap Y}{X \cup Y} \quad (2)$$

4 Performance Analysis and Accuracy Measurement

4.1 Accuracy Measurement

To understand whether the movie recommendation system is effective and usable to users, the precision rate and recall rate are represented as evaluation metrics of recommendation service for users.

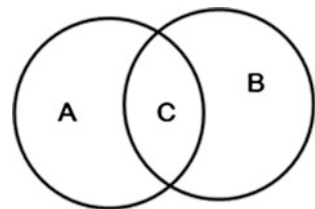
Figure 2 illustrates items for user preference. After user feedback, C is the number of suggested items that the user has interested in. When A is denoted as the number of the item that the user does not interest, A plus C represents the number of items that the system suggests to one user. When B is interested to items not recommended to the user, B plus C represents the number of all items in the service that the user interested in. The precision rate for recommendation service is defined in Eq. (3).

$$\text{Precision} = \frac{C}{A + C} \times 100\% \quad (3)$$

In order to calculate the number of interesting items to user that can be recommended, the recall rate is defined in Eq. (4).

$$\text{Recall} = \frac{C}{B + C} \times 100\% \quad (4)$$

Fig. 2 Illustration of all of items related to user preference



Precision measurement in some situations is conflicted between precision rate and recall rate. By combing these two rate metrics, F-score also called F-measure is a measure of a test’s accuracy in statistical analysis of binary classification. The traditional F-score is the harmonic mean of precision and recall, defined as follows:

$$F\text{-Score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

The higher the F-Score, better the evaluation results of recommendation algorithms is.

4.2 Recommender Experimentation

A prototype website is proposed to gain user feedback, while the recommendation service is implemented by Apache Mahout on Hadoop. Mahout with item-based CF algorithms provides fixed input formats, from which we can select different similarity mechanisms for running recommendation service. The only three chosen similarity mechanisms, Cosine, Euclidean, and Tanimoto are utilized to the recommendation service and the prototype, as depicted in Fig. 3. A user gains

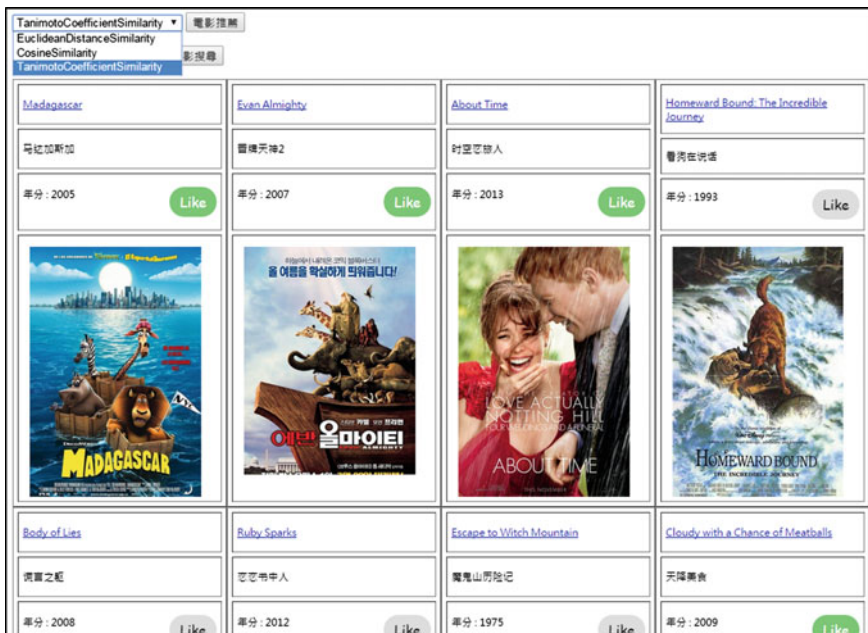


Fig. 3 Suggested movies to a user from the recommendation service

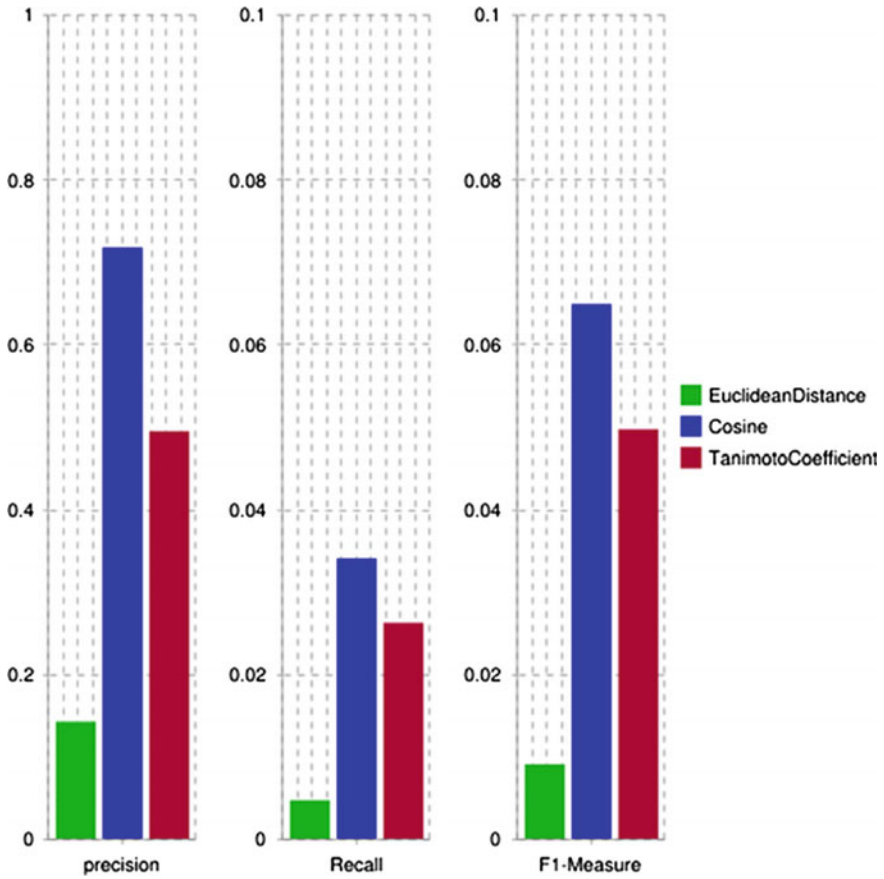


Fig. 4 Evaluation of three recommendation services

suggested movies after authenticated by his/her Facebook account. The user can return a feedback to the system about the movies he likes or not.

The prototype has been evaluated that gave feedback to the movies to three recommendation mechanisms independently. Figure 4 represents the three bar charts after measuring each of three metrics, precision rate, recall rate, and F-score with the recommendation service. The first chart is the statistics of precision rate about 18 % for Euclidean similarity, 71 % for Cosine similarity, and 50 % for Tanimoto similarity. The second depicts the recall rate about 0.5 % for Euclidean similarity, 3.4 % for Cosine similarity, and 2.6 % for Tanimoto similarity. The final one is for F-score, where the measurements of Euclidean, Cosine, and Tanimoto similarities are about 0.9, 6.5, and 5 %, individually.

The chosen dataset from MovieLens consists of 6000 films, so the subjects only draw a small part of all interesting movies of the system in this short evaluation duration. It is not possible to calculate the number of all interesting films in the

system for any user. Therefore, the system calculates the number of all films with the categories that user ever interested instead of the number of the films that the user liked in practice. The film number in the dataset is huge, so the evaluation results of Recall and F-score got the rates much smaller than Precision. Consequently, the little rate value did not influence the comparison of the three similarities. In this experience, the Cosine similarity in our movie recommendation system is an accurate prediction to users.

5 Conclusion

This paper proposes a recommender system to suggest films to users during their time for entertainment, since recommendation is an alternative way to help users searching and finding their interested items. As users rate the films, the system will predictably recommend interesting films implicit to them. Through a prototype website interacting to users for feedback data, the tree metrics are adopted to compare with three similarity mechanisms operating in recommender system. It is possibly that only few subjects has joined in the system during short evaluation period, and the Cosine similarity mechanism is better than the others. In the future work, given that the development of suitable mobile context-aware clients to the recommendation service is a trend, the recommender may be required to supply real-time data adaptively based upon to their current environment and situation. Taking into consideration privacy issues, secured MapReduce computations for recommendation is aimed in task design model [12], as well as review MapReduce computations over multi-GPU environments [13, 14].

Acknowledgments This investigation was supported by Ministry of Science and Technology, Taiwan, under grant no. MOST 103-2221-E-126-010, and Providence University research grant.

References

1. Hsieh MY, Huang T-C, Hung JC, Li K-C (2015) Analysis of gesture combos for social activity on Smartphone. *Lect Notes Electr Eng* 329:265–272
2. Hsieh MY, Yeh CH, Tsai YT, Li KC (2014) Toward a mobile application for social sharing context. *Lect Notes Electr Eng* 274:93–98
3. Haspra M (2011) Scalable real-time product recommendation based on users activity in a social network. Systems Group, Department of Computer Science, ETH Zurich
4. Grivolla J, Badia T, Campo D, Sonsona M, Pulido J-M (2014) A hybrid recommender combining user, item and interaction data, IEEE
5. Bhutkar B, Aghav J, Dorwani S (2013) Data management using apache cassandra. Department of Computer Engineering and, Information Technology, College of Engineering, Pune, India
6. Dede E, Sendir B, Kuzlu P, Hartog J, Govindaraju M (2013) An evaluation of cassandra for hadoop. Grid and Cloud Computing Research Laboratory SUNY Binghamton

7. Haspra M (2011) Scalable real-time product recommendation based on users activity in a social network. Systems Group, Department of Computer Science, ETH Zurich
8. Senthil Kumar T, Neetha Susan T, Johnpaul CI (2013) Performance analysis of various recommendation algorithms using apache hadoop and mahout. *Int J Sci Eng Res* 4 (12):279–287
9. Vito CO et al (2013) Mobile movie recommendations with linked data. Availability, reliability, and security in information systems and HCI. *Lect Notes Comput Sci* 8127: 400–415
10. Middleton SE et al Capturing knowledge of user preferences: ontologies in recommender systems. In: K-CAP '01 proceedings of the 1st international conference on knowledge capture, ACM, New York, pp 100–107
11. Movie Trailer [online]. Available: <http://www.traileraddict.com/trailerapi>
12. Lin H-Y, Hsieh M-Y, Li K-C (forthcoming) Secured map reduce computing based on virtual machine using threshold secret sharing and group signature mechanisms in cloud computing environments. *Telecommunication systems*, Springer, Berlin
13. Jiang H, Chen Y, Qiao Z, Li K-C, Ro W, Gaudiot J-L (2014) Accelerating MapReduce framework on multi-GPU systems. *Cluster Comput* 17(2):293–301
14. Chen Y, Qiao Z, Jiang H, Li K-C, Ro WW (2013) MGMR: multi-GPU based MapReduce. In: GPC'2013 the 7th international conference on grid and pervasive computing. LNCS 7861, Springer, Korea

A Method of Event Ontology Mapping

Xu Wang, Wei Liu, Yujia Zhang, Yue Tan and Feijing Liu

Abstract Ontology mapping is an important solution to ensure interoperability while integrating heterogeneous and distributed data sources. This paper proposes an approach for ontology mapping based on event, which enable the mapping between event-based information with more abundant semantics. Firstly, this paper gives the definition of event ontology mapping, and then proposes an comprehensive semantic similarity calculation model based on the similarity of events and event structures. Experiments show that the proposed model can effectively find the semantic relations of event in two event ontologies.

Keywords Event ontology mapping · Semantic similarity · Semantic neighbor

1 Introduction

Ontology is the core of semantic web, and it is widely used in the areas such as data integration, metadata management and data exchange. However, due to various cultural backgrounds and comprehensions of ontology, different usage habits of terms leads to an ocean of ontologies with different structures, which causes the limited integration and limited sharing of data. In order to accomplish interoperability of semantics ontology, it is necessary to establish mapping relations among heterogeneous ontologies. Whereas most current ontology matching methods are based on calculation of concepts correspondence [1], it will result in semantic information loss while processing event-based data because of “Tennis Problem” [2]. In recent years, research on the use of events as a key concept for representing knowledge, organizing and structuring media on the web is surging, especially in

X. Wang · W. Liu (✉) · Y. Zhang · Y. Tan · F. Liu
School of Computer Engineering and Science, Shanghai University,
Shanghai 200444, China
e-mail: liuw@shu.edu.cn

X. Wang
e-mail: wangx89@126.com

semantic web community. Event ontology is a shared, formal and explicit specification of an event class system model that exists objectively, which has become new paradigm for describing and reasoning event-based knowledge in web. Event-based ontology matching is necessary for integrating heterogeneous data sources in event-centered domain application, such as emergency response, public opinion monitoring, history and cultural heritage, etc.

Aims at the event ontology mapping in semantic web application, this paper firstly gives the definition of event ontology mapping, and then proposes a semantic similarity calculation model based on the similarity of event classes and event class structure. The similarity of event classes is calculated by the similarity of common elements in different event. The similarity of event class structure is obtained by calculating the similarity of semantic neighbor sets. The semantic neighbors of event class could be found in a certain semantic radius in event ontology network structure.

The paper is organized as follows. Section 2 reviews the related work. Section 3 introduces concepts about event ontology. Section 4 proposes the mapping approach of event ontology. We present experimental results and analysis in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Related Work

The ontology mapping problem has been researched extensively in the past decade, yet, despite this research, it is still considered an “unsolved” problem [3]. Currently, basic matcher is a similarity function of a pair of entities, $\sigma : o \times o \rightarrow R$ where R is $[0,1]$. In point-to-point approach, matching uses lexical or structural similarity of labels or instances. Current ontology mapping methods can be categorized as: terminological mapping [4], structural mapping [5] and semantic technique [6]. Ontology mapping have gained a large amount of attention and regarded as an important solution that enables interoperability across heterogeneous systems and semantic web applications. A single mapping method maybe not afford all mapping tasks, consequently, different methods sometimes are combined to realize concept mapping in ontologies.

According to [4], currently methods of ontology mapping face several difficulties, such as difficulty in processing word variations in the same ontology or across ontologies while terminological mapping, difficulty in processing many kinds of variations that occur in ontologies while structural mapping. At the same time, ontology requires inductive inputs but semantics technique is deductive in nature, currently, there is a lack of interoperability between inductive technique and deductive semantic techniques.

From our perspective, event-based concept mapping is arduous to be accomplished by using traditional conceptualized mapping technologies. Event-based information involves objects, action, time, location and so on, the traditional concept mapping usually results in loss of semantics. In additional, relations between events

contains more semantic information than the hierarchical relationships between concepts. How to use these semantic relations to enhance the mapping between the event ontologies is worthy of study. This is also the motivation of this paper.

3 Concepts About Event Ontology

Definition 1 (Event) We defined event as a thing happening in a certain time and place, which is involved in some actors, objectives and action features with statuses changing. Event e is defined as a 6-tuple formally:

$$Event ::= \langle A, O, T, P, S, L \rangle$$

where, A means an action happen in an event, O means actors and objects involved in an event, T means instant and interval time of an event, P means the place that an event happens in, S means statuses of object O before and after an event, L indicates linguistic expressions of text-based event, it includes the language expression of events and event elements. L is also the key parameter to calculate the similarity between event classes [7].

Definition 2 (Event Class) Event class is an abstract event that represents a set of events with some common characteristics, denoted as EC :

$$EC = (E, C_A, C_O, C_T, C_P, C_S, C_L)$$

$C_i = \{c_{i1}, c_{i2}, \dots, c_{im}, \dots\}$ ($i \in \{A, O, T, P, S, L\}$, $m \geq 0$) where E means an event set, called extension of the event class, C_i denotes the common characteristics set of certain event element (element i), called intension of the event class and event element class, c_{im} denotes one of the common characteristics of event factor i .

Definition 3 (Event Ontology) An Event ontology EO is a shared, formal and explicit specification of an event class system model that exists objectively. Event ontology EO can be defined as a 4-tuple formally:

$$EO = \langle UECs, ECs, R, Rules \rangle,$$

where:

1. $UECs$ is a set of upper event classes in the event ontology, each UEC represents a category, all of the UEC constitute a tree category structure of an event ontology;
2. $ECs = \{EC_1, EC_2, \dots, EC_n\}$ is a set of event classes;
3. $R = \{r \mid r \text{ is the relation of } \langle EC_i, EC_j \rangle\}$, $r \in \{R_{is_a}, R_{compOf}, R_{cause}, R_{follow}, R_{concur}\}$. R_{is_a} means subsumption relations, R_{compOf} means composition relations, R_{cause} means causality relations, R_{follow} means follow relations, R_{concur} means concurrency relations.

4. *Rules* is set of rules be expressed in logic languages, including rules for inference of upper level event categories and rules for relations inference among events.

4 Event Ontology Mapping

4.1 Definitions About Event Ontology

Event ontology mapping is to specify how the event classes in various event ontologies map to each other. In this section, we defined two concepts about event ontology.

Definition 4 (*Event Class Mapping Consistency*) Event ontologies EO_1 , EO_2 , we use C to denote the consistency of the two event classes EC_1 and EC_2 in EO_1 and EO_2 , defined as a 4-tuple:

$$C ::=_{def} \langle id, EC_1, EC_2, S \rangle$$

where id means the unique identification of the consistency, EC_1 and EC_2 represent event classes, S represents semantic similarity between event class EC_1 and EC_2 , matching $S \in [0, 1]$. User can set the threshold value of S according to the requirement of application. It means EC_1 is consistence with EC_2 while the similarity is above the value of S .

Definition 5 (*Event Ontology Mapping*) Event ontology EO_1 , EO_2 , event ontology mapping M is represented by a set of consistency of mapping.

$$M ::=_{def} (c_1, c_2, \dots, c_n) \quad (0 < n \leq |EO_1 \leftrightarrow EO_2|)$$

where $|EO_1 \leftrightarrow EO_2|$ is the pairs count of mapping between EO_1 and EO_2 .

4.2 Event Ontology Comprehensive Semantic Similarity Calculation Model

According to the event class definition above, an event class is composed of six elements. So, the similarity between two event classes basically results from the similarities between their common elements (such as action, participants and places). As well, the relationships among event classes provide context semantics for event classes, and it is rather easy to notice that two event classes with high similarity always have similar relation structure with their neighbors in event ontology. In this section, we present an event ontology comprehensive semantic similarity calculation model based on features of event inner structure and event

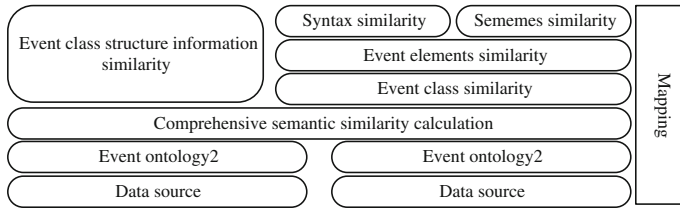


Fig. 1 A simple high level view of a mapping process

relationship network structure. As shown in Fig. 1, the calculation process of event class similarity includes two parts. First is the calculation of the event class similarity, which can be calculated through the elements similarity of event classes. Second is the calculation of event class structures similarity. As mentioned above, the relations between the events are divided into two categories: taxonomic relations and non-taxonomic relations. We can use semantic neighbors of event classes to express the semantic relationships among event classes. According to the event class network structure in event ontology, taking the event class as the center and setting a semantic radius r . In this range, the value of r reflects the close or distant relation between event classes. While the calculation of event classes similarities in the first step, we got a similarity distribution table. Some of the similarity value might be very small or even zero, which means sparse relation between event classes, but in fact, some implicit semantic relations between them can be extracted in the semantic neighbor sets. We take the biggest similarity of event classes in the semantic neighbor sets as the similarity of semantic neighbor sets. It would eventually get a similarity distribution table. Furthermore, two similarity distribution tables are integrated together. Finally the biggest similarity value corresponding to the event class is taken as the final similarity value.

4.2.1 Event Class Similarity Calculation

Event class is an abstract event that represents a set of events with common characteristics. The event class EC_1 is denoted as $EC_1 = \{e_{11}, e_{12}, \dots, e_{1m}\}$, while the event class EC_2 is denoted as $EC_2 = \{e_{21}, e_{22}, \dots, e_{2m}\}$. The similarity between EC_1 and EC_2 can be calculated as follows:

$$\text{sim}(EC_1, EC_2) = \frac{1}{m * n} \sum_{i=1 \dots n, j=1 \dots m} \text{sim}(arg_{1i}, arg_{2j}) (arg \in \{a, o, t, p, s\}) \quad (1)$$

where $\text{sim}(arg_{1i}, arg_{2j})$ means the similarity of event elements. The calculation of elements similarity will be introduced in detail below.

While modeling an event class with 6-tuple event model, event elements described with natural language (*objects, action and place*) or assertion expressions (*time and status*), which may be a simple word, a phrase or a sentence. So, the

calculation of elements similarity include the syntax similarity and semantics similarity. We utilize the method in [8] to calculate the syntax similarity of elements.

Definition 6 (*Syntax Similarity Calculation*)

$$\text{sim}_{\text{syntax}}(\text{con}_1, \text{con}_2) = \frac{2 \sum_j \text{length}(\max \text{SameSubString}_j(\text{con}_1, \text{con}_2))}{\text{length}(\text{con}_1) + \text{length}(\text{con}_2)} \quad (2)$$

where $\max \text{SameSubString}_j(\text{con}_1, \text{con}_2)$ means that the j th longest common substring of the two elements: con_1 and con_2 . Finding the longest common substring of these elements and remove it from the original string. Then continue to find the longest common substring from the rest string, until there is no longest common substring.

While calculating semantics similarity, we query the concept from HowNet [9], which has rich semantics information about concepts and inter-conceptual. Similarity between simple words is related to the calculation of the sememe similarity and concept similarity. We utilize the method in [10] to calculate the sememe similarity of words.

Definition 7 (*Similarity Calculation of Sememes*)

$$\text{sim}(p_1, p_2) = \frac{\text{comLevel}}{\text{height}_{\text{tree}} + \text{dis}_{p_1, p_2}} \quad (3)$$

where the “comLevel” is the minimum common parent node level, the $\text{height}_{\text{tree}}$ is the depth of sememes-tree, $\text{dis}_{\{\{p_1, p_2\}\}}$ is the path distance of the sememes. The “notional word” in HowNet described as follows:

$$\text{semantic item:} \left(\begin{array}{l} \text{The first basic semene Description} = \text{basic semene} \\ \text{The other basicsemene} = \{\text{basic semene}_b, \text{basic semene}_c, \dots\} \\ \text{Relation semene description} = \left(\begin{array}{l} \text{Relation semene}_1 = \text{basic semene}_x | \text{word}_x \\ \text{Relation semene}_2 = \text{basic semene}_y | \text{word}_y \\ \dots \end{array} \right) \\ \text{Relation symbol description} = \left(\begin{array}{l} \text{Relation symbol}_1 = \text{semene}_u | \text{word}_u, \text{semene}_v | \text{word}_v, \\ \text{Relation symbol}_2 = \text{semene}_s | \text{word}_s, \text{semene}_t | \text{word}_t, \\ \dots \end{array} \right) \end{array} \right)$$

Each part of the description is a set of sememes. It is necessary to calculate the sets of sememes’ similarity before calculating the similarity between two concepts. The similarity of sememes’ similarity can be expressed as follows.

Definition 8 (*Similarity Calculation of Sememe Sets*)

$$\text{sim}(\text{set}_1, \text{set}_2) = \frac{|\text{sem}_{\text{set}_1 \leftrightarrow \text{set}_2}|}{|\text{set}_1| + |\text{set}_2|} \left(\frac{\sum_{i=1}^{|\text{pair}_{\text{set}_1 \leftrightarrow \text{set}_2}|} \text{sim}(p_{1i}, p_{2i})}{|\text{pair}_{\text{set}_1 \leftrightarrow \text{set}_2}|} \right) \quad (4)$$

where set_1 and set_2 denote the set of sememes, $|set_1|$ is the count of sememes in sets, $|sem_{set_1 \leftrightarrow set_2}|$ is the total count of sememes in two sets which have semantic relations, $|pair_{set_1 \leftrightarrow set_2}|$ is the count of pairs which have semantic relations in two sets. The semantic similarity between two concepts is composed of the four portions in notional word described above in HowNet. The similarity calculation is as follows.

Definition 9 (Concept Similarity Calculation)

$$\text{sim}(con_1, con_2) = \sum_{i=1}^4 \beta_i \prod_{j=1}^i \text{sim}_j(set_1, set_2) \quad (5)$$

where β_i is the degree of each part's influence in the notional word's concept, the values of the degree decrease gradually, and $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$. We can get the empirical value of β_i : $\beta_1 = 0.5$, $\beta_2 = 0.2$, $\beta_3 = 0.17$, $\beta_4 = 0.13$.

There are several semantic descriptions for $SimpleWord_1$ and $SimpleWord_2$ in HowNet. If $SimpleWord_i = \{con_{i1}, con_{i2}, \dots, con_{im}\} (i = 1, 2)$, we can get the words' similarity as follows:

$$\text{sim}(simpleWord_1, simpleWord_2) = \max_{i=1 \dots n, j=1 \dots m} \text{sim}(con_{1i}, con_{2j}) \quad (6)$$

The result will be the maximum value of semantic concepts' similarity. Or a set of sequences, $sequence_i = \{simpleWord_{i1}, simpleWord_{i2}, \dots, simpleWord_{im}\} (i = 1, 2)$, then the formula of word sequences' similarity calculation as follows:

Definition 10 (Word Sequence Similarity Calculation)

$$\text{sim}(seq_1, seq_2) = \frac{|sem_{set_1 \leftrightarrow set_2}|}{|seq_1| + |seq_2|} \left(\frac{\sum_{i=1}^{|pair_{set_1 \leftrightarrow set_2}|} \text{sim}(simpleWord_{1i}, simpleWord_{2i})}{|pair_{set_1 \leftrightarrow set_2}|} \right) \quad (7)$$

where $|seq_1|$ is the count of simple words in set, $|sem_{seq_1 \leftrightarrow seq_2}|$ is the total count of words which have semantic relationships in set_1 and set_2 . $|pair_{seq_1 \leftrightarrow seq_2}|$ is the pairs count of sets which has semantic relations in two sets. If there is only a simple word in both of the sequences set, it is equivalent to the formula (6).

Definition 11 (Elements Similarity Calculation)

$$\text{sim}(arg_1, arg_2) = \text{wgt}_{arg_syn} * \text{sim}_{syntax}(arg_1, arg_2) + \text{wgt}_{arg_sem} * \text{sim}_{semantic}(arg_1, arg_2) \quad (arg \in \{a, o, t, p, s\}) \quad (8)$$

where wgt_{arg_syn} and wgt_{arg_sem} are weights of syntax similarity and semantic similarity, $\text{wgt}_{arg_syn} + \text{wgt}_{arg_sem} = 1$, we take sigmoid to calculate the values of wgt_{arg_syn} and wgt_{arg_sem} . The calculation of event similarity could be obtained as follows:

$$sim(event_1, event_2) = \sum_{i=1}^5 w_{arg} * sim(arg_1, arg_2) (arg \in \{a, o, t, p, s\}) \quad (9)$$

$w_a + w_o + w_t + w_p + w_s = 1$, the weight assignment is based on sigmoid function. In the formula above, each element is assigned a weight which means its importance in the event. $\sum_{i=1}^5 weight = 1$, $arg_{i(i=1,2,\dots,5)}$ means elements of an event.

4.2.2 Calculation of Event Class Structure Information Similarity

Event ontology is a network that composed of event classes and relations. The similarity of event classes structure information defined as $sim_S(ECS_1, ECS_2)$. ECS_1 denotes an event class structure with a set of semantic nodes, it taking EC_1 as the center and a set of neighbor nodes with a semantic radius r (each element in ECS_1 has a distance p with EC_1 , and $p \leq r$). The node in ECS_1 representation as a triple, $\langle pre_eventClass, relation, eventClass \rangle$, where $pre_eventClass$ denotes its preceding node, $relation$ means event relation between these two nodes, $eventClass$ denotes event class in this node. ECS_2 has the same structure.

The algorithm of calculation for event class structure information similarity is given as below:

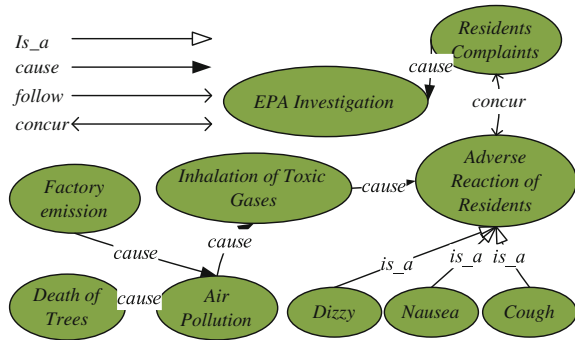
```

path = 1; r = 5; ThresholdValue tv; sim(ECS1, ECS2) = 0; Node preA = EC1; preB = EC2;
ECS1 = {EC1, neighbor node set}; ECS2 = {EC2, neighbor node set};
while(path <= r){//5
  List nodesA = preA.directNeighborNodes; List nodesB = preB.directNeighborNodes;
  for(i = 0; i < nodesA.length; i++){//4
    Node A = nodesA[i];
    maxSimilarityValue = 0;
    for(j = 0; j < nodesB.length; j++){//3
      B = nodesB[j];
      if(sim(A.eventClass, B.eventClass) >= tv && sim(A.eventClass, B.eventClass) >
maxSimilarityValue){ //2
        similarity = sim(A.eventClass, B.eventClass)/path;
        if(A.relation == B.relation){//1
          similarity = similarity * (1 + (path / r * 10));
        }//1
        maxSimilarityValue = sim(A.eventClass, B.eventClass);
      }//2
    }//3
    simess(ECS1, ECS2) = Simess(ECS1, ECS2) + similarity;
  }//4
  path = path + 1; preA = A; preB = B;
}//5

```

In the algorithm, the `directNeighborNodes` means a function to obtain all neighbor nodes with path is 1. Eventually, the comprehensive semantic similarity formula is $max\{sim(ECS_1, ECS_2), sim(EC_1, EC_2)\}$, and the similarity between EC_1 and EC_2 can be obtained by using the formula (1).

Fig. 2 Event ontology of air pollution caused by factory gas emissions



5 Experiment and Analysis

In this section, we present the process of the similarity calculation between two sample event ontology, and try to find semantics relation mapping between them.

Before the similarity calculation of event classes, by analyzing and event-oriented annotating more than 60 articles about air pollution and water pollution from internet, and under the instruction of environmental experts, we create two ontology models as shown in Figs. 2 and 3. The air pollution event ontology describes factory gas emission would result in air pollution; and air pollution would result in inhalation of toxic gas and death of trees; inhalation of toxic gas would cause health hazard and complaints from residents; health hazard includes adverse reaction of residents, such as dizzy, nausea and cough, and severely health injury, such as poisoning and death; complaints would result in investigation from environmental protect administration (EPA). The water pollution ontology describes a series of event classes caused chemical leakage of vehicle, as shown in Fig. 3.

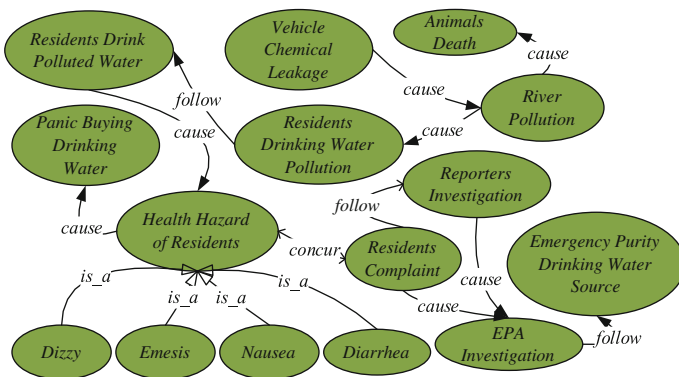


Fig. 3 Event ontology of water pollution caused by vehicle chemical leakage

5.1 Event Class Similarity and Event Class Structure Information Similarity

The events are from the news report on the internet, and the elements of event are annotated manually. Event ontology A represents “*air pollution incident caused by factory gas emissions*”; event ontology B represents “*water pollution incident caused by vehicle chemical leakage*”, the number of events are listed in Table 1 as follows.

According to the calculation methods of the event class similarity, the results of similarity among event classes are shown in Table 2.

For the calculation of event class structure similarity, we need to get the semantic neighbor sets of each event class, calculating the semantic similarity between them by using the given formula given in Sect. 4.2.1. From the event ontology figures above, the semantic neighbor sets of each event class can be found out, such as “*air pollution*” semantic neighbor sets are {*factory emission, inhalation of toxic gases, death of trees*}. The similarity of structure information is calculated by using the algorithm given in Sect. 4.2.2 (Table 3).

When the results of two calculations are tabulated, choosing the largest value as the semantic similarity between event classes, then get the comprehensive semantic similarity calculation results as shown in Table 4.

Table 1 Events number of event ontology A and event ontology B

Name of event classes in event ontology A	Event num	Name of event classes in event ontology B	Event num
Factory emission EC _{A1}	20	Vehicle chemical leakage EC _{B1}	15
Air pollution EC _{A2}	20	River pollution EC _{B2}	20
Inhalation of toxic gases EC _{A3}	15	Residents drinking water pollution EC _{B3}	15
Death of trees EC _{A4}	15	Residents drink polluted water EC _{B4}	20
Adverse reaction of residents EC _{A5}	15	Animals death EC _{B5}	20
Residents' complaints EC _{A6}	20	Panic buying drinking water EC _{B6}	20
EPA investigation EC _{A7}	25	Health hazard of residents EC _{B7}	40
Dizzy EC _{A8}	10	Residents complaint EC _{B8}	20
Nausea EC _{A9}	10	Reporters investigation EC _{B9}	15
Cough EC _{A10}	10	EPA investigation EC _{B10}	25
		Emergency purity drinking water source EC _{B11}	20
		Dizzy EC _{B12}	10
		Emesis EC _{B13}	10
		Nausea EC _{B14}	10
		Diarrhea EC _{B15}	10

Table 2 Similarity results of event class

	EC _{A1}	EC _{A2}	EC _{A3}	EC _{A4}	EC _{A5}	EC _{A6}	EC _{A7}	EC _{A8}	EC _{A9}	EC _{A10}
EC _{B1}	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
EC _{B2}	0.000	0.342	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
EC _{B3}	0.000	0.101	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000
EC _{B4}	0.000	0.000	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
EC _{B5}	0.000	0.000	0.000	0.303	0.000	0.000	0.000	0.000	0.000	0.000
EC _{B6}	0.000	0.000	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001
EC _{B7}	0.000	0.000	0.001	0.000	0.734	0.001	0.000	0.001	0.001	0.001
EC _{B8}	0.000	0.001	0.001	0.000	0.001	0.802	0.000	0.001	0.001	0.001
EC _{B9}	0.001	0.000	0.000	0.000	0.000	0.001	0.341	0.000	0.000	0.000
EC _{B10}	0.001	0.002	0.000	0.000	0.000	0.001	0.872	0.000	0.000	0.000
EC _{B11}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EC _{B12}	0.000	0.000	0.001	0.000	0.201	0.001	0.000	0.803	0.324	0.121
EC _{B13}	0.000	0.000	0.001	0.000	0.114	0.001	0.000	0.213	0.847	0.092
EC _{B14}	0.000	0.000	0.001	0.000	0.201	0.001	0.000	0.191	0.207	0.108
EC _{B15}	0.000	0.000	0.001	0.000	0.215	0.001	0.000	0.102	0.133	0.103

Table 3 Distribution of event class structure similarity

	EC _{A1}	EC _{A2}	EC _{A3}	EC _{A4}	EC _{A5}	EC _{A6}	EC _{A7}	EC _{A8}	EC _{A9}	EC _{A10}
EC _{B1}	0.342	0.000	0.342	0.342	0.000	0.001	0.000	0.000	0.000	0.000
EC _{B2}	0.101	0.303	0.101	0.303	0.000	0.001	0.001	0.000	0.000	0.000
EC _{B3}	0.342	0.001	0.342	0.342	0.001	0.001	0.000	0.001	0.001	0.001
EC _{B4}	0.101	0.001	0.734	0.101	0.001	0.734	0.001	0.734	0.734	0.734
EC _{B5}	0.342	0.000	0.342	0.342	0.000	0.001	0.000	0.000	0.000	0.000
EC _{B6}	0.000	0.001	0.734	0.000	0.001	0.734	0.000	0.734	0.734	0.734
EC _{B7}	0.001	0.001	0.215	0.001	0.803	0.847	0.802	0.215	0.215	0.215
EC _{B8}	0.002	0.001	0.734	0.002	0.001	0.734	0.001	0.734	0.734	0.734
EC _{B9}	0.002	0.002	0.002	0.002	0.802	0.872	0.802	0.001	0.001	0.001
EC _{B10}	0.001	0.001	0.001	0.001	0.802	0.001	0.802	0.001	0.001	0.001
EC _{B11}	0.002	0.000	0.002	0.001	0.001	0.872	0.001	0.000	0.000	0.000
EC _{B12}	0.000	0.001	0.734	0.000	0.001	0.734	0.001	0.734	0.734	0.734
EC _{B13}	0.000	0.001	0.734	0.000	0.001	0.734	0.001	0.734	0.734	0.734
EC _{B14}	0.000	0.001	0.734	0.000	0.001	0.734	0.001	0.734	0.734	0.734
EC _{B15}	0.000	0.001	0.734	0.000	0.001	0.734	0.001	0.734	0.734	0.734

Table 4 Comprehensive semantic similarity value

	EC _{A1}	EC _{A2}	EC _{A3}	EC _{A4}	EC _{A5}	EC _{A6}	EC _{A7}	EC _{A8}	EC _{A9}	EC _{A10}
EC _{B1}	0.342	0.000	0.342	0.342	0.000	0.001	0.001	0.000	0.000	0.000
EC _{B2}	0.101	0.342	0.101	0.303	0.000	0.001	0.001	0.000	0.000	0.000
EC _{B3}	0.342	0.101	0.342	0.342	0.001	0.001	0.001	0.001	0.001	0.001
EC _{B4}	0.101	0.001	0.734	0.101	0.001	0.734	0.001	0.734	0.734	0.734
EC _{B5}	0.342	0.000	0.342	0.342	0.000	0.001	0.000	0.000	0.000	0.000
EC _{B6}	0.000	0.001	0.734	0.000	0.001	0.734	0.000	0.734	0.734	0.734
EC _{B7}	0.001	0.001	0.215	0.001	0.803	0.847	0.802	0.215	0.215	0.215
EC _{B8}	0.002	0.001	0.734	0.002	0.001	0.802	0.001	0.734	0.734	0.734
EC _{B9}	0.002	0.002	0.002	0.002	0.802	0.872	0.802	0.001	0.001	0.001
EC _{B10}	0.001	0.001	0.001	0.001	0.802	0.001	0.872	0.001	0.001	0.001
EC _{B11}	0.002	0.000	0.002	0.001	0.001	0.872	0.001	0.000	0.000	0.000
EC _{B12}	0.000	0.001	0.734	0.000	0.201	0.734	0.001	0.803	0.734	0.734
EC _{B13}	0.000	0.001	0.734	0.000	0.114	0.734	0.001	0.734	0.734	0.734
EC _{B14}	0.000	0.001	0.734	0.000	0.201	0.734	0.001	0.734	0.734	0.734
EC _{B15}	0.000	0.001	0.734	0.000	0.215	0.734	0.001	0.734	0.734	0.734

6 Conclusion

In this paper, we defined the concept of event ontology mapping, and proposed a comprehensive semantic similarity calculation model based on the event classes and event class structure information to accomplish the mapping between event ontologies. The experimental results show that the proposed calculation model can be used to find semantic mapping between event classes from different ontologies effectively. Because of the complexity of event ontology, there are still some problems need to be solved in the future. Firstly, the limitation of vocabulary coverage in HowNet affected the precision of semantic similarity calculation. Secondly, the semantic neighbors of event classes are not always be selected accurately.

Acknowledgements This paper is supported by the Natural Science Foundation of China, No. 61305053 and No. 61273328.

References

1. Seddiqui MH, Aono M (2009) An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size. *Web Semant Sci Serv Agents* 7(4):344–356
2. Espinoza M, Gómez-Pérez A, Mena E (2008) *Enriching an ontology with multilingual information*. Springer, Berlin
3. Noy NF (2004) Semantic integration: a survey of ontology-based approaches. *ACM Sigmod Record* 33(4):65–70

4. Hooi YK, Hassan MF, Shariff AM (2014) A survey on ontology mapping techniques. In: *Advances in computer science and its applications*. Springer, Berlin, pp 829–836
5. Euzenat J, Shvaiko P (2007) *Ontology matching*. Springer, Heidelberg
6. Kotis K, Vouros GA, Stergiou K (2006) Towards automatic merging of domain ontologies: The HCONE-merge approach. *Web Semant* 4(1):60–79
7. Zongtian L, Meili H et al (2009) Research on event-oriented ontology. *Computer Science* 36 (11):189–192 (in Chinese)
8. Shan Bi (2011) *Ontology mapping algorithm based on the concept of similarity calculation*. Beijing Jiaotong University, Beijing (in Chinese)
9. Dong Z, Dong Q (2006) *HowNet and the computation of meaning*. World Scientific, Singapore
10. Jia L (2007) *Chinese ontology mapping based on HowNet*. Beijing University of Posts and Telecommunications, Beijing (in Chinese)

A Research on Multi-dimensional Multi-attribute String Matching Mechanism for 3D Motion Databases

Edgar Chia-Han Lin

Abstract Due to the development of computer technology and the mature development of 3D motion capture technology, the applications of 3D motion databases become more and more important. How to analysis the huge data stored in the database and efficiently retrieved the matched data is an important research issue. 3D animation design is one of the important applications of 3D motion databases. Based on our teaching experience, the bottleneck of the students' learning of 3D animation is the motion animation of the 3D characters. Therefore, the 3D motion database can be used to assist the design of the motion for 3D characters. However, it is still a difficult problem because of the high complexity of the matching mechanism and the difficult of user interface design. In this paper, the 3D motion data can be represented as multi-dimensional multi-attribute sequences while the corresponding index structures and query processing mechanism are proposed for efficiently processing the 3D motion queries. Moreover, Microsoft Kinect is used in this project as the user interface. The captured data can be used as the user query and the further comparison will be performed to find the matched motion data.

Keywords 3D motion database · Index structure · Query processing · Kinect

1 Introduction

Due to the great improvement of technology and the mature development of 3D motion capture technology, the applications of 3D motion database become more important in recent years. How to analyze the large amount of 3D motions recorded in the databases to efficiently retrieve the desired motions become an important

This work was partially supported by Asia University (Project No. 103-asia-09).

E.C.-H. Lin (✉)

Department of Information Communication, Asia University, Taichung, Taiwan
e-mail: edgarlin@asia.edu.tw

research issue. 3D motion data is a time sequence data which is formed by the series of reference coordinate values in different locations on the human body. Due to the large amount of reference coordinate values and the characteristics of time series data, the analysis and retrieval of 3D motion data is time consuming. There are many researches which focus on the complexity reduction of the motion data or the index structure construction for the motion data to enhance the efficiency of data comparison.

In [1], the complex multidimensional time series data is separated into several smaller segments to reduce the complexity of data. Furthermore, the segments are used to reduce the difficulty of the matching mechanism. In these years, many researches focus on the representation of the motion data. In [2], the geometric relationships between reference points on the body are defined as features which are used to represent different body poses. In [3], the piecewise-linear model is used to classify different motion types. Moreover, the index structure is constructed based on the linear models. In [4] and [5], the features of the motion poses will be extracted for each motion frame which map to a multi-dimension vector. The motion content is represented by those vectors while the index structure is also constructed. In [6] and [7], the motion poses are represented by a hierarchical structure and the key frames are extracted to find the motion information. However, the extraction of key frames is time consuming. In [8–10], various motion database representation mechanisms are proposed.

In [11], a method to find the similar motions is proposed. Based on the index structure, the partial queries, i.e., the motions of some specific body parts, can be processed. Five index structures are constructed for the motions performed by different body parts. Moreover, a hierarchical structure of the human body is used to integrate the index structures such that the partial motion queries can be processed. A new similarity measurement is proposed in [12] to enhance the efficiency and effective of similar matching mechanism.

Since the motion data can be represented as multiple attribute (multiple reference points) time series data, many researches focus on the similarity search of multiple attributes databases. In [13], the Dynamic Time Warping (DTW) and Longest Common Subsequence (LCSS) approaches are extended as the similarity measurement of the multiple attribute data. In [14], a pivot-based index structure for combination of feature vectors is proposed. The query processing problem is transformed to a searching problem of multiple attributes data set. In [15], an aggregate nearest-neighbors retrieval algorithm is proposed for multi-points query problem. iDistance [16] is an index structure based on distance. In [17], a pre-processing procedure is proposed to construct the neighbor graph for motion database such that the nearest neighbor search can be efficiently performed. However, the preprocessing procedure is time consuming.

In our previous works [18–21], the index structures and corresponding query processing mechanisms for multiple attribute time sequence are proposed for exact or approximate matching problem.

This paper proposes multi-dimensional multi-attribute index structure for 3D motion data, and the corresponding query processing mechanism is proposed to efficiently find the matched motions.

2 Multi-dimensional Multi-attribute Sequence Model for Motion Data

In this section, the data model of 3D motion is proposed to represents the content of 3D motions in the database. Since the motion data can be represented by the movement of multiple reference points. The movement of each reference point can be represented by the movement in 3 coordinates. That is, the movement of each reference point can be represent as a 3 attribute sequence. Moreover, each reference point can be considered as a dimension of the motion data. Therefore, the motion data can be represented by a multi-dimensional 3-attribute sequence.

There are two different data models which are proposed to represent the semantic and geometric meanings of 3D motions. The semantic model is constructed based on the content information of the 3D motions while the geometric data model is constructed based on the coordinates of the reference points in 3D spaces.

To construct the semantic model, the hierarchy of human body will be defined first. In the proposed semantic model, the semantic meanings of human body are defined as the following different types: Whole Body, Half Body, Hand and Foot. The type of motions corresponding to each type of human body will be further defined as: Stand, Squat, Walk, Run, Jump, Wave and Kick. Each motion in the database will be analyzed and hierarchically classified into some particular semantic class such as Hand → Wave. Therefore, each motion in the database will be structurally annotated based on the semantic data model. The annotated metadata can be used as the query criteria specified by users.

The geometric data model is constructed based on the motion properties of the reference points in the 3D spaces. The movement of each reference point can be represented by the feature values of 3 dimensions. That is, the feature values are used to represent the movement characteristic of three different dimensions. The feature values used to represent the motion data are shown as follows.

X-dimension	Left	Steady	Right
Y-dimension	Up	Steady	Down
Z-dimension	Forward	Steady	Backward

For example, the movement of the reference point corresponding to left wrist in Fig. 1 can be represented as the following 3-attribute string.

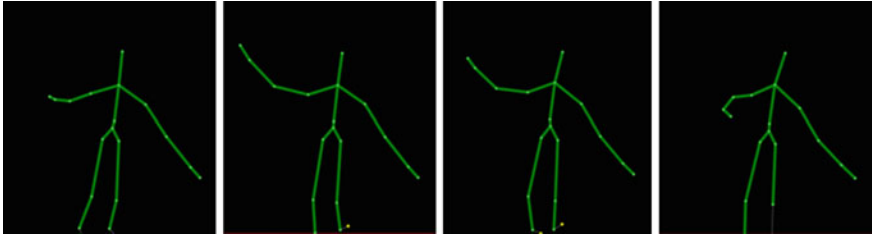


Fig. 1 Example of a 3D motion

L	R	R
U	D	D
F	F	B

Therefore, the 3D motions can be represented as a set of 3-attribute strings which can be used to construct the index structure for further query processing mechanism.

3 Index Structure and Query Processing

Since the 3D motion data is represented as a multiple 3-attribute strings, the suffix tree based index structure proposed in our previous work [18] can be modified and applied to efficiently find the matched motion data. On the other hand, the semantic hierarchy can be used to record the semantic meaning of each motion which can be used to find the matched results by traverse the hierarchy. The architecture of the index structure construction can be shown in Fig. 2.

A kinect based query interface is also designed for user to specify queries. The feature values of the query motion will be extracted and represented as a query string. The query processing mechanism will be applied to process the query string and the match results will be efficiently found via the index structure. The architecture of the proposed query processing mechanism is shown in Fig. 3.

The semantic query can be specified by users via the interface to find the corresponding motions. The geometric queries will be specified by using the Kinect device. The query motion can be acted by user and the Kinect device will capture the user motion and transform into the multiple 3-attribute strings as the query string. The multiple 3-attribute strings will be decomposed and traverse the corresponding index tree to find the matched data. At last, the matched data will be compared and the matched results will be found. Since only the movement of each reference point is considered, the absolute location of the human body captured by

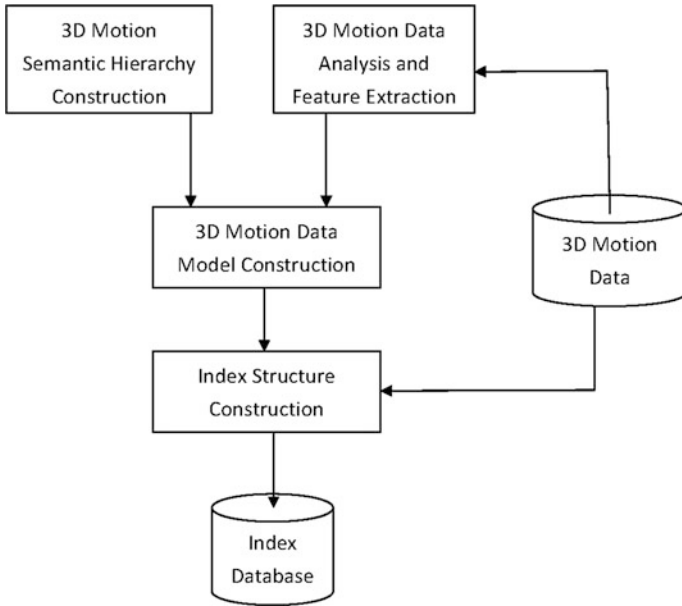


Fig. 2 Index structure construction

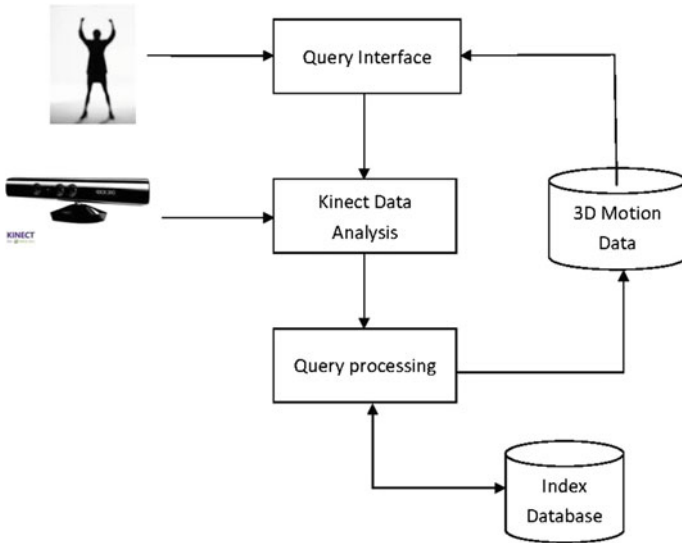


Fig. 3 The query processing mechanism

Kinect device won't affect the matching results. The experiment results show the flexibility of specifying queries.

4 Experimental Results

In this paper, a 3D data management query system is proposed. The desired motions can be specified semantically or geometrically by users. A semi-automatically annotation system is developed to annotate the semantic metadata of each 3D motion recorded in the database. Moreover, the multiple 3-attribute strings used to represent each 3D motion are automatically generated by considering the relationship of the coordinates between reference points. Then, the multiple index structures corresponding to multiple reference points are constructed according the corresponding 3-attribute string. The query interface is designed for user to specified desired motions as queries. The semantic queries can be manually specified via the interface. Moreover, the geometric queries can be captured via the Kinect device and the corresponding multiple 3-attribute strings will be transformed and specified as query strings. Therefore, the index structures are traversed and the matched results can be found.

The experiments show that the desired motions can be easily specified by users and the matched results can efficiently found.

5 Conclusion

In this paper, a 3D motion data management query system based on Kinect is proposed. The semantic meaning of 3D motions will be hierarchically annotated. Moreover, the geometric meanings of the 3D motions will be represented as multiple 3-attribute strings and the corresponding index structure are proposed to efficiently find the matched motions. Moreover, the Kinect device is used to capture the user queries and represented as a query string for further query processing mechanism. The experimental results show that the desired motions can be easily specified and the matched results can be found efficiently. Since the user queries may not exactly describe the desired motions, the approximate results should be further considered. We are currently working on extending the proposed methodology to find the approximate results. The similarity measurement and the corresponding matching algorithm are currently under development.

References

1. Lu C, Ferrier NJ (2003) Automated analysis of repetitive joint motion. *IEEE Trans Inf Technol Biomed* 7(4):263–273
2. Muller M, Roder T, Clausen M (2005) Efficient content-based retrieval of motion capture data. *ACM Trans Graphic (TOG)* 24:667–685

3. Liu G, Zhang J, Wang W, McMillan L (2005) A system for analyzing and indexing human-motion databases. In: Proceedings of the 2005 ACM SIGMOD international conference on management data, pp 924–926
4. Chao S-P, Chiu C-Y, Chao J-H, Ruan Y-C, Yang S-N (2003) Motion retrieval and synthesis based on posture features indexing. In: Proceedings of the 5th international conference on computational intelligence multimedia applications, pp 266–271
5. Chiu C-Y, Chao S-P, Wu M-Y, Yang S-N, Lin H-C (2004) Contentbased retrieval for human motion data. *J Vis Commun Image Representation* 15:446–466
6. Liu F, Zhuang Y, Wu F, Pan Y (2003) 3D motion retrieval with motion index tree. *Comput Vis Image Underst* 92:265–284
7. Gu Q, Peng J, Deng Z (2009) Compression of human motion capture data using motion pattern indexing. *Comput Graph Forum* 28(1):1–12
8. Wang J, Fleet D, Hertzmann A (2008) Gaussian process dynamical models for human motion. *IEEE Trans Pattern Anal Mach Intell* 30(2):283–298
9. Tang K, Leung H, Komura T, Shum HPH (2008) Finding repetitive patterns in 3D human motion captured data. In: Proceedings of 2nd international conference ubiquitous information management communication. Suwon, Korea, pp 396–403
10. Wang X, Yu Z, Wong H (2008) 3D motion sequence retrieval based on data distribution. In: Proceedings of 2008 IEEE international conference multimedia expo. pp 1229–1232
11. Gaurav NP, Balakrishnan P (2009) Indexing 3-D human motion repositories for content-based retrieval. *IEEE Trans Inf Technol Biomed* 13(5):802
12. Kruger B, Tautges J, Weber A, Zinke A (2010) Fast local and global similarity searches in large motion capture databases. In: Proceedings of 2010 ACM SIGGRAPH/eurographics symposium on computer animation
13. Vlachos M, Hadjieleftheriou M, Gunopulos D, Keogh E (2003) Indexing multi-dimensional time-series with support for multiple distance measures. In: Proceedings SIGMOD. pp 216–225
14. Bustos B, Keim D, Schreck T (2005) A pivot-based index structure for combination of feature vectors. In: Proceedings of SAC 2005. New York, pp 1180–1184
15. Papadias D, Tao Y, Mouratidis K, Hui CK (2005) Aggregate nearest neighbor queries in spatial databases. *ACM Trans Database Syst* 30(2):529–576
16. Yu C, Ooi BC, Tan K-L, Jagadish HV (2001) Indexing the distance: an efficient method to KNN processing. In: Proceedings VLDB. San Francisco, CA, pp 421–430
17. Chai J, Hodgins JK (2005) Performance animation from low-dimensional control signals. *ACM Trans Graph* 24(3):686–696 (SIGGRAPH 2005)
18. Lin C-H, Chen ALP (2006) Indexing and matching multiple-attribute strings for efficient multimedia query processing. *IEEE Trans Multimedia* 8(2):408–411
19. Lin C-H, Chen ALP (2006) Approximate video search based on spatio-temporal information of video objects. In: The first IEEE international workshop on multimedia databases and data management
20. Lin EC-H (2013) Research on sequence query processing techniques over data streams. *Appl Mech Mater* 284–287:3507–3511
21. Lin EC-H (2013) Research on multi-attribute sequence query processing techniques over data streams. In: 2nd international conference on advanced computer science applications and technologies

An Novel Web Service Clustering Approach for Linked Social Service

Wuhui Chen, Banage T.G.S. Kumara, Takazumi Tanaka,
Incheon Paik and Zhenni Li

Abstract It is considered that Web services have had a tremendous impact on the web as a potential silver bullet for supporting a distributed service-based economy on a global scale. However, despite the outstanding progress, their uptake on a web scale has been significantly less than initially anticipated due to higher usage thresholds. For instance, it is a hard task for service provider to seek appropriate semantic information such as OWL ontologies for service annotation in the service publication stage due to the fact that nowadays we are suffering from serious lack of available and ubiquitous ontologies for global consensus. Also it is not realistic for query users who do not possess much semantic knowledge to specify their requests with associated semantic information in the service discovery stage. In this paper, we propose an approach to publish services based on Linked data principles and discover services by service cluster with visualization for reducing the using thresholds. First, we propose Linked social service which is published on the open web by following Linked data principles with social link, and then we suggest a new method to calculate service similarity with tree structure. Then, a spatial clustering algorithm is proposed to enable visualization for reducing the using thresholds. Finally, experiment is conducted to show the effectiveness of our proposed approach.

W. Chen (✉) · B.T.G.S. Kumara · T. Tanaka · I. Paik · Z. Li
The School of Computer Science and Engineering,
The University of Aizu, Aizuwakamatsu, Japan
e-mail: chenwuhui21@gmail.com

B.T.G.S. Kumara
e-mail: btgsk2000@gmail.com

T. Tanaka
e-mail: himura.eco.ttt.sw@nifty.com

I. Paik
e-mail: paikic@u-aizu.ac.jp

Z. Li
e-mail: lizhenni2012@gmail.com

1 Introduction

Web services have been considered to have a tremendous impact on the web, as a potential solution for supporting a distributed service-based economy on a global scale. However, despite outstanding progress, uptake on a Web scale has been significantly less than initially anticipated. On the one hand, the number of services available on the web is far less than the expectation. Today, Seekda.com provides a site that has one of the largest indexes of publicly available Web services, currently accounting for 28,500 Web services with their corresponding documentation. The number of publicly available services contrasts significantly with the billions of Web pages available. Interestingly it is not significantly greater than the 4000 services estimated to be deployed internally within Verizon. Other academic enquiries into crawling and indexing Web services on the Web have found far smaller numbers of services [1]. On the other hand, the handicap of service discovery and automatic service composition results in a lack of applications for using the services in the computer industry. Most services published on the web are never used; only few of services on the web have ever been discovered, composed or invoked [2–5]. The merger condition with the handicap in the service environment results in a vicious circle of creation, publication, location, and composition of services in the computer industry. From investigation in several technological perspectives of Web services, the reasons can be mainly described by the following:

First, a lack of available and ubiquitous ontologies for service annotation results in higher using threshold for service provider in service publication stage. To better support service discovery, composition and execution, Semantic web services has been proposed as a key to maximize a higher level of automation by enriching services with semantic annotation and has already shown their benefits [6]. However, up until now, the impact of Semantic web services on the open web has been minimal due to lack of available and ubiquitous ontologies for service annotation in service publication stage. In current ontology engineering field, there is still a large deficiency of uniform and ubiquitous ontologies in many application domains [7, 8]. This is due to the fact that creating ontology usually requires many engineers to cooperate with each other.

Second, it is not realistic for query users who do not possess much semantic knowledge to specify their requests with associated semantic information in the service discovery stage. Traditional approaches have not provided visualization of the clusters. They show clusters with service groups on distance base. The conceptual clusters are mainly useful for machine. But visualization helps for human manipulation of the service clusters and gives inspiration for a specific domain from visual feedback. Density variation of the services within cluster varies according the similarity of services. Most similar services are tied together. However in traditional algorithms there is not any method to get the measurement or clue to identify the density variation within cluster and cluster position relative to the other clusters on the space. Another issue of traditional algorithms is, in iterative steps these algorithms consider about the similarity of limited number of services (e.g., similarities

of cluster representatives like cluster centers of intermediate clusters). So if there are any false positive members in intermediate clusters, then it will affect to the cluster performance. Furthermore traditional clustering algorithms are failed to achieve higher noise isolation.

In order to address the aforementioned issues, we propose an approach to publish services based on Linked data principles and discover services by service cluster with visualization for reducing the using thresholds. To reduce user's usage thresholds with service consumer, we apply spatial clustering technique called the Associated Keyword Space (ASKS) [9] with projection from a 3D sphere to a 2D spherical surface for 2D visualization. To support the semantic service annotation, Linked social service is built on a web of data which is an outstanding body of knowledge (light weight ontologies and data expressed in their terms) that can help to significantly reduce the effort for creating semantic annotations for services.

The remainder of this paper is structured as follows: in Sect. 2 we propose Linked social service to connected distributed services with social link. Then in Sect. 3 we suggest a new method to calculate service similarity with tree structure based on Linked data term distance. Then, a spatial clustering algorithm is proposed to enable visualization for reducing the using thresholds in Sect. 4. And then the evaluations of effectiveness of our approach are done in Sect. 5. The final section gives the conclusion and future work.

2 Linked Social Service

Web services are nowadays mostly used within controlled environments such as large enterprises rather than on the Web. One could argue that a reason is the fact that the lack of success of UDDI which have several drawbacks, for instance, syntactic discovery returns results with low precision, Web services are treated as independent elements in UDDI, and present registries do not record the services' peer services. Research on SWS has managed to alleviate some of the technical drawbacks of existing Web services technologies by enriching them with semantic annotation. However, building ontologies is a huge and complicated task requiring most knowledge experts to cooperate with each other resulting in the shortage of consensus and ubiquitous ontologies.

We believe that the advent of the Web of Data together with social principles constitute the final necessary ingredients that will ultimately lead to a widespread adoption of services on the Web. Firstly, the evolution of the Web of Data is highlighting the fact that light weight semantics yield significant benefits that justify the investment in annotating data and deploying the necessary machinery. This initiative is contributing to generating an outstanding body of knowledge (light weight ontologies and data expressed in their terms) that can help to significantly reduce the effort for creating semantic annotations for services. Secondly, the recent evolution around Linked data has shown that linking data over the Web can lead to

large quantities of very useful data with a low cost. Rather than isolated data islands, connecting distributed structured data into a single data space can lead to reused data, discover data from relevant data and integrate data from large numbers of formerly unknown data sources [10–13]. This new scenario provides suitable technologies and data, as well as the necessary economic and social interest for the wide application of services technologies on a Web scale. Rather than isolated service islands, connecting distributed services into a single global service space can lead to more effective service discovery and composition.

Our previous work [14] proposed Linked social service to construct a global social service network based on linked data principle for better quality of service discovery and service composition. In the global social service network, services described in lightweight ontologies are interlinked to related services from different sources functionally across the Web and in turn external services may link to them functionally using social link. However, our previous work has ignored the calculation of service similarity and service cluster with visualization for reducing the using thresholds. In this paper, we focus on service discovery by service cluster with visualization for reducing the using thresholds.

3 Calculating Service Similarity with Tree Structure

Current approaches, such as ontology concept-based techniques, and information retrieval-based techniques only consider service's input/output as a simple datatype for service similarity calculation. However real-world services published on the web always have input/output parameters with complex datatype. Therefore it is significantly important to discover services considering complex datatype. In this section, we propose an algorithm considering complex datatype of service's input/output by mapping services to tree structures.

Definition 1 The Input tree and output tree is a hierarchical tree structure with a set of linked nodes $T = (Nr, Np, Nc, L)$, where

- Nr is the root node, and its value is input or output.
- Np is a parent node, and its value represents a complex datatype or abstract datatype.
- Nc is a child node, and its value represents a primary datatype or concrete datatype.
- L is a link between Np and Nc , and it represents the relationship of Np and Nc , such as a subclass of, aggregation, and generation.

Note that Nr , Np , and Nc are subjects and L is a predicate when they are described in RDF.

3.1 Mapping Service to Tree Structure

In order to improve the performance of matching services, it is important to calculate service similarity considering complex datatype. In this subsection, we explain the tree structure in service annotation for mapping service into tree structure. A service described based on a conceptual model, such as WSML, OWL-S and WSMO [15], can be considered into a tree structure and can be mapped into an abstract level like in Fig. 1. Figure 1 shows that a service represents Interface and contains some Operations. An Operation includes some Outputs and Inputs message, which is completely hierarchical tree structure. Figure 2 shows a RDF-based annotation mapping in detailed. It means that an operation contains an input tree and an output tree defined above, and an input/output tree is containing some complex datatype which is a parent node in tree structure and primary datatype which is a child node.

3.2 Service Tree Mapping Algorithm

We propose an algorithm to mapping services to tree structure which is described in the above section. The mapping algorithm is defined as Algorithm 1. The algorithm is including three parts: first, it finds *OperationList* from service annotation and

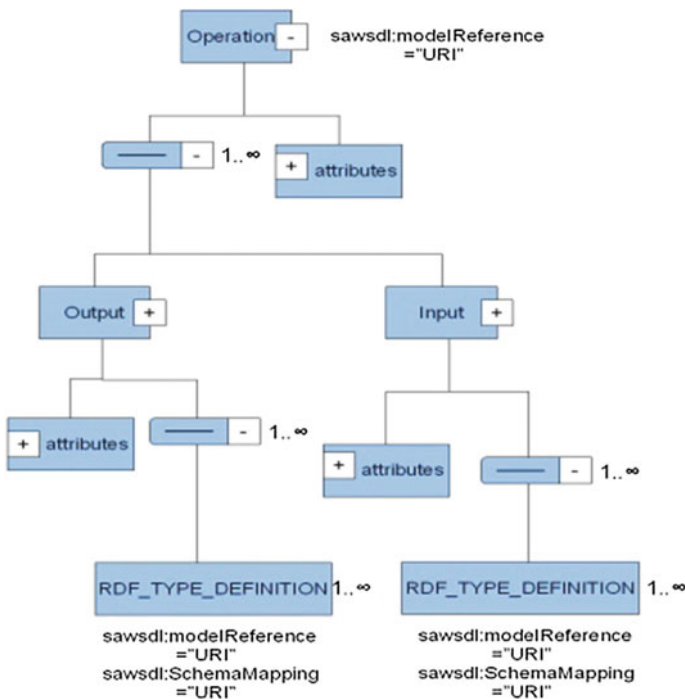


Fig. 1 Abstract level mapping

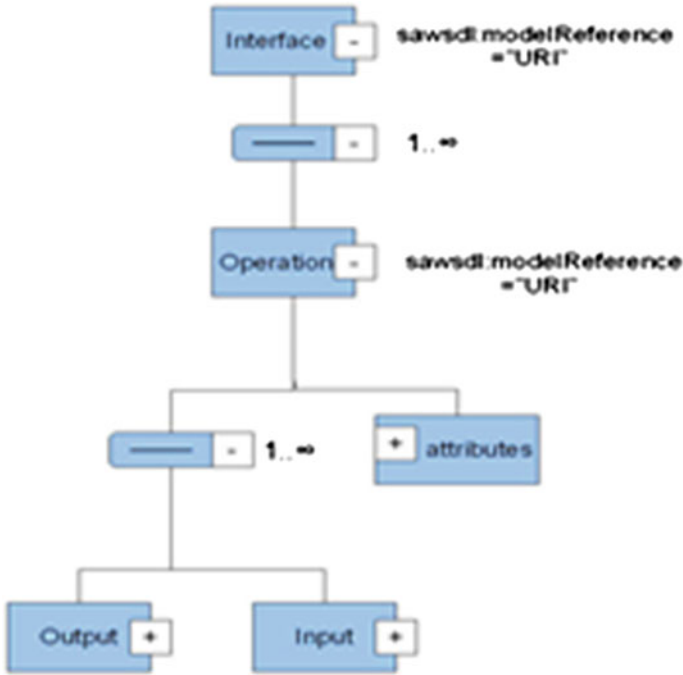


Fig. 2 RDF-based annotation mapping

creates new Operation node; second, it translates *InputList* in service annotation into Input tree and creates new Input tree (lines 7–16); third, it translates *OutputList* in service annotation into Output tree and creates new Output tree (lines 17–27).

3.3 Service Similarity Calculations Based on Tree Structure

We proposed an algorithm to mapping service to tree structure, so that complex datatype can be considered as a subtree of service tree in above section. Based on tree structure, we propose an algorithm to calculate service similarity in this section. We can calculate using the tree structure made from the previous section. Consider that we want to calculate the similarity of service 1 and service 2, service 1 and service 2 as shown in Fig. 3. We model all services participating as a tree which has been defined. As a first step, we compute initial similarity values between all pairs of resources (1, 2, 3) from resource service and (5, 6, 7) from target services. Such similarity values might be calculated by a string similarity algorithm comparing literals directly attached to these resources. In our example, this produces the results in Table 1. Next, we construct a tree similarity measure. In our example, we consider the possible tree mappings in Table 2. Then, we associate a measure with such mappings: we sum the similarity values associated with each pair (x; y) and we normalize it by the number of

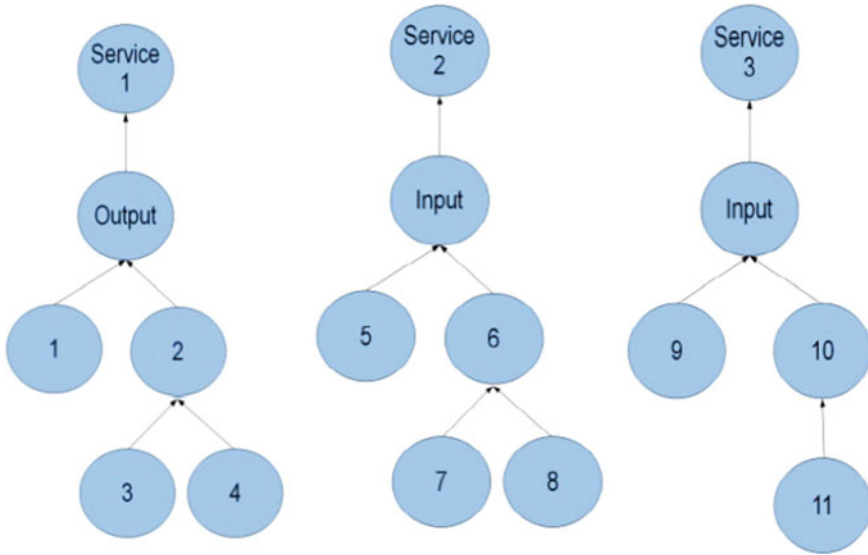


Fig. 3 Matching between input and output tree

pairs in the mapping. In our example, the resulting measures are in Table 3. Finally, we choose the mapping whose similarity measure is the highest, optionally thresholding to avoid making mappings between graphs which are highly dissimilar.

```

1: while Service.read != null do
2:   ServiceInfo = getServiceInformation
3:   createChildNode(ServiceInfo)
4:   while OperationList != null do
5:     OperationInfo = getOperationInformation
6:     createChildNode(OperationInfo)
7:     while InputList != null do
8:       InputInfo = getInputInformation
9:       createChildNode(InputInfo)
10:    loop while ParameterList != null do
11:      ParameterInfo = getParameterInformation
12:      createChildNode(ParameterInfo)
13:      if(ParameterInfo == complexType)
14:        goto loop;
15:      end while
16:    end while
17:    while OutputList != null do
18:      OutputInfo = getOutputInformation
19:      createChildNode(OutputInfo)
20:    loop while ParameterList != null do
21:      ParameterInfo = getParameterInformation
22:      createChildNode(ParameterInfo)
23:      if(ParameterInfo == complexType)
24:        goto loop;
25:      end while
26:    end while
27:  end while
28: end while
    
```

Algorithm 1. Service Tree mapping

Table 1 Legend for Fig. 3

Num	Property
1, 5, 9	Dbpedia-ont: ADRESSTYPE
2, 6, 10	Dbpedia-ont: clinicinfo
3, 7, 11	Dbpedia-ont: cliniccity
4, 8	Dbpedia-ont: clinicstate

Table 2 Initial similarity mapping

Resource 1	Resource 2	Resource 3
1	5	1
1	6	0.4
1	9	1
1	10	0.4
2	5	0.1
2	6	1
2	9	0.1
2	10	1
3	7	1
3	8	0.2
3	11	1
4	7	0.3
4	8	1
4	11	0.3

Table 3 Possible tree matching measures

Graphs		Mapping	Measures
G1	G2	MG1:G2a = {(1, 5), (2, 6), (3, 7), (4, 8)}	1
G1	G2	MG1:G2b = {(1, 5), (2, 6), (3, 8), (4, 7)}	0.6
G1	G2	MG1:G2c = {(1, 6), (2, 5), (3, 7), (4, 8)}	0.7
G1	G2	MG1:G2d = {(1, 6), (2, 5), (3, 8), (4, 7)}	0.3
G1	G3	MG1:G3a = {(1, 9), (2, 10), (3, 11)}	0.8
G1	G3	MG1:G3b = {(1, 9), (2, 10), (4, 11)}	0.6
G1	G3	MG1:G3c = {(1, 10), (2, 9), (3, 11)}	0.55
G1	G3	MG1:G3d = {(1, 10), (2, 9), (3, 11)}	0.3

4 Spatial Clustering

Current clustering approaches use traditional clustering algorithms such as agglomerative [16] and k-means as the clustering algorithms. Traditional approaches have not provided visualization of the clusters. They show clusters with service groups on distance base. The conceptual clusters are mainly useful for machine. But visualization helps for human’s manipulation of the service clusters and gives inspiration for a specific domain from visual feedback. In this research we

apply spatial clustering technique called Spherical Associated Keyword Space (SASKS) which we proposed in our previous work [9]. SASKS algorithm is modified version of Associated Keyword Space (ASKS). ASKS is an extended multidimensional scaling algorithm and able to represent services in 3-D space by using the service’s similarity. Another advantage of ASKS is that, it can achieve higher noise isolation. This result to increase the precision of service clusters.

1. Distance Measure of SASKS

Let k denote the dimension of the space in which services are located. Distance between two services is given by D_{ij} ;

$$D_{ij} = -f(x_j^{(k)} - x_i^{(k)}) \tag{1}$$

where x_i and x_j are locations of the service i and j respectively. f has a parameter a and is defined using (2).

$$f(x) = \begin{cases} |x|^2, & |x| < a \\ 2a|x| - a^2, & |x| \geq a \end{cases} \tag{2}$$

where parameter a is denscontrol peter. Clustering efficiency and the calculation load are both strongly influenced by the parameter a .

2. Iterative solution of nonlinear optimization

The criterion function of SASKS is given by (3).

$$J(x_1, x_2, \dots, x_n) = \sum_i \sum_j \left\{ -M_{ij} f(x_j^{(k)} - x_i^{(k)}) \right\} \rightarrow \max \tag{3}$$

where M_{ij} is the affinity value between services i and j . Here we use similarity score between services as the affinity value. The partial derivative of J with respect to x_i provides the formula for determining the values of x_i that maximize J :

The following (1) iterative computation converges to the solution $x_i : ij = 1, 2, \dots, n, k = 1, 2, 3, \dots, p$ and $t = 1, 2, \dots$

$$x_i^{(k)}(t+1) = \frac{\sum_{j=1}^n M_{ij} \left\{ D(x_j^{(k)}(t) - x_i^{(k)}(t)) (x_j^{(k)}(t)) \right\}}{\sum_{j=1}^n M_{ij} D(x_j^{(k)}(t) - x_i^{(k)}(t))} \tag{4}$$

ASKS plots services to a 3D sphere. However, this 3D form is difficult to visualize on a 2D screen. We therefore apply the SASKS by modifying the uni-formalization part of ASKS for our clustering approach. The SASKS technique plots services onto a 2D spherical surface for easy visualization on a 2D screen. The affinity calculation part of SASKS is the same as for ASKS. In SASKS, after calculating the service position by using affinity calculations, a KL transform is

used to fit the origin and distribution of service positions. The service position is then fitted to a spherical surface using a diagonal from the center. At this stage, the service distribution is a temporary fit to the spherical surface, and it may involve deviations. After several iterations of recalculating via the KL transform and fitting to the spherical surface, SASKS can achieve a stable distribution.

5 Implementation and Evaluation

5.1 Service Similarity with Tree Structure

This experiment was conducted using a set of services that consisted of complex datatypes and simple datatypes. The experiments have been conducted on intel(R) Core2 CPU, 2.4 GHz, and 4 GB RAM. We prepared a set of services that had a complex datatype as shown in Fig. 4. Our experiment result is shown in Table 4. According to Table 4, DT1 is similar to DT4 by considering simple datatype.

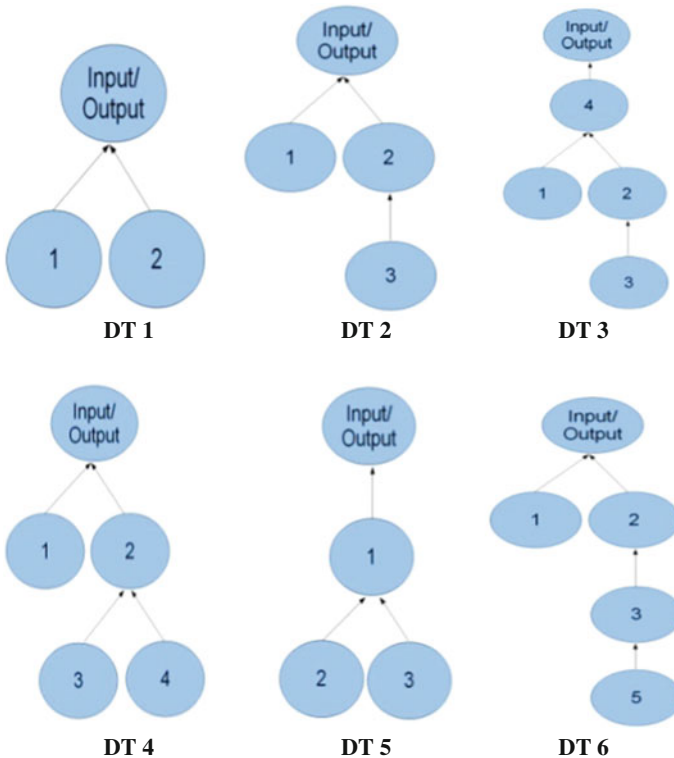


Fig. 4 Data tree (*DT*) of complex datatype. DT 1, DT 2, DT 3, DT 4, DT 5, DT 6

Table 4 Experiment result

DT compared	Mapping tree structure	Simple datatype
DT1/DT1	1	1
DT1/DT2	0.66	1
DT1/DT3	0.5	0
DT1/DT4	0.5	1
DT1/DT5	0.33	0.5
DT1/DT6	0.5	1
DT2/DT2	1	1
DT2/DT3	0.75	0
DT2/DT4	0.75	1
DT2/DT5	0.33	0.5
DT2/DT6	0.75	1
DT3/DT3	1	1
DT3/DT4	0.75	0
DT3/DT5	0.25	0
DT3/DT6	0.75	0
DT4/DT4	1	1
DT4/DT5	0.25	0.5
DT4/DT6	0.75	1
DT5/DT5	1	1
DT5/DT6	0.25	0.5
DT6/DT6	1	1

But DT1 is not similar to DT4 by considering complex datatype. In this instance, the latter is correct. In DT1, parameter 1 and 2 are in the first hierarchy. In DT4, parameter 1 and 2 are in hierarchy of first. But, parameter 3 and 4 is in the second hierarchy. Since a parameter exists in another hierarchy, the high similarity is not correct. Therefore, we can eliminate the mistake in the calculation caused by structure.

DT1 and DT2, DT2 and DT6 have high similarity. DT1 and DT2, DT2 and DT6 have common parameters in the first hierarchy. When a parameter in the hierarchy is the same, it turns out that the similarity becomes higher.

5.2 Clustering Visualization

This experiment was conducted on Microsoft Windows 7, Intel core i7-3770, 3.40 GHz and 4 GB RAM. ASKS algorithm was implemented using MATLAB. WSDL documents were gathered from the real-world Web service providers and Web service repositories. We performed manual classification in order to categorize the Web service data set to compare the results. Book, Medical, Food, Film and Vehicle were the identified categorizes. As we mentioned, cluster efficiency

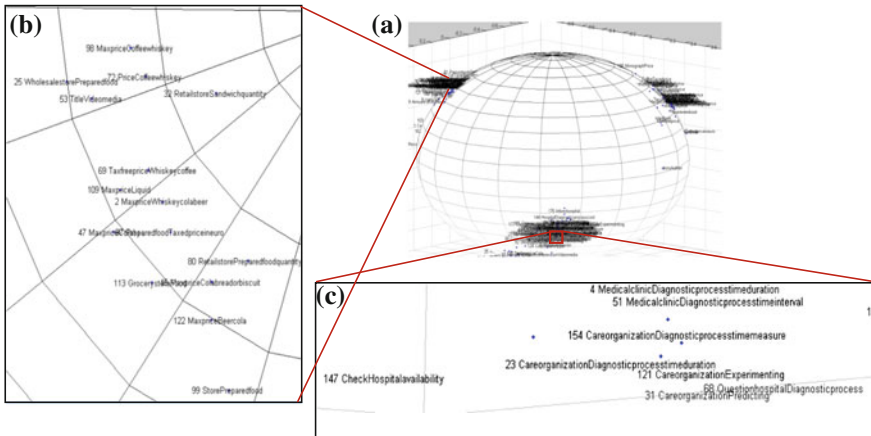


Fig. 5 Result of spatial clustering and visualization. **a** Clustering surface, **b** part of food cluster, **c** part of medical cluster

strongly influenced by the density control parameter a . We have done the experiments with $a = 0.2$ and with 100 iterations. Figure 5 shows result of spatial clustering approach. On the spherical surface the services are distributed according to their similarity. When analyzing the spherical surface, we observed five main regions where services are placed and we show that similar services in same domain are placed into one region. We observed clear separation of regions and density variation of services within the region that can be considered these regions as service clusters. Highlighted areas in Fig. 5a show some similar services. Figure 5b, c show parts of Food and Medical clusters respectively. Highlighted area in Fig. 5c shows that more similar services are placed in same area within the cluster. For example *CareorganizationExperimenting* service is placed inside the cluster more closely to the *careorganizationDiagnosticprocesstimeduration* and *careorganizationpredicting* service than *checkHospitalavailability* service.

6 Conclusion

In order to reduce the using thresholds for both service provider and service consumer, we proposed an approach to publish services based on Linked data principles and discover services by service cluster with visualization for reducing the using thresholds. Linked social service was proposed to publish service on the open web by following Linked data principles with social link, and then a new method was proposed to calculate service similarity with tree structure based on Linked data term distance. Then, a spatial clustering algorithm was proposed to enable visualization for reducing the using thresholds. Finally, experiment was conducted to show the effectiveness of our proposed approach. In the future work, by connecting

isolated service islands or services repositories into service social network, we expect that our approach can make service requirement for service-based economy at global scale clear so that our approach can impel service providers to publish their services on the web as a piece of service social network and motivate service consumer to use services.

References

1. Petrie C (2009) Practical web services. *IEEE Internet Comput* 13(6):94–96
2. Jiang W, Lee D, Hu S (2012) Large-scale longitudinal analysis of SOAP-based and RESTful web services. In: *Proceedings 19th IEEE international web service conference*. pp 218–225
3. Erl T (2007) *SOA principles of service design*. The prentice hall service-oriented computing series. Prentice Hall, Upper Saddle River
4. Hadley M (2009) Web application description language. Member submission, W3C
5. Bellwood T, Capell S, Clement L, Colgrave J, Dovey JM, Daniel F, Hately A, Kochman R (2004) UDDI. Technical report, OASIS
6. Pilioura T, Tsalgatidou A (2009) Unified publication and discovery of semantic web services. *ACM Trans Web* 3:1–44
7. Brogi A, Corfini S, Popescu R (2008) Semantics-based composition-oriented discovery of web services. *ACM Trans Internet Technol* 8(4):1–39
8. Fujii K, Suda T (2009) Semantics-based context-aware dynamic service composition. *ACM Trans Auton Adapt Syst* 4(2):1–31
9. Yaguchi Y, Oka R (2012) Spherical visualization of image data with clustering. In: *2012 4th international conference on awareness science and technology (iCAST)*. IEEE, pp 200–206
10. Bizer C, Heath T, Lee TB (2009) Linked data—the story so far. *J Semant Web Inf* 5(3):1–22
11. Bizer C, Heath T, Lee TB (2008) Linked data: principles and state of the art. In: *17th international world wide web conference*
12. Pedrinaci C, Domingue J (2010) Toward the next wave of services: linked services for the web of data. *J Univ Comput Sci* 16(13):1694–1719
13. Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z (2008) DBpedia: a nucleus for a web of open data. In: *Proceedings 6th IEEE international semantic web conference*. pp 722–735
14. Chen W, Paik I (2012) Improving efficiency of service discovery using linked data-based service publication. *Inf Syst Front*. doi:[10.1007/s10796-012-9381-x](https://doi.org/10.1007/s10796-012-9381-x)
15. Srinivasan N, Paolucci M, Sycara K (2004) Adding OWL-S to UDDI: implementation and throughput. In: *Proceedings 1st international semantic web services and web process composition conference*
16. Kumara BTGS, Paik I, Chen W (2013) Web-service clustering with a hybrid of ontology learning and information-retrieval-based term similarity. In: *Proceedings international conference on web services*

Cloud Computing Adoption Decision Modelling for SMEs: From the PAPRIKA Perspective

Salim Alismaili, Mengxiang Li and Jun Shen

Abstract The popularity of cloud computing has been growing among enterprises since its inception. It is an emerging technology which promises competitive advantages, significant cost savings, enhanced business processes and services, and various other benefits. The aim of this paper is to propose a decision modelling using Potentially All Pairwise RanKings of all possible Alternatives (PAPRIKA) for the factors that have impact in SMEs cloud computing adoption process.

Keywords Potentially all pairwise RanKings of all possible alternatives (PAPRIKA) · Cloud services · Small and medium enterprises (SMEs)

1 Introduction

Cloud computing is an emerging technology which introduced various services and resources across the network. With all the claimed benefits for organisations; cloud computing services possess significant technical, economic, ethical, legal, and managerial issues [11]. Existing studies investigated more on the technical aspects of cloud computing, with limited focus on issues related to business perspective about the adoption of cloud computing [11]. Moreover, there is shortage of detailed studies on decision support systems and cloud computing adoption process from business view [37]. This paper identifies the relevant themes affecting SMEs (Small and Medium Enterprises) adopting cloud computing in Australia. These themes

S. Alismaili (✉) · M. Li · J. Shen
School of Computing and Information Technology, University of Wollongong,
Wollongong, Australia
e-mail: szaai787@uowmail.edu.au

M. Li
e-mail: mli@uow.edu.au

J. Shen
e-mail: jshen@uow.edu.au

provide a glance for this research and will be the base in which the main topic is investigated, analyzed, discussed, and the researcher view is presented.

The objective of this paper is to propose and produce an initial decision model which is intended to assist its potential users (SMEs decision makers) in their prioritizing and selection process for cloud services. The constructs used in this model were derived from a review of relevant literature. This step will be followed by an exploratory qualitative phase then with an in-depth quantitative study in later stages. The exploratory study is being undertaken to empirically unfold the influential factors of the adoption of cloud computing from SMEs perspectives.

2 Related Work

2.1 Background

Decision making in adopting of any technology can be a difficult process even with its promises of various advantages and enhancement of business processes [4]. Cloud computing paradigm can have similar complications. In recent years there has been a demand for more holistic examination of the adoption of ICT (Information and Communication Technology). This is because such an approach can combine more than one theoretical framework in order to understand the phenomenon from different perspectives [25, 36]. The decision making situation is complex in this regard. A more comprehensive understanding of a decision case from different angles creates a better and more accurate final decision and therefore gives more positive benefits, outcomes, and from which results can be obtained.

With all the claimed benefits for organizations, cloud computing has significant technical, economic, ethical, legal, and managerial issues [21, 34]. Existing studies investigated more on the technical aspects of cloud computing, with limited focus on issues related to business perspective about the adoption of cloud computing [37]. Moreover, there is shortage of detailed studies on decision support systems and cloud computing adoption process from a business viewpoint [33, 37]. It is very important for the planning, assessment, and evaluation of cloud computing adoption decision to be done systematically taking into consideration the needs of the firm [18].

2.2 Decision Support Systems

There is an extent of variation in focus of the existing studies for cloud selection models. Han et al. [14] proposed automated system for cloud selection based on tangible and easy measurable parameters such as Quality of Service (QoS) and Virtual Machine (VM) performance based on SaaS category. The study did not take into consideration other relevant variables in the context. An alternative approach

was used by Li et al. [20] proposed an evaluation tool based on IaaS and PaaS services such as storage, network, and processing performance as selection criteria for different cloud computing services providers. Multi-Criteria Decision Making (MCDM) techniques have been considered by other researchers like Godse and Mulik [13] using Analytical Hierarchy Process (AHP). It provided a wider dimension for studying various subjective criteria but was limited to analyze SaaS services. Hussain and Hussain [17] further advanced and developed a general and complex model which was less practical especially if it is intended to be used by SMEs with limited technical capabilities. Under Multi-Criteria Decision Analysis (MCDA), there are different preference presentations and scoring methods, all of which have their own benefits and drawbacks.

Deciding about the most appropriate cloud computing deployment model and selecting suitable cloud services for businesses is not an easy task. This is because there are many technological solutions provided by cloud computing services providers and also various direct and indirect factors that influence this decision and need to be considered carefully for efficient judgment. There are various approaches of ranking, prioritizing, and weighting selections that can be provided as tools for decision maker in their selection process for their right alternatives of services which will be discussed in the coming sections. In this research a MCDA framework is implemented by combining 1000Minds software [26] and the “Potentially All Pairwise RanKings of all possible Alternatives” (PAPRIKA) scoring method [15], to determine the factors that influence the adoption of cloud computing decision to make trade-offs between different alternatives to design a model which will help decision makers in making complex decisions. PAPRIKA is a method that uses a concept of multi-MCDM or conjoint analysis for establishing decision-makers’ preferences through using pairwise rankings of alternatives [15].

The proposed model for this research was originated from a methodology that attempted to address the limitations in the previous studies. It will contribute in modelling decision making for both prioritizing and selection process in order to help enterprises make their optimal and efficient decision of the right cloud computing services that is most suitable to their business objectives. PAPRIKA arguably was selected as it closer to human logic of choice, simple, and at the same time have the complexity feature of analyzing different criteria and attributes including qualitative and quantitative data types. Moreover, PAPRIKA provides more preference comparison than most other scoring methods [15], such as direct rating [35], Simple Multi-Attribute Rating Technique (SMART) [9], Simple Multi-Attribute Rating Technique Extended to Ranking (SMARTER) [10], and the Analytical Hierarchy Process (AHP) [31]. It is implemented from 1000Minds software (www.1000Minds.com) [26]. This mechanism compares two criteria at a time which offers more accurate results in opposing to other pairwise comparison systems. This method is a useful tool for subjective and incomplete information and therefore it has the ability to produce practical solutions for real world use. The method involves prioritizing ranking of competing alternatives through evaluating all possible undominated pairs of attributes, presenting the final results in a beneficial model [15]. This will assist organizations in their decision making process.

2.3 Rationality of Using PAPRIKA Method

In PAPRIKA method each choice requires a decision-maker to trade-off one characteristics for the other (Fig. 1). Decision-makers express a preference by choosing between two things. The software automatically changes the order of the trade-off questions for each survey. This strategy of swapping the order of questions helps in reducing or eliminating the potential order biases [6, 19, 27].

On of powerful features of PAPRIKA is it ability in surveying any number of criteria and levels; as these numbers increase, the number of potential alternatives (combinations) increases exponentially. For example, six criteria and four levels creates 4096 possible alternatives [15]. The PAPRIKA method largely reduces the number of selection the decision-maker have to make by reducing ‘dominant’ pairwise comparisons and use the transitivity feature to implicitly respond to other questions. Domination occurs when a decision is not required for certain alternatives due to the high rate of some alternatives in comparison with others. Then, the ‘undominated’ pairs are to be analyzed by the software. The ‘undominated’ pair occurs when one alternative has at least one criterion with higher rate and a least one criterion with lower rate in comparison with other alternatives. The software eliminates all the redundant choices when comparing two ‘undominated’ pairs via transitivity. For example, if choice A is ranked higher than choice B and choice B is higher than choice C, then by transitivity, choice A is ranked higher than choice C. After the two choices, the third choice becomes redundant. Then the software progress in selecting another choice and the process continues until all ‘undominated’ pairs processed and ranked.

PAPRIKA method and the software have been used by researchers in different disciplines such as health-care, management, agriculture, and commerce to study various phenomena (e.g. [5, 22, 24]). This research will use PAPRIKA scoring

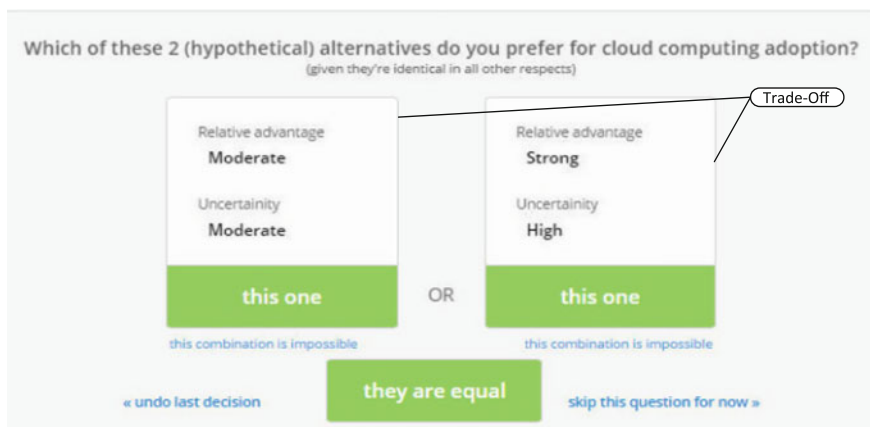


Fig. 1 Example of a pairwise-ranking trade-off question for scoring the value model presented in graphical user interface

method through its running environment 1000Minds software and not other methods for the following reasons:: (1) User friendly (2) Less complex as pairwise comparison is defined on two criteria (3) Less complex as pairwise comparison is defined on two criteria (4) Generates individual weights for every decision-maker which can be easily combined (5) Decision survey designed is clear, direct, and cost-effective (6) The survey format is robust, clear, and easy to follow.

There are numbers of decision analysis software, for more details you can visit this website link www.orms-today.org/surveys/das/das.html. 1000Minds is the only software that supports PAPRIKA method [26]. Sullivan [32] presented in his study the comparison between different scoring methods used in making decision process more easier such as SMART/SWING (Simple Multi-attribute Rating Technique), DCEs (Discrete Choice Experiments), CA (Conjoint Analysis), ACA (Adaptive conjoint analysis), AHP (The Analytic Hierarchy Process), PAPRIKA (Potentially all pairwise rankings of all possible alternatives), and Outranking methods. All these methods are based on the simple additive model except the outranking method. MCDA methods are suitable for formulating decision maker's preferences than non-compensatory methods Baltussen and Niessen [1]. Outranking models can compensate the high performance on some criteria for poor performance on others with no consideration of the resulted differences [7]. Simplicity, predictive power, and preferences evaluation capabilities are elements that determine the effectiveness of the method [16].

PAPRIKA method uses only two criteria selection, whereas SWING/SMART, outranking, and some CA methods use ranking, direct rating, weighting to rank alternatives. In these methods, scoring the criteria is based on individuals, experts, and public opinion. Rating the criteria and alternatives by decision makers can introduce confusion in data interpretation. This is becoming obvious of the different interpretation of the rating scale by different people in a specific research focused group. Hence, Forman and Selly [12] stated that the scoring of alternatives is depending on decision maker's opinion and understanding of the scoring scale.

From the other hands, the method that provides selection system between two alternatives at a time is less complicated, less interpretation errors, and demanding less knowledge and task in ranking or scoring alternatives. The choice-based methods between two alternatives have advantage over selecting from the methods that use scale; it is more fitting to human experience situation [8]. In non-trade-off choice mechanisms; there is a possibility of equal ranking or scoring occurrence. Choice modelling, permits decision-makers to establish trade-offs between criteria.

The AHP method presents the decision-makers with framework of making pairwise comparisons at each hierarchal level for the presented criteria or alternatives. It has been argued that selecting preference based on methods other than cardinal form generates consistence and reliable results [23]. Sullivan [32] discussed in his study about three methods that elicit preference information in ordinal form namely: PAPRIKA, ACA, and DCE/CA. In ACA and DCE/CA methods, however, usually two or more choice sets are presented which can include more than two criteria for each choice set [29]. The more the number of criteria, the more complex the choice becomes. Additionally, focusing on some criteria and eliminating the

other for the purpose of simplification can lead to inaccuracy in estimating criteria weights [3]. On the other hands, PAPRIKA method offers larger number of choices for decision-makers for a value model in comparing with other methods [15]. For example, DCE/CA offers smaller number of choice sets in corresponding with the number of scenarios presented [28]. The smaller number of choice sets presented by this method can be good in terms of reducing the effort that takes decision-makers for attempting to the preferences; however it can cause unreliability issues in the results. ACA method also present limited scenarios to the decision-makers which can make the preferences process of various choice sets inefficient.

The criteria weight describes the relative significance of the criteria and the intention of the decision-maker(s) (represented as individual or as a sub-group or as a complete sample) to trade off one criterion for another substitute. AHP and PAPRIKA are unique methods that produce individual criteria weights for every single decision-maker. In other methods such as SWING/SMART and outranking decision-makers determine the weight points directly to criteria. DCE/CA and ACA generates a group of weights for the whole sample. PAPRIKA method can compare criteria weights of one decision-maker with another in the trading-off the same criteria basis. However, AHP method can do the same only if decision-makers have used the same attributes and/or levels [2]. The aggregation of weight in this method depends on setup agreed by decision-makers, if it is to combine their judgement, then a geometric mean is used. Additionally, 'experts' can combine their results and geometric mean is also used and it is further can be used to rank the 'experts' themselves [30].

Selection of cloud computing providers, services, and deployment models is not an easy process for organizations. Various factors need to be considered as the decision can have a significant impact in the business. There are different approaches for rating, ranking, prioritizing, and selection of cloud computing services and its providers. One of the approaches is by using MCDA which can help decision-makers in choosing the most appropriate cloud computing deployment model and selecting suitable cloud services for their businesses. Under the category of MCDA there are various scoring and preference elicitation methods, each have its own benefits and drawbacks. In this research PAPRIKA method which is supported by 1000Minds software will be used to understand SMEs willingness in trading-off the different factors that influence them in adoption of cloud computing services.

3 Modelling the Cloud Adoption Process

3.1 Model Design

The development of a decision model for cloud adoption decision-making process was implemented based on two methods: (1) Literature review and (2) 15 semi-structured interview including 4 cloud computing services providers, 4 SMEs cloud computing adopters, 4 prospectors, and 3 not intend to adopt cloud

Table 1 Conceptual components of the proposed model

Variable	Definitions from cloud computing perspective
Relative advantage	The extent to which cloud computing is perceived as being better than the idea of other computing paradigm it supersedes
Complexity	The degree to which cloud computing is perceived as being relatively difficult to understand and use
Compatibility	The degree to which cloud computing is perceived as consistent with the existing values, past experience, and needs of potential users
Uncertainty	The degree to which cloud computing is perceived as more secure than other computing paradigms
Security concern	The perceived security and privacy concerns of cloud computing due to the occurrence of data loss
Cost savings	The extent of users perceived total cost of using cloud computing services
Privacy risk due to geo-restriction	The extent of privacy risk due to geo-restriction of cloud computing
Adoption decision	Investigated status of cloud computing services adoption decision

computing. The purpose is to identify the relevant influential factors in the adoption of cloud computing. The outcome of the interview shall confirm a more solid framework of these factors. The framework will be used to form the building components of the decision model. The initial conceptual variables are illustrated in Table 1. The final conceptual framework will be developed from the outcomes of first phase of semi-structured interviews with the 15 organizations. The qualitative study will be the basis of the second quantitative study and then the decision model design and experiment will follow. This paper will present the initial model design and its simulation based on the previous studies and industrial reports.

4 Method

The PAPRIKA method uses pairwise preferences evaluation based on trade-off process through selection one of the three options: 1—pair one is better than pair two 2—pair two is better than pair one 3—both pairs are equal. The value model or the preference values are represented by the relative importance ‘weight’ of the criteria which is calculated via mathematical methods (i.e. linear programming). The relative importance of each criterion is obtained from its highest ranked category, and the total of all the highest categories in each criterion is equal to 100 %. Cost-benefits calculations are other useful measure that can be considered in alternatives scoring through Pareto analysis which provides an additional “value for money” evaluation tool for final selection of alternatives. PAPRIKA pointing system allows the use of criteria which can be either of quantitative nature (e.g. number of employees and experience) or qualitative nature (technological,

organizational, and environmental influential factors in the adoption of cloud computing). Non-categorical criteria can also be represented with different as appropriate to the case study (e.g. low rank, medium rank, and high rank).

PAPRIKA uses ‘pairwise ranking’ method for ranking of alternatives. This is in contrast with most other decision facilitator methods which use ‘scaling’ or ‘ratio’ measurements for ranking of preferences. For example, AHP is relying on a scaling method which is based on 1–9 points and valuating which of the two defined criteria are more important in this scale system. With PAPRIKA method, users are allowed to choose one alternative between just two which is easier and natural as in the human life daily decision. PAPRIKA can process any number of pairwise rankings of the hypothetical alternatives required by decision makers. Therefore, PAPRIKA method presents better confidence in decision making. Figure 2 illustrates “The Cloud Computing Decision Model Design Process”. This activity involves.

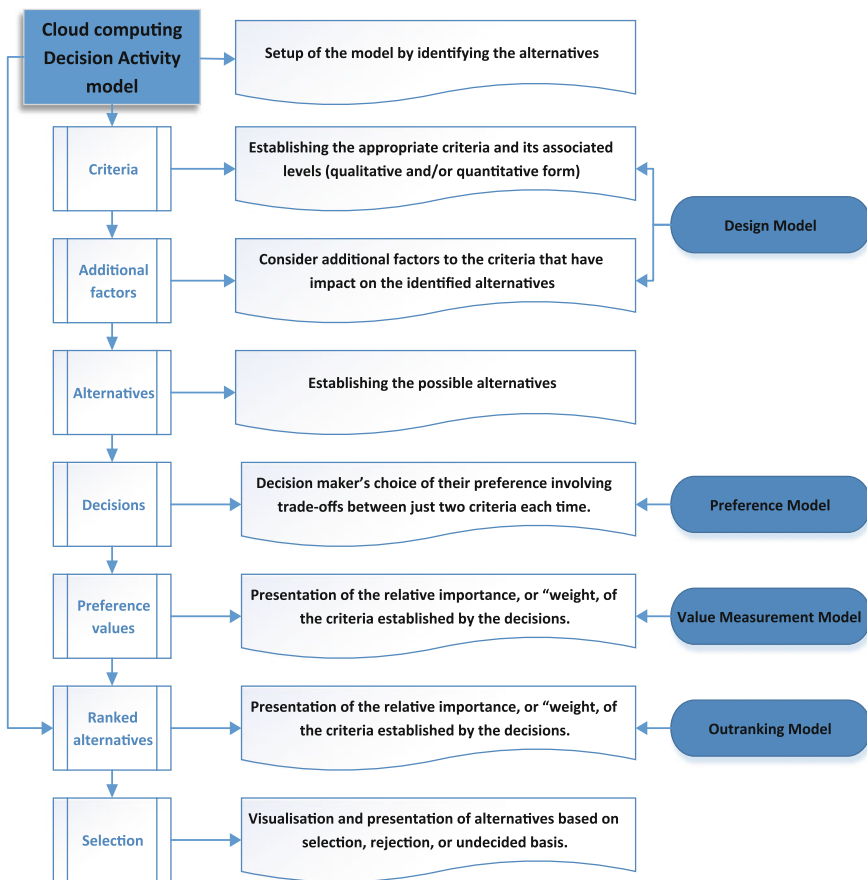


Fig. 2 The cloud computing ‘decision model’ design process

5 Simulation

At this stage the decisions are simulated using two cases. A preference survey was distributed to two theoretical emails for simulation. The survey was answered by the author taking into account two cases for conjoint analysis. This analysis estimates the expected values of the adoption decision associated with each cloud services. The first case was about “Decision to adopt cloud services with companies that have tendency towards low concern to security and privacy”. The second case was about “Decision to adopt with high concern of security and privacy preferences”. The base-case analyses ranked all the alternatives using estimates for all the best expected of the input parameters based on their intuitive judgment. The process involved elicitation the opinions of the business decision makers (imitated in this case) about the relative values of attributes within cloud services. The study also simulated a discrete choice experiment (conjoint analysis) to prioritize the cloud services criteria of the two cases by using a preference survey (answers simulated) to reveal individual’s preference values, or ‘weight’, and the average of the group. These values were then used to rank the alternatives. In summary, for simulation the study conducted two activities: 1-Ranking of alternative survey 2-Preference survey (conjoint survey).

6 Results

The study started with the creation of the initial decision model and ranked the alternatives according to literature understanding and authors intuitive knowledge (the model illustrated in Appendix 1-A). First of all, we have to mention that the ranking of alternatives of this study resulted from two simulation cases. The ranking results of the 11 alternatives are presented in Appendix 1: B and C. The report classifies the results as followings:

1. Preference Values and Criterion Rankings

Preference values represent the relative importance, or ‘weights’, of the criteria—summarized by the criterion rankings (Table 2). Each criterion’s weight corresponds to the % value for its highest level (Table 2). These values—weights—sum to 100 % (i.e. 1). For a given case 1, the value of the highest-ranked level (Table 3) for each criterion represents that criterion’s importance relative to the other criteria. The criteria weight values in Table 3 represent the importance of the criterion to the participants. For example it can be observed that ‘relative advantage’ with a value of 0.267 has the highest level of importance among other criteria. Median, mean values and rankings are the average for both cases. Standard deviation (SD) used to calculate the cases values using the ‘n’ method. Fig. 3 visualize of the criteria mean preference values.

Table 2 Relative importance of criteria (mean weights); 'marginal rate of substitution' (ratio) of the column criterion for the row criterion

	Relative advantage	Cost savings	Uncertainty	Compatibility	Security concerns	Complexity	Privacy risk due to geo-restriction
Relative advantage		1.3	1.7	1.8	2.7	4.5	4.7
Cost savings	0.8		1.3	1.4	2.1	3.5	3.7
Uncertainty	0.6	0.7		1.1	1.6	2.6	2.8
Compatibility	0.6	0.7	0.9		1.5	2.5	2.6
Security concerns	0.4	0.5	0.6	0.7		1.7	1.7
Complexity	0.2	0.3	0.4	0.4	0.6		1.1
Privacy risk due to geo-restriction	0.2	0.3	0.4	0.4	0.6	0.9	

Table 3 Normalized criterion weights and single criterion scores (means); a more traditional, though equivalent, representation of the preference values above

Criterion	Criterion weight (sum to 1)	Level	Single criterion score (0–100)
Security concerns	0.099	High	0
		Medium	51.6
		Low	100
Cost savings	0.211	Low	0
		Medium	66.7
		High	89.7
		Very high	100
Relative advantage	0.267	Weak	0
		Low	26.5
		Moderate	66.3
		Strong	100
Uncertainty	0.157	High	0
		Moderate	88.8
		Low	100
Privacy risk due to geo-restriction	0.057	High	0
		Medium	79.9
		Low	100
Compatibility	0.148	Weak	0
		Good	71.8
		Strong	100
Complexity	0.06	High	0
		Medium	78.3
		Low	100

2. Rankings of Alternatives

The spearman’s rank correlation coefficient (rs) measures the extent of similarity of 2 rankings of alternatives (Table 4), and ranges between 1 and -1. The value of 0.870 for participants indicates that there is a tendency towards an identical agreement in ranking of alternatives. Below is the spearman’s formula; where *i* = paired score.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{1}$$

Spearman’s rank correlation coefficient = 0.870 (1 = identical rankings, 0 = unrelated rankings, -1 = identical reverse rankings).

Fig. 3 Criterion value functions (mean preference values)

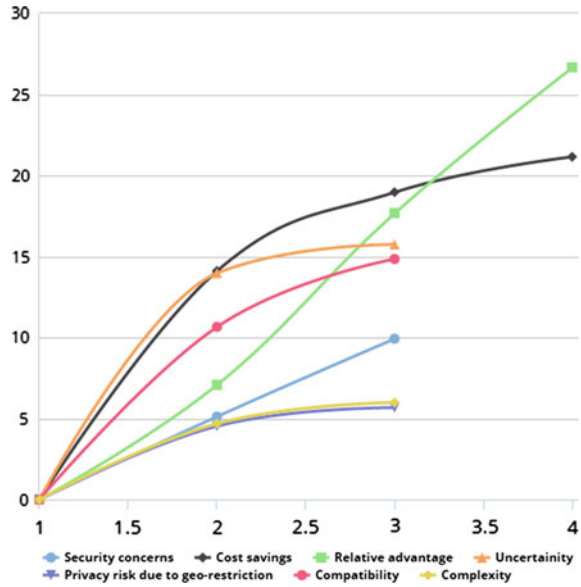


Table 4 Rankings (mid-ranks) of the 11 alternatives

Ranks		Cases		
		Simulation case 2	Simulation case 1	Mean
Alternative	Public IaaS-system	-1.5	1.5	2.5
	Private IaaS	2	-2	3
	Public IaaS-storage	-1.5	1.5	3.5
	Private PaaS	2	-2	4
	Public PaaS	-1.5	1.5	4.5
	Private SaaS	2	-2	5
	Public SaaS	-1.5	1.5	5.5
	Hybrid IaaS	0	0	8
	Hybrid PaaS	0.5	-0.5	9.5
	Hybrid SaaS	-0.5	0.5	9.5
	Status quo (not to adopt)—legacy IT	0	0	11
Spearman's rank correlation with mean ranking	0.870	0.870	1.000	

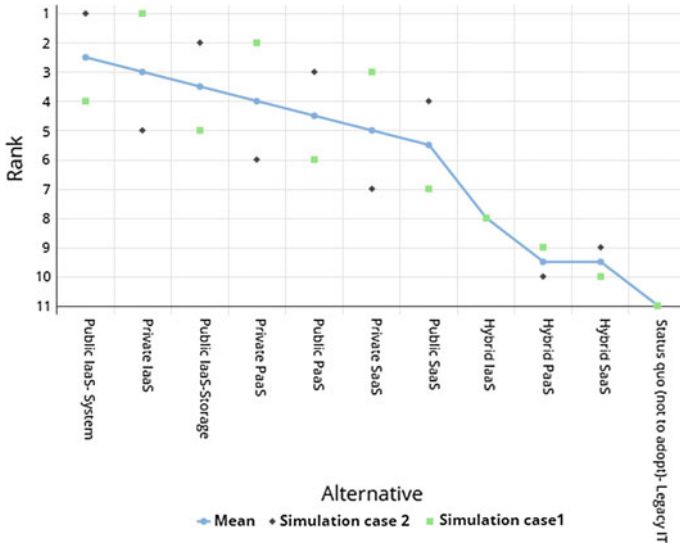


Fig. 4 Chart of 2 cases rankings of the 11 alternatives

Mid-ranks have been presented for tiered ranks (Fig. 4). The highest ranked alternates are: 1st Public IaaS-Systems; 2nd: Private IaaS; 3rd Public IaaS-Storage (Table 4). Hybrid PaaS and Hybrid SaaS have the same rank of number 9 in the list sharing the same mean value of 9.5. The model also provides a result checker tool to increase confidence in the results.

3. The Value for Money Chart

Relevant assessment of the alternative options available to SMEs in adoption of cloud computing. The results generated are relevant in understanding how and why the alternatives were ranked and also to prioritize cloud services solutions to SMEs according to their need and based on their resources. The chart (Fig. 5) represents variables in the 2 axis and additional variables represented by bubble size and bubble color. It is a useful tool for decision makers. The model further provides a budget constraint variable as an additional parameter for evaluating between alternatives. The Pareto (efficiency) frontier line in the chart identifies alternatives that ‘dominate’ all others.

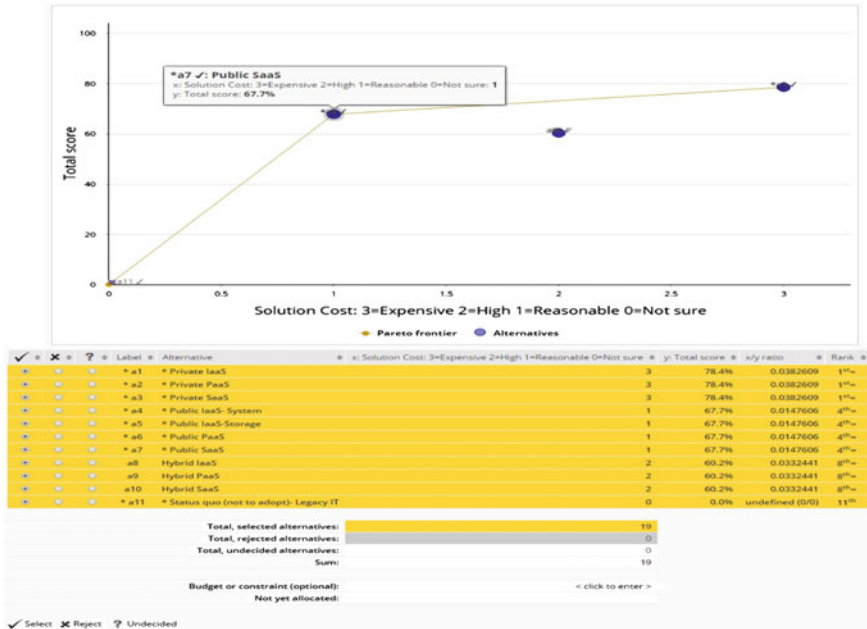


Fig. 5 Example of value for money model (simulation case1)

7 Discussion

This paper described a preference based method for ranking, prioritizing, and selection of cloud services and presented an initial decision model. This was achieved through simulation of two decision-making cases. At the 2nd phase of the study, the choice experiments will be conducted with real world cases of SMEs decision makers aiming to produce a collaborative web-based tool for businesses which will facilitate the cloud computing adoption decision process. The alternatives evaluation shows that Public IaaS-System, Private IaaS, Public IaaS, Private PaaS are the top ranked options for SMEs to adopt in cloud services. PAPRIKA method provides a useful tool for ranking the possible decisions based on decision-maker’s judgement of the importance of the criteria to their specific situation. The tool provides a consistency mechanism checker to ensure meeting the objective. PAPRIKA approach helps in reducing the complexity of the multidimensional influential variables decisions to a simple series of trade-offs choices of only two variables at a time. Making choice between two is easier and closer to human nature of judgement and selection.

8 Limitations

The presented decision model will not be static; it has to be dynamic due to the change in socio-technical nature, business environment, etc. Therefore, the model needs to be kept in continuous review and update. There are further potential in developing decision modelling in different contexts and with different environmental characteristics.

9 Conclusion

Cloud computing popularity is in continues growing among SMEs. As a result, it is very useful to understand the entire scene behind the process of cloud computing adoption. Apparently, a simple, advance, and easy to use decision making tool is useful for businesses to increase their productivity and leverage country economic. This paper proposed a new method and designed an initial cloud computing decision model based on assumptions from the authors considering simulated cases of decisions. In the next stage of the research, the proposed model will be tested experimentally with several real-world scenarios of SMEs decision makers.

It is believed that more case scenarios can help in improving the model. Finally, possible expansion of the model can be investigated to include more parameters representing the influential adoption factors at the specific time and within a specific environment. The rapid advancement of technologies requires reviewing, refining, and modifying the concepts and the parameters of the model.

Acknowledgments The authors thank 1000minds decision-making software (the software that supports PAPRIKA method) for providing us a free license and open access for the duration of the research, and Paul Hansen for his suggestions to our thinking in this area.

Appendix 1: Ranking of Alternatives

Alternative	Security concerns	Cost savings	Relative advantage	Uncertainty	Privacy risk due to geo-restriction	Compatibility	Complexity	Rank	Mid-rank	Total score (%)	Solution Cost: 3 = Expensive 2 = High 1 = Reasonable 0 = Not sure	Benefits: 3 = High 2 = Average 1 = Low 0 = No benefit	Service trust: 3 = High 2 = Average 1 = Low 0 = Not sure	Quality of Service: 3 = Very High 2 = High 1 = Average 0 = Not sure
<i>A. Model ranking</i>														
Public IaaS-system	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	79.6	1	1	1	1
Public IaaS-storage	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	79.6	1	1	1	1
Public PaaS	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	79.6	1	1	1	1
Public SaaS	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	79.6	1	1	1	1
Hybrid IaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	5th=	6	65.9	2	2	2	2
Hybrid PaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	5th=	6	65.9	2	2	2	2
Hybrid SaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	5th=	6	65.9	2	2	2	2
Private IaaS	High	Very high	Strong	Low	High	Strong	High	8th=	9	55.7	3	3	3	3
Private PaaS	High	Very high	Strong	Low	High	Strong	High	8th=	9	55.7	3	3	3	3
Private SaaS	High	Very high	Strong	Low	High	Strong	High	8th=	9	55.7	3	3	3	3
Status quo (not to adopt)-legacy IT	High	Low	Weak	High	High	Weak	High	11th	11	0	0	0	0	0
<i>B. Simulation case 1 rank</i>														
Alternative														
Private IaaS	High	Very high	Strong	Low	High	Strong	High	1st=	2	85.5	3	3	3	3
Private PaaS	High	Very high	Strong	Low	High	Strong	High	1st=	2	85.5	3	3	3	3
Private SaaS	High	Very high	Strong	Low	High	Strong	High	1st=	2	85.5	3	3	3	3
Public IaaS-system	Low	High	Moderate	High	Medium	Good	Low	4th=	5.5	57.5	1	1	1	1
Public IaaS-storage	Low	High	Moderate	High	Medium	Good	Low	4th=	5.5	57.5	1	1	1	1

(continued)

(continued)

Alternative	Security concerns	Cost savings	Relative advantage	Uncertainty	Privacy risk due to geo-restriction	Compatibility	Complexity	Rank	Mid-rank	Total score (%)	Solution Cost: 3 = Expensive 2 = High 1 = Reasonable 0 = Not sure	Benefits: 3 = High 2 = Average 1 = Low 0 = No benefit	Service inst: 3 = High 2 = Average 1 = Low 0 = Not sure	Quality of Service: 3 = Very High 2=High 1 = Average 0 = Not sure
Public PaaS	Low	High	Moderate	High	Medium	Good	Low	4th=	5.5	57.5	1	1	1	1
Public SaaS	Low	High	Moderate	High	Medium	Good	Low	4th=	5.5	57.5	1	1	1	1
Hybrid IaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	8th=	9	53.1	2	2	2	2
Hybrid PaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	8th=	9	53.1	2	2	2	2
Hybrid SaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	8th=	9	53.1	2	2	2	2
Status quo (not to adopt)- legacy IT	High	Low	Weak	High	High	Weak	High	11th	11	0	0	0	0	0
<i>C: Simulation case 2 rank</i>														
Alternative														
Public IaaS-system	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	77.9	1	1	1	1
Public IaaS-storage	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	77.9	1	1	1	1
Public PaaS	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	77.9	1	1	1	1
Public SaaS	Low	High	Moderate	High	Medium	Good	Low	1st=	2.5	77.9	1	1	1	1
Private IaaS	High	Very high	Strong	Low	High	Strong	High	5th=	6	71.3	3	3	3	3
Private PaaS	High	Very high	Strong	Low	High	Strong	High	5th=	6	71.3	3	3	3	3
Private SaaS	High	Very high	Strong	Low	High	Strong	High	5th=	6	71.3	3	3	3	3
Hybrid IaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	8th=	9	67.2	2	2	2	2
Hybrid PaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	8th=	9	67.2	2	2	2	2
Hybrid SaaS	Medium	Medium	Low	Moderate	Medium	Good	Medium	8th=	9	67.2	2	2	2	2
Status quo (not to adopt)- legacy IT	High	Low	Weak	High	High	Weak	High	11th	11	0	0	0	0	0

References

1. Baltussen R, Niessen L (2006) Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Allocat* 4(1):14
2. Bolloju N (2001) Aggregation of analytic hierarchy process models based on similarities in decision makers' preferences. *Eur J Oper Res* 128(3):499–508
3. Cameron TA, DeShazo J (2010) Differential attention to attributes in utility-theoretic choice models. *J Choice Model* 3(3):73–115
4. Clemen R, Reilly T (2013) *Making hard decisions with DecisionTools*. Cengage Learn, Boston
5. de Lautour H, Dalbeth N, Taylor W (2014) Development of preliminary remission criteria for Gout using Delphi and 1000Minds consensus exercises. Wiley-Blackwell, Hoboken
6. Dillman DA, Smyth JD, Christian LM (2014) Internet, phone, mail, and mixed-mode surveys: the tailored design method. Wiley, Hoboken
7. Doumpos M, Marinakis Y, Marinaki M, Zopounidis C (2009) An evolutionary approach to construction of outranking models for multicriteria classification: the case of the ELECTRE TRI method. *Eur J Oper Res* 199(2):496–505
8. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL (2005) *Methods for the economic evaluation of health care programmes*. OUP Catalogue, Oxford
9. Edwards W (1977) How to use multiattribute utility measurement for social decisionmaking. *IEEE Trans Syst Man Cybern* 7(5):326–340
10. Edwards W, Barron FH (1994) SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. *Organ Behav Hum Decis Process* 60(3):306–325
11. El-Gazzar RF (2014) *A literature review on cloud computing adoption issues in enterprises*. Springer, Berlin
12. Forman EH, Selly MA (2001) *Decision by objectives: how to convince others that you are right*. World Scientific, Singapore
13. Godse M, Mulik S (2009) *An approach for selecting software-as-a-service (SaaS) product*. IEEE, USA
14. Han S-M, Hassan MM, Yoon C-W, Huh E-N (2009) Efficient service recommendation system for cloud computing market. ACM, New York
15. Hansen P, Omblér F (2008) A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *J Multi-Criteria Decis Anal* 15(3–4):87–107
16. Hastie R, Dawes RM (2010) *Rational choice in an uncertain world: the psychology of judgment and decision making*. Sage, Thousand Oaks
17. Hussain FK, Hussain OK (2011) *Towards multi-criteria cloud service selection*. IEEE, USA
18. KPMG (2013) *The cloud takes shape*. KPMG, New York
19. Landon EL (1971) Order bias, the ideal rating, and the semantic differential. *J Market Res* 8:375–378
20. Li A, Yang X, Kandula S, Zhang M (2010) *CloudCmp: comparing public cloud providers*. ACM, New York
21. Marston S, Li Z, Bandyopadhyay S, Zhang J, Ghalsasi A (2011) Cloud computing—the business perspective. *Decis Support Syst* 51(1):176–189
22. Martín-Collado D, Byrne T, Amer P, Santos B, Axford M, Pryce J (2015) Analyzing the heterogeneity of farmers' preferences for improvements in dairy cow traits using farmer typologies. *J Dairy Sci* 98:4148–4161
23. Moshkovich HM, Mechitov AI, Olson DL (2002) Ordinal judgments in multiattribute decision analysis. *Eur J Oper Res* 137(3):625–641
24. Nielsen H, Amer P, Byrne T (2014) Approaches to formulating practical breeding objectives for animal production systems. *Acta Agric Scand A: Anim Sci* 64(1):2–12
25. Oliveira T, Martins MF (2011) Literature review of information technology adoption models at firm level. *Electron J Inf Syst Eval* 14(1):110–121
26. Omblér F, Hansen P (2012) 1000Minds software

27. Perreault WD (1975) Controlling order-effect bias. *Publ Opin Q* 39:544–551
28. Raghavarao D, Wiley JB, Chitturi P (2010) *Choice-based conjoint analysis: models and designs*. CRC Press, Boca Raton
29. Ryan M, Gerard K (2003) Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy* 2(1):55–64
30. Saaty TL (2008) Decision making with the analytic hierarchy process. *Int J Serv Sci* 1(1):83–98
31. Saaty TL (1990) How to make a decision: the analytic hierarchy process. *Eur J Oper Res* 48(1):9–26
32. Sullivan T (2012) Using MCDA (multi-criteria decision analysis) to prioritise publicly-funded health care, University of Otago, Dunedin
33. Timmermans J, Stahl BC, Ikonen V, Bozdag E (2010) The ethics of cloud computing: a conceptual review
34. Venters W, Whitley EA (2012) A critical review of cloud computing: researching desires and realities. *J Inf Technol* 27(3):179–197
35. Von Winterfeldt D, Edwards W (1986) *Decision analysis and behavioral research*. Cambridge University Press, Cambridge
36. Wu Y, Cegielski CG, Hazen BT, Hall DJ (2013) Cloud computing in support of supply chain information system infrastructure: understanding when to go to the cloud. *J Suppl Chain Manage* 49(3):25–41
37. Yang H, Tate M (2012) A descriptive literature review and classification of cloud computing research. *Commun Assoc Inf Syst* 31(2):35–60

Cost Analysis Between Statins and Hepatocellular Carcinoma by Using Data Mining Approach

Yu-Tse Tsan, Yu-Wei Chan, Wei-Chen Chan and Chin-Hung Lin

Abstract Statin use for cancer may be a potential protective effect in patients with chronic hepatitis B virus infection, according to our research. Statin use reduced the associated risk of liver cancer. With people eating westernized in Taiwan, the use of statins increases year by year. In addition to prevention of cardiovascular disease, also for patients with chronic hepatitis on the preventive effect of hepatocellular carcinoma, the economic benefits of statin use is worth studying. We used the National Health insurance data to explore so far, the statin use since 1997 and the clinical efficacy, including adverse effects. According to the results of our study, statin use for HCC had a potential protective effect in patients with chronic hepatitis B virus infection. Statin use reduced the associated risk of HCC, in particular, our studies found that statin use significantly reduced risk of HCC with cost-effectiveness.

Keywords Cost-effectiveness · Hepatitis B virus · Hepatocellular carcinoma

1 Introduction

The prevalence rates of viral hepatitis are very high in Taiwan; more than 90 % of the general population having contacted HBV infection; and the prevalence of chronic infections is as high as 15–20 %. Additionally, liver cancer is the first of 10

Y.-T. Tsan

Division of Occupational Medicine, Department of Emergency Medicine,
Taichung Veterans General Hospital, Taichung, Taiwan

W.-C. Chan

School of Medicine, Chung Shan Medical University, Taichung, Taiwan

Y.-W. Chan

Department of Hospitality Management, Chung Chou University of Science
and Technology, Yuanlin, Taiwan

C.-H. Lin (✉)

Department of Information Management, Chung Chou University of Science
and Technology, Yuanlin, Taiwan

leading cancers for males and the forth for females. Carriers of HBV infection have a substantial risk of HCC and liver-related death compared with individuals not infected with HBV in Taiwan.

3-Hydroxy-3-methylglutaryl-CoA reductase inhibitors, commonly referred to as statins, have therapeutic as well as primary and secondary preventative effects in cardiovascular disease and stroke [1–4]. Recently, there has been emerging interest in use as anticancer agents based on preclinical evidence of their anti-proliferative, proapoptotic, anti-invasive, and radiosensitizing properties [5, 6]. Inhibition of 3-hydroxy-3-methylglutaryl-CoA reductase by statins interferes with the rate-limiting step of the mevalonate pathway, leading to reduced levels of mevalonate and its downstream products, so that statins may reduce tumor initiation, growth, and metastasis. In fact, statins have demonstrated growth-inhibiting activity in cancer cell lines and preclinical tumor models in animals [7–14]. Observational studies have raised the possibility that statin use may decrease overall risk of cancer and of specific cancers [15–18]. National Health Insurance Bureau of Statistics National Health Insurance for the total drug costs last year to 142.2 billion NT dollars, further broken down into the various diseases, lipid-lowering drug costs about 6.811 billion NT dollars, accounted for the fourth of the Taiwan National Health Insurance drug. In recent years, the average life expectancy is gradually increased, and changes in dietary patterns of life, so people suffered from high blood pressure, high cholesterol and other cardiovascular diseases. The cost-effectiveness between statin use and the risk of hepatocellular carcinoma in patients with chronic hepatitis is needed for further surveillance.

Previous foreign studies pointed out that the use of statin drugs for secondary prevention of cardiovascular disease with economic clinical benefit (cost-effectiveness). statin use for each quality adjusted life—years (quality adjusted life years, QALYs) can be more than about \$20,000 harvest in terms of the costs of health care is a very good suggestion. Literature in particular has a number of controlled clinical trials confirmed that carried in the lower low-density lipoprotein cholesterol (LDL-C) will reduce coronary heart disease. Statin use in reducing LDL-C, generally known as “bad cholesterol” the most effective. However, in groups of low risk for cardiovascular disease in the statin use whether to remain economically clinical benefit worthy of discussion. But the cost-effectiveness of statin use on HCC prevention is still unknown, so our study is focused on this important issue. In addition to prevention of cardiovascular disease, also for patients with chronic hepatitis on the preventive effect of hepatocellular carcinoma, the economic benefits of statin use is worth studying.

2 Data Mining Method

2.1 Study Design

The National Health insurance data was used to explore so far, the statin use, and non-frequent users from 1999 to 2010. We explored statin drugs since 1999 to the date of actual usage. Besides, we also evaluate the clinical efficacy and economic assessment between statin users and non-users.

2.2 The Database of Taiwanese National Health Insurance

The Taiwanese National Health Insurance (NHI) program provides compulsory universal health insurance, implemented on March 1, 1995, which covers all forms of health-care services in 98 % of the island's population. In cooperation with the Bureau of NHI, the National Health Research Institute (NHRI) of Taiwan randomly sampled a representative database of 1,000,000 subjects from the year 2005 registry of all NHI enrollees using a systematic sampling method for research purposes, as the Longitudinal Health Insurance Database (LHID). There were no statistically significant differences in age, gender, or healthcare costs between the sample group and all enrollees, as reported by the NHRI.

The importance of the National Health Insurance database cannot be underestimated, since, not only does it contain virtually all of the health insurance medical records for all citizens in Taiwan, but it is possibly the most comprehensive record of statin users available anywhere in the world. The study was mainly from a computerized database, which is population based and is highly representative. Because the subjects were selected from a simple random sampling of an insured general population, we can rule out the possibility of selection bias [19]. As the data of statin use were obtained from an historical database that collects all available prescription information before the date of HCC, we can rule out the possibility of recall bias. We utilized databases for admissions and outpatient visits of the sample cohort, both of which included information on patient characteristics, including sex, date of birth, date of admission, date of discharge, dates of visits, and up to five discharge diagnoses or three outpatient visit diagnoses (by International Classification of Diseases, Ninth Revision (ICD-9) classification) [20]. The data files also contained information on patient prescriptions, including the names of prescribed drugs, dosage, duration, and total expenditure. So the NHIRD provides a good chance and the material to evaluate the cost-effectiveness between statin use and the risk of hepatocellular carcinoma in patients with chronic hepatitis.

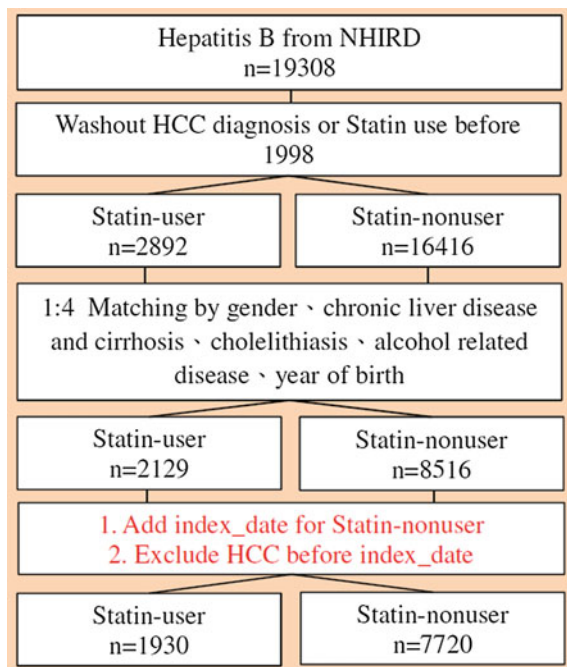
2.3 Identification of Study Sample

We conducted a population-based cohort study in which all patients older than 18 years who had a first-time diagnosis of HBV infection (ICD-9 codes 070.2, 070.3, V02.61) without hepatitis C virus (HCV) infection (ICD-9 codes 070.7, 070.41, 070.44, 070.51, 070.54, V02.62) between January 1, 1999 and December 31, 2010 formed the study cohort. To select potential case patients in this cohort, HCC cases (ICD-9 code 155.0) were identified in the admission files and the date of the first diagnosis of HCC was used as the index date. The American Association for the Study of Liver Diseases Practice Guidelines are recommended by the NHI Bureau for the diagnosis of HCC. We included only admitted HCC cases under the stringent criteria to affirm the diagnosis. We limited eligible cases to those newly diagnosed between January 1, 1999, and December 31, 2010.

2.4 Statin User and Non-User

We identified patients who filled prescriptions for statins in the inpatient and ambulatory care order files between January 1, 1999 to December 31, 2010. We collected the dates of prescriptions, the daily dose, the number of days supplied, and the number of pills per prescription. In accordance with the Anatomical Therapeutic Chemical classification of drugs, we selected simvastatin, lovastatin, atorvastatin,

Fig. 1 Flowchart of statin user and non-user



fluvastatin, pravastatin, and rosuvastatin as the major study drugs of interest. We defined statin user as the case group, which compared with the statin non-user as comparison group by 1:4 matching age, sex, liver cirrhosis, cholelithiasis, alcohol related diseases. The index date of comparison group was the same with case group defined as the first date to use statins (Fig. 1). We also divided the both groups into the HCC and non-HCC groups according to the incidence of HCC.

2.5 Potential Confounders

We systematically identified potential confounding risk factors for HCC as the following diagnoses recorded between January 1, 1999, and the HCC index date: alcohol-related disease (ARD) (ICD-9 codes 291, 303.0, 303.9, 305.0, 571.0, 571.1, 571.2, or 571.3), cirrhosis (571.2, 571.5, 571.6, 572.2, 572.3, 572.4, 572.8, or 573.0), chronicobstructive pulmonary disease (COPD) (491, 492), and diabetes (250). Sociodemographic characteristics (age, sex, smoking, alcohol habits) were also considered.

2.6 Statistical Analyses

Cox proportional hazard models were used to compute the hazard ratios (HRs) accompanying 95 % confidence intervals (CIs) for HCC after adjustment for the variables mentioned. Because the case group included the cost of all medications, the expenditure was higher than comparison group expectedly. We collected the cost per person-year in statin users and non-users. In addition, we also calculated the cost between HCC and non-HCC groups. All the above analyses were conducted using SAS statistical software (version 9.2; SAS Institute Inc, Cary, NC).

3 Conclusion

3.1 Results of Demographic Characteristics

After matching the index date, age, sex, liver cirrhosis, cholelithiasis, alcohol related diseases between statin user and non-user groups, the overall number of HBV carriers was 9650, including 1930 statin users and 7720 non-users. There were no statistical difference on age, sex, socio-economics between both groups. But hypertension (Statin: 52.33 %; Non-Statin: 28.03 %), Disorders of lipid metabolism (Statin: 79.53 %; Non-Statin: 16.37 %), Diabetes Mellitus (Statin: 41.66 %; Non-Statin: 15.71 %), Coronary Artery Diseases (Statin: 26.37 %;

Table 1 Demographic characteristics of statin users and non-users in chronic hepatitis B patients

		Total		Statin (n = 1930)		Non-statin (n = 7720)	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
Hepatitis B (n = 9650)							
Demographic							
<i>Age</i>							
Age < 20	30	0.31	9	0.47	21	0.27	
20 ≤ Age < 40	2029	21.03	392	20.31	1637	21.2	
40 ≤ Age < 65	6603	68.42	1344	69.64	5259	68.12	
Age ≥ 65	988	10.24	185	9.59	803	10.4	
Mean (95 % CI)	49.37	(49.13-9.10)	49.25	(49.73-49.78)	49.4	(49.14-49.67)	
<i>Gender</i>							
Female	3195	33.11	639	33.11	2556	33.11	
Male	6455	66.89	1291	66.89	5164	66.89	
<i>Income</i>							
0	1164	12.06	232	12.02	932	12.07	
1-15,840	1401	14.52	278	14.4	1123	14.55	
15,841-25,000	4246	44	872	45.18	3374	43.7	
>25,001	2839	29.42	548	28.39	2291	29.68	
<i>Urban</i>							
I	2937	30.44	597	30.93	2340	30.31	
II	4652	48.21	937	48.55	3715	48.12	
III	1425	14.77	264	13.68	1161	15.04	
IV (rural area)	570	5.91	114	5.91	456	5.91	
Missing	66	0.68	18	0.93	48	0.62	

(continued)

Table 1 (continued)

		Total			Statin (n = 1930)			Non-statin (n = 7720)		
		Frequency	Percent		Frequency	Percent		Frequency	Percent	
Morbidity	Liver disease	762	7.9		151	7.82		611	7.91	
	Other disorders of muscle, ligament, and fast	134	1.39		31	1.61		103	1.33	
	Drug induced liver injury	66	0.68		10	0.52		56	0.73	
	Acquired immuno deficiency syndrome	16	0.17		2	0.1		14	0.18	
	Alcohol related disease	360	3.73		72	3.73		288	3.73	
	Chronic obstructive pulmonary disease	811	8.4		198	10.26		613	7.94	
	Cholelithiasis	620	6.42		124	6.42		496	6.42	
	Chronic renal failure	300	3.11		127	6.58		173	2.24	
	Hypertension	3174	32.89		1010	52.33		2164	28.03	
	Disorders of lipid metabolism	2799	29.01		1535	79.53		1264	16.37	
	Heart failure	328	3.4		109	5.65		219	2.84	
	Diabetes mellitus	2017	20.9		804	41.66		1213	15.71	
	Chronic liver disease and cirrhosis	670	6.94		134	6.94		536	6.94	
	Nonalcoholic steatohepatitis	1066	11.05		292	15.13		774	10.03	
	Coronary artery disease	1443	14.95		509	26.37		934	12.1	
	Cancer	328	3.4		42	2.18		286	3.7	

Table 2 Cost-effectiveness analysis between statin users and non-users

Hepatitis B								
HCC	Statin	N	Cost	Time (mean cost/time)		95 % CI		Ratio
0	0	7434	308,363,725.77	5.49	7555.59	7027.59	8083.59	
1	0	286	8,323,281.77	4.41	6599.18	5017.61	8180.76	0.87
0	1	1888	251,280,618.62	5.52	24,111.15	21,644.73	26,577.58	
1	1	42	4,058,133.73	5.42	17,826.98	10,271.43	25,382.53	0.74

Non-Statin: 12.1 %) had statistical meaning. The hazard ratio of statin for HCC in hepatitis B was 0.58 (95 %CI: 0.42–0.80). After adjusting for liver cirrhosis, diabetes mellitus, urbanization, income, the hazard ratio for HCC in hepatitis B was 0.53 (95 %C: 0.38–0.74) (Table 1).

3.2 Results of Cost-Effectiveness Analysis

The average cost of each HBV-infected statin non-user without HCC was NT 7556 per year, and 6599 for each non-user with HCC. The cost of each statin user without HCC per year was 24,111, and statin user with HCC was 17,827. In comparison group, the ratio of expenditure was lower in HCC patients (Ration = 0.87), and 0.74 in statin users. After logging the cost, the hazard ratio for HCC was 0.78 (95 %CI 0.72–0.85). The result showed that the risk declined 22 % while added per Log (Cost) (Table 2).

3.3 Discussion

Statin use for cancer may be a potential protective effect in patients with chronic hepatitis B virus infection, according to our research. Statin use reduced the associated risk of liver cancer. With people eating westernized in Taiwan, the use of statins increases year by year. In addition to prevention of cardiovascular disease, also for patients with chronic hepatitis on the preventive effect of hepatocellular carcinoma, the economic benefits of statin use is worth studying. The results supported that statin users had benefit on cost-effectiveness of incidental HCC.

References

1. Goldstein JL, Brown MS (1990) Regulation of the mevalonate pathway. *Nature* 343:425–430
2. Keyomarsi K, Sandoval L, Band V et al (1991) Synchronization of tumor and normal cells from G1 to multiple cell cycles by lovastatin. *Cancer Res* 51:3602–3609

3. Kusama T, Mukai M, Iwasaki T et al (2002) 3-Hydroxy-3-methylglutaryl-coenzyme a reductase inhibitors reduce human pancreatic cancer cell invasion and metastasis. *Gastroenterology* 122:308–317
4. Weis M, Heeschen C, Glassford AJ et al (2002) Statins have biphasic effects on angiogenesis. *Circulation* 105:739–745
5. Gauthaman K, Fong CY, Bongso A (2009) Statins, stem cells, and cancer. *J Cell Biochem* 106:975–983
6. Chan KK, Oza AM, Siu LL (2003) The statins as anticancer agents. *Clin Cancer Res* 9:10–19
7. Dimitroulakos J, Marhin WH, Tokunaga J et al (2002) Microarray and biochemical analysis of lovastatin-induced apoptosis of squamous cell carcinomas. *Neoplasia* 4:337–346
8. Jacobs EJ, Newton CC, Thun MJ et al (2011) Long-term use of cholesterol-lowering drugs and cancer incidence in a large United States Cohort. *Cancer Res* 71:1763–1771
9. Friis S, Poulsen AH, Johnsen SP et al (2005) Cancer risk among statin users: A population-based cohort study. *Int J Cancer* 114:643–647
10. Poynter JN, Gruber SB, Higgins PD et al (2005) Statins and the risk of colorectal cancer. *N Engl J Med* 352:2184–2192
11. Karp I, Behloul H, Leloir J et al (2008) Statins and cancer risk. *Am J Med* 121:302–309
12. Research, N. H. R. I. N. H. I., database. http://www.nhri.org.tw/nhird/date_01.html#_edn1. Accessed 1 Jan 2012
13. Centers for Disease Control and Prevention, Atlanta, Georgia (1979) International Classification of Diseases, Ninth Revision (ICD-9). <http://www.cdc.gov/nchs/icd/icd9.htm>. Accessed 10 Jan 2012
14. Khan BH (2000) A framework for web-based learning. Educational Technology Publications, Englewood Cliffs
15. Liaw S-S, Huang H-M, Chen G-D (2007) Surveying instructor and learner attitudes toward e-learning. *Comput Educ* 49:1066–1080
16. Wang Y-S, Wang H-Y, Shee DY (2007) Measuring e-learning systems success in an organizational context: scale development and validation. *Comput Hum Behav* 23:1792–1808
17. Chikh A, Berkani L (2010) Communities of practice of e-learning, an innovative learning space for e-learning actors. *Procedia Soc Behav Sci* 2:5022–5027
18. Sultan N (2010) Cloud computing for education: a new dawn? *Int J Inf Manage* 30:109–116
19. Arshad J, Townend P, Xu J (2011) A novel intrusion severity analysis approach for Clouds. *Future Gener Comput Sys*. doi:10.1016/j.future.2011.08.009
20. Loganayagi B, Sujatha S (2012) Enhanced cloud security by combining virtualization and policy monitoring techniques. *Procedia Eng* 30:654–661
21. Baigent C, Keech A, Kearney PM et al (2005) Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* 366:1267–1278
22. Hebert PR, Gaziano JM, Chan KS et al (1997) Cholesterol lowering with statin drugs, risk of stroke, and total mortality. An overview of randomized trials. *JAMA* 278:313–321
23. Kjekshus J, Apetrei E, Barrios V et al (2007) Rosuvastatin in older patients with systolic heart failure. *N Engl J Med* 357:2248–2261
24. Amarenco P, Bogousslavsky J, Callahan A 3rd et al (2006) High-dose atorvastatin after stroke or transient ischemic attack. *N Engl J Med* 355:549–559
25. Sassano A, Plataniias LC (2008) Statins in tumor suppression. *Cancer Lett* 260:11–19
26. Newman TB, Hulley SB (1996) Carcinogenicity of lipid-lowering drugs. *JAMA* 55–60
27. Wong WW, Dimitroulakos J, Minden MD et al (2002) HMG-CoA reductase inhibitors and the malignant cell: the statin family of drugs as triggers of tumor-specific apoptosis. *Leukemia* 16:508–519

Hospital Service Queue Management System with Wireless Approach

Manoon Ngorsed and Poonphon Suesaawaluk

Abstract This paper presents a proposed alternative system for queuing management that could reduce inconvenience to the public. The motivation of this system is depicted from an observation on the people queuing for services in the hospitals and the government offices without committing to the estimated time for their demand. Waiting for the service is counterproductive which consumes an unacceptable amount of productive time for the patients. We develop the system to manage the queue without physically lining up and allow people to monitor their queue status by their wireless handheld devices. The project accomplishes its objective as a tool to manage the hospital queue online where customers, patients and stakeholder can access their queues remotely over the Internet through a web application. The results benefit to both stakeholder to manage their time for other desire activities and hospitals in utilizing its spacious area for other business proposes.

Keywords Hospital queuing management system · Web application

1 Introduction

The innovation of technologies could bring support to the quality of life for human in various aspects and objectives. However, in order to apply and implement technology system to be used requires the costly investment for itself. This constraint leads to the inescapable archaic management methods, and the systems still coexist alongside the advances in procedures. One of the unavoidable significances is the hospital service for the people, especially among the undeveloped country

M. Ngorsed (✉) · P. Suesaawaluk (✉)
Graduate School of eLearning in Information Communication Technology,
Assumption University, Bangkok, Thailand
e-mail: g5371304@au.edu

P. Suesaawaluk
e-mail: poonphonssw@au.edu

and developing country. The public hospitals likely support the poor and middle classes which have to patronize the public services in the state hospitals.

A growing population base will continue having a pressure to the existing hospital facilities. With the cycle of limited facilities, it leads to the coupled staffing shortages which will guarantee that long queues to remain synonymous anytime visiting a hospital and other public service facilities. The people must take a queue as long as they need the services. Whether the problem is caused by staff shortages, equipment shortages, or the hospital capacity is not sufficient for the population area they serve. Long queues are an unwanted and unnecessary burden to the public as well as the hospital staffs. Long queues are then associated with a negative image of the hospital experience, but most people can't avoid to be under this present system.

For this project, we propose the system with the main objective as to create a visual queue for hospital online where people can access and reserve their queue wirelessly over the Internet. The system allows people to monitor their queuing status from the web service application. This beneficial system is designed to offer the options for people who are waiting for the service; they can go anywhere while they are in the queue rather than standing and presenting themselves in front of the service area.

2 Literature Review

The traditional queuing management methods mostly used in the hospital are queue card and smart queue as described it featured by Fig. 1. When using queue card system, the people in the queue are assigned by numbers according to the arrival order. This method allows the patients to be able to manage their time based on an estimation of the time available until their number is called. Venturing outside of the immediate area is a constant gamble. The queue number may guarantee service according to the number priorities; however, a delay in returning may still result in the loss of a queue position.

Most of the private hospitals provide a smart queue system as well as helpdesks and counter services for their customers. The smart queue system provides automatic queue numbers along with automatic voice calling and LED display panels on a progressive basis. However, this system still requires patients to congregate in the immediate area to monitor the progress of queue numbers being serviced. This service only eliminates the need to stand in an organized line, but does not address a more productive method for time utilization

Based on a survey, people waiting in a queue get a service from public hospitals in rural area in Thailand reveal that they are compelled to endure the endless waits. They lined up at the service counter. Any abandonment results in their requirement to return to the back of the line and an even longer wait. With such a long queue and waiting period, it represents a considerable amount of time wasted for the people involved. Any desire to venture outside the immediate area is outweighed by the uncertainty of not having information regarding the progress of the queue. They

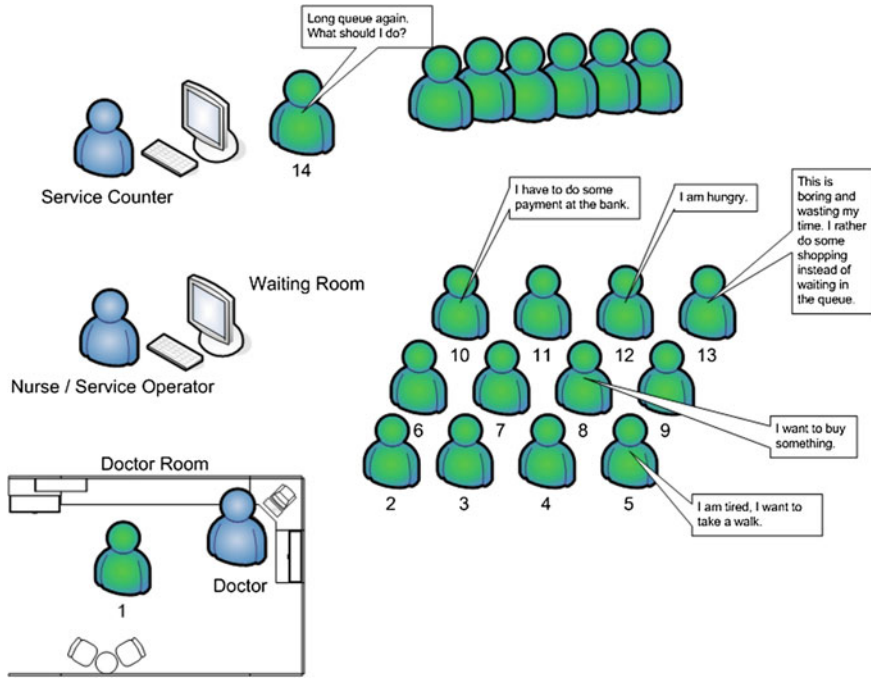


Fig. 1 Typical state hospital queue management system

simply cannot miss their position due to a lack of information. This problem motivated us to develop a method to manage the reserved queue to alleviate on minimizing the number of people in the physical queue

3 Service Queue Management System with Wireless Approach

3.1 System Boundary and Architecture

The new approach of the hospital queue management system will provide stakeholder with tools to manage their queue status wirelessly [1]. The system would allow them to know what is going on with the queue wherever they go. As can be seen in the Fig. 2 a new comer arrives at the service counter before booking into the hospital queue. With their wireless devices, the queue status can be accessed through the Internet, and it provides information to everyone in the queue.

The proposed system, the boundary and its functionality are described in a form of UML concepts [1, 2] shown in Fig. 3. The system's functionality is demonstrated and explained as the role of four actors and seven use cases as following:

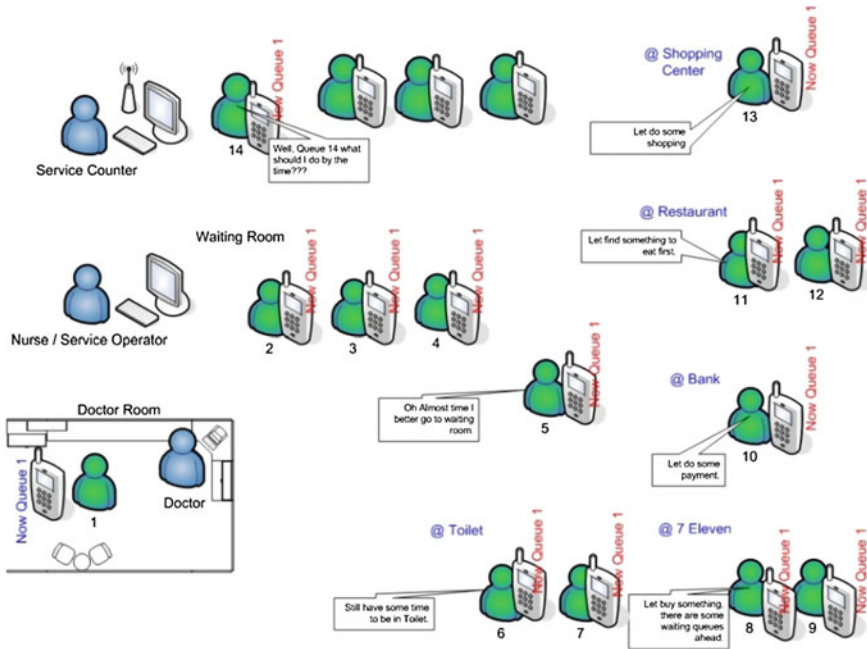


Fig. 2 Existing hospital service queue management system

Actors role;

- **System Admin:** represents an administrator who grants access to all system features; the role is to register a new hospital and queue administration to the system.
- **Queue Admin:** represents a hospital queue administrator, the role is to create queues and operators to the system.
- **Queue Operator:** represents a person who takes care of each queue. The role is to register queue client to the system and to manage all activities in the queue.
- **Queue Client:** represents the person who requires hospital service and is seated in the queue. The role is to view the queue status in order to know when to be in the service.

Use case role;

- **Register Hospital:** Describes a behavior for the system administration to register hospital details into the hospital queue system.
- **Register Queue Admin:** describes a behavior that a queue administration is created by the system admin.
- **Create Queue:** Describes a queue that is created by queue administration.
- **Register Queue Operator:** Describes a behavior that a queue operator is created by the queue administration.

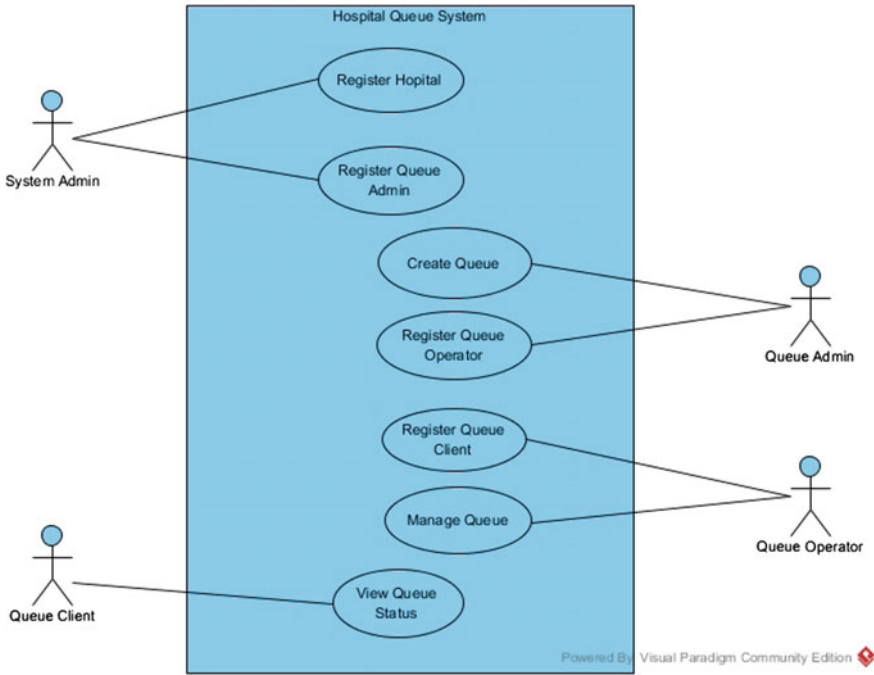


Fig. 3 Hospital queue system use case diagram

- **Manage Queue:** Describes how the queue operator manages all activities that happen in a queue, including the insertion of a client to the queue, put queued client into a service and end the client from the queue after the service is complete.
- **View Queue Status:** Describes a behavior where a queue client can check or view their queue status during the queue process.
- **View Queue Status:** Describes the behavior where a queue client can check or view their queue status during the queue process.

3.2 Database and Development Tools

We consider using Java programming run on Java EE environment [3], Glassfish as its web server and running web service on WSDL (Web Services Description Language) model running with XML to view and exchange data. A system based on Java technology could be strong in term of security and great in term of performance and implement with RESTful [4] architecture. The set of tools is an advantage of ease to access and deployment.

Database is created by using PHPMyAdmin [5] which is MySQL [6] management program. It comes together with XAMP [7]. The entities and relationships system is deployed with relational database [8] principle which consists of five tables. Qbusiness table stores the entity of the hospital that is registered to the hospital queue system. Quser table stores the entity of user who works with the system. Qqueue table stores the entity of a queue which is created by queue administration. Qtran table stores the entity of a queue transaction which is generated through queuing process. Customer table stores the entity of a queue client or a patient who requires hospital services and seating in the queue.

3.3 *Queuing Management Mechanism*

In this project “First Come, First Serve” concept and queuing theory with Little’s law [9, 10] is deployed as the system discipline to manage service queue. Given λ is the average number of items arriving per unit time; W is average waiting time per for an item, and L is an average number of items in the queuing system, so $L = \lambda W$.

The Arrival Rate (λ) is formulated by a division of Total arrival (N) by Total Time (T) as $\lambda = N/T$. This means that at the time interval T the system has been observed, the number of arrival N entering to the system queue.

Finding individual waiting time: In order to find the time remaining or waiting time for an individual in the queue, we need to know the average waiting time W of the system at the period time T by being calculated from Eq. (1).

$$W = \frac{1}{N} \sum_{i=1}^N W_i \quad (1)$$

To calculate the waiting time for the N th queue number to be in service, the average waiting time needs to be calculated onward to get the most likely average time. For instance, the queue may have an average L customers waiting in the queue with arrival rate λ , so calculating the average waiting time is $W = L/\lambda$. Therefore, the expected waiting time for the N th queue to reach the service is

$$WN = \sum_{i=1}^N W_i \quad (2)$$

According to the Eq. (2), an individual waiting time could be rewritten as the average waiting time multiply by the number of individual a queue number approximately. As of continuous system, the estimate time waiting could be denoted as the Eq. (3).

$$W(\mathbf{n}, \mathbf{m}) = mW_n \quad (3)$$

where \mathbf{m} is the queue number of an individual queue and \mathbf{n} is the number of queues included in average, an estimated waiting time of an individual, $W_{(n,m)}$ could be suggested to the customer of the queue as the multiplication of queue number m with the average waiting time W_n .

4 System Prototype Implementation

The proposed hospital queue system is required to run over the Internet or intranet; therefore, the stakeholder, system administrative users and patients can use their smart phones and Internet access devices to view their queue status. The system prototype is demonstrated by testing with a set of tools and equipment as described below:

- Locally testing with XAMP
The web server needs to be set up and tested on Windows environment and running on XAMP v3.2.1 [7], which is a bundle package of Apache, MySQL and PHP [11, 12]. However, the system cannot fully operate locally since the customer/client/patient must be able to view a queue status over their wireless device. Therefore, the system has to be online to serve this requirement.
- Online hosqueue.com
To take the system online, a domain name needs to be registered. Also, it has been named as hosqueue.com. The system domain is also hosted with one of a domain hosting providers which gives the system space and requires server environment for the system to run.

The system has been done on top of the previous code taken from the open source software called Complain Management System written by Tousif Khan [13]. The system is built on top of pre-coding and structure with a new database design and new business processing. The system is built on PHP [11, 14] and Java Script [4], coding example is shown in the following page. The code represents part of PHP requesting query to the database before converting into JSON data format which performs RESTful web service. As it can be seen in the code, one important requesting element is AvgTime (Average Time). The system allows the queue admin to modify the number of samples of individual waiting time as a set of average time waiting. AvgTime is then to be used to calculate time remaining for the next remaining queue.

```

$sql = "SELECT
qtran.QueueID, queueno, CustID, arrive, tstatus, qqueue.AvgTime
FROM ".$dbname.".qtran INNER JOIN ".$dbname.".qqueue ON
qtran.QueueID = qqueue.QueueID WHERE qtran.QueueID
='".$qid."'";
$result=mysql_query($sql); $rows = array();
while($r = mysql_fetch_assoc($result)) {
$rows['Queue'][] = $r; }
printjson_encode($rows);

```

[Example of PHP script on data conversion by using json_encode function yields requesting queue data output into JSON data format.]

The main operation is on queue management system on PHP demonstration. When the customer/client/patient queue viewer is mainly on Android [3, 15, 16] application, coding example is shown below. This part of the code allows the application to retrieve JSON data format from the PHP web service. AvgTime abruptly calculates time remaining equivalent to the sequential order of the patient queue number. This part of the system allows the user to access data over their wireless device.

```

public void ListDrawer() {
    try{JSONObject jsonResponse =
        newJSONObject(jsonResult);
        JSONArray jsonMainNode =
            jsonResponse.optJSONArray("Queue");
        rowQueue.clear();
        for (inti = 0; i<jsonMainNode.length(); i++) {
            JSONObject jsonChildNode =
                jsonMainNode.getJSONObject(i);
            columnQueue.set(0, jsonChildNode.optString("QueueID"));
            columnQueue.set(1, jsonChildNode.optString("queueno"));
            columnQueue.set(2, jsonChildNode.optString("CustID"));
            columnQueue.set(3, jsonChildNode.optString("arrive"));
            columnQueue.set(4, jsonChildNode.optString("tstatus"));
            columnQueue.set(5, jsonChildNode.optString("AvgTime"));
            columnQueue.set(6, String.valueOf(Integer.valueOf(
                jsonChildNode.optString("AvgTime"))*60*(i+1)));
            columnQueue.set(7, String.valueOf(Integer.valueOf(
                jsonChildNode.optString("AvgTime"))*60*(i+1)));
            rowQueue.add(new ArrayList<String>(columnQueue));
        }
    } catch (JSONException e) { ... }}

```

[Example of Android programming function called ListDrawer, which retrieves JSON data format and displays on Android client application.]

- **Installation Client Application with Android**

An Installation client program for Android application hosqueue.com is stored in an APK file after its compilation. The customer can download the file and install it to an android device. The program requires running on Android 4.0.3 (Ice Cream Sandwich) and above (Fig. 4).

The system function will display all the queue and find queue by ID as shown in Fig. 5.

- **Display All Queue:** The Display All Queue button leads to a view by queue selected screen where the customer can view the queue by choosing a particular queue that they want to view. Therefore, the customer is required to know which queue to look for.
- **Find Queue by ID:** The Find Queue by ID helps the customer in searching the queue in case that the customer does not know which queue it is. However, the customer is still required to know their customer ID to be used as a finding key to the queue.

The queue displays the queue number, customer ID., queue status, and estimated time of service. This can help the customer go anywhere nearby or do other activities while still knowing the queue status.

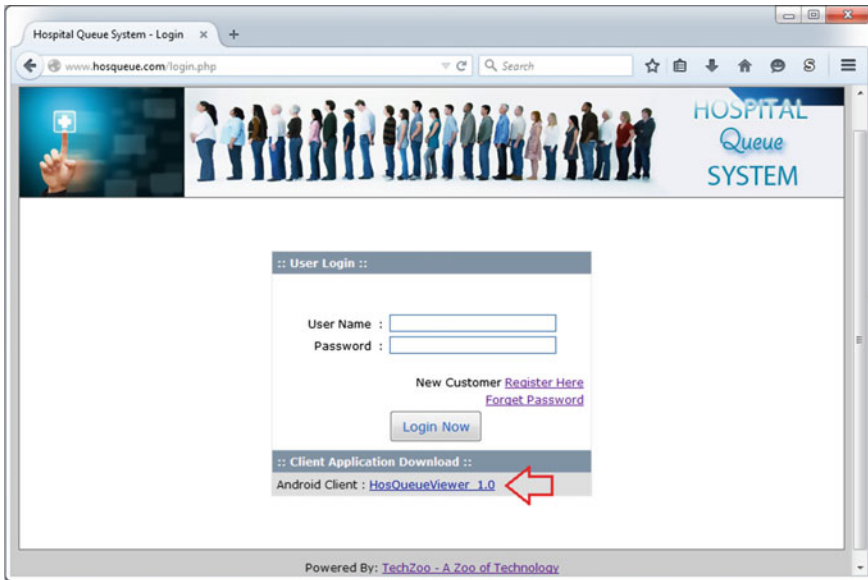


Fig. 4 Available download of hospital queue viewer 1.0

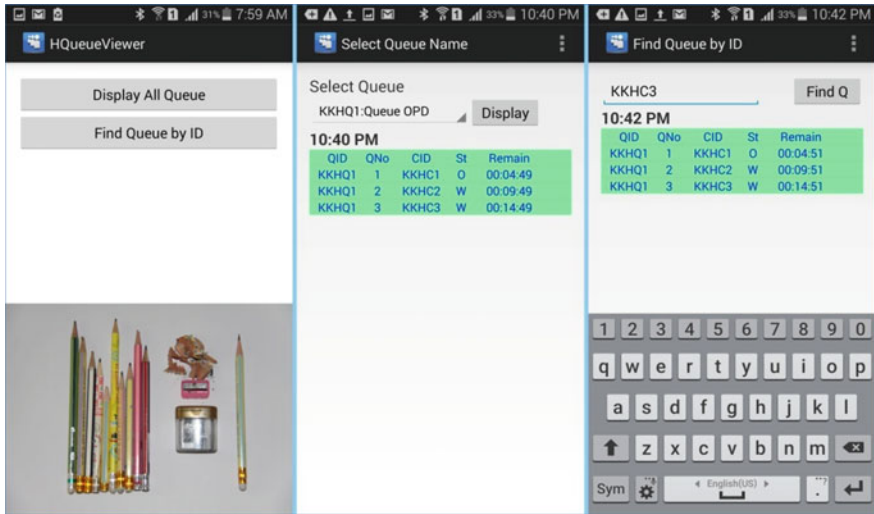


Fig. 5 Queue display by searching customer ID

5 Conclusion

Hospital Service Queue System is a project to eliminate the traditional physical queue and replace it with a convenient management. This project is designed to help the public who suffers from long queues in hospitals, especially the public hospitals. The main system functionalities which are constructed and implemented online are ready for hospital queue services; hence, the customer/patient/client can view a queue status over their wireless.

The contribution of this system does not only serve the people requesting the service in hospital but also utilize their time to do other activities. Also, the advantage of using open source it could benefit to community as a whole. Not only one hospital can benefit with the current system design and setting, but multiple hospitals can be served at the same time. An individual hospital can manage its own queues with a given power user as a queue administration. With this design, a cost sharing arrangement is possible amongst hospitals without having any budget to spend for the extra development.

References

1. Kendall KE, Kendall JE (2011) Systems analysis and design, 8th edn. Pearson Education, Harlow
2. Bruegge B, Dutoit AH (2010) Object-oriented software engineering using UML, Patterns, and Java, 3rd edn, International Edition. Pearson Education, Upper Saddle River
3. Java Software—Oracle. <https://www.oracle.com/java>. Accessed 15 Apr 2015

4. The World Wide Web Consortium (W3C). <http://www.w3.org>. Accessed 12 May 2015
5. phpMyAdmin. <http://www.phpmyadmin.net>. Accessed 23 Feb 2015
6. MySQL Community. <http://www.mysql.org>. Accessed 23 Feb 2015
7. XAMPP Installers and Downloads for Apache Friends. <https://www.apachefriends.org>. Accessed 23 Feb 2015
8. Hoffer JA, Mary S, Heikki T (2011) Modern database management, 10th edn. Prentice-Hall, Upper Saddle River
9. Chhajed D, Lowe TJ (2008) Building intuition: insights from basic operations management models and principles. Springer, Heidelberg, pp 81–84
10. Cooper RB (1981) Introduction to queueing theory, 2nd edn. Elsevier North Holland, Inc., New York, pp 178–185
11. PHP—Hypertext Preprocessor. <http://php.net>. Accessed 12 Feb 2015
12. The Apache Software Foundation. <https://www.apachefriends.org>. Accessed 23 Feb 2015
13. A Zoo of Technology. <http://www.techzoo.org>. Accessed 15 Nov 2014
14. Welling L, Thomson L (2008) PHP and MySQL® web development, 4th edn. Addison-Wesley Professional, Boston
15. Meier R (2012) Professional android 4 application development, updated for android 4. Wiley, Indiana
16. Android. <https://www.android.com>. Accessed 17 Jan 2015

A Smartphone Based Hand-Held Indoor Positioning System

Lingxiang Zheng, Zongheng Wu, Wencheng Zhou, Shaolin Weng and Huiru Zheng

Abstract In this paper, we present a smartphone-based hand-held indoor positioning system. The system collects data using the accelerometer, gyroscope and gravity virtual sensor sensors embedded in the smartphone. The accelerometer and gravity data are used to detect zero vertical speed and calculate the vertical displacement of each walking step, and then the Pythagorean Theorem is applied to calculate the step length of every step. Gyroscope data is used to estimate the direction angle. The step length and the direction angle of each step is combined to determine the coordinates of each step. A Kalman filter is used to reduce the vertical speed offset caused by accelerometer drift errors. The testing results show good performance of the proposed system.

Keywords Smartphone · Indoor positioning · Kalman filter

1 Introduction

In many applications, indoor positioning is required, in particular where the Global Positioning System (GPS) is not available, for example, tracking a firefighter in fire scenes, looking for a car in an underground parking, tracking health-care workers and instruments in a hospital, and so on.

Currently, most indoor positioning methods use infrastructure-based approaches and rely on the signals of external devices. It limits their usability, especially when there are not any external sources of signals or these signals are hard to

L. Zheng (✉) · Z. Wu · W. Zhou · S. Weng · H. Zheng
School of Information Science and Engineering, Xiamen University, Xiamen, China
e-mail: lxzheng@xmu.edu.cn

H. Zheng
e-mail: h.zheng@ulster.ac.uk

L. Zheng · Z. Wu · W. Zhou · S. Weng · H. Zheng
School of Computing and Mathematics, Ulster University, Jordanstown Campus,
Shore Road, Newtownabbey, Co. Antrim, UK

setup. Moreover, the wireless signal source is vulnerable to interruptions or interferences, which makes it harder to establish an accurate and stabilized indoor positioning system.

Infrastructure-free-based approaches eliminate the needs of external signals. Most of these approaches are based on accelerometers and gyroscopes. Depending on where the sensors are placed, previous studies generally could be classified into three types: foot mounted [1–8], waist mounted [9–12] and hand-held [13–15]. In our previous research [16, 17], we proposed a 3D indoor positioning system using low cost MEMS sensors mounted on foot, algorithms were developed to calculate the position using a Kalman filter. In [18], the author put a smart phone in pocket and estimated the vertical displacement by double integral vertical velocity, then calculate the step length through the Pythagorean Theorem, and obtained the attitude information through map matching. In [15], the author presented an activity sequence-based indoor pedestrian localization approach using smartphones held in hand in front of body. Nonetheless, these methods still use specialized sensors, require a floor plan, and have to go through a training phase or suffer from low accuracy.

To overcome the above limitations, we developed an indoor positioning system using a smartphone held on hand static to body and the system doesn't require training nor a floor plan. This method uses a smartphone application to collect the data of phone's built-in accelerometer, gyroscope and virtual gravity sensor. A 5-dimensional Kalman filter is designed in accordance with the motion equation to calculate the vertical velocity. The vertical displacement is the integral of vertical velocity, step lengths are then obtained using the Pythagorean Theorem given the length of legs. The heading of each step can be estimated by the angular velocity measured by gyroscope.

The sensors can accurately collect data in harsh environment. However, the drift and bias errors of the sensors cause serious problems. Such errors can accumulate over time, so the measured vertical acceleration over time accumulated an increasingly large error. We use a zero velocity update method (ZUPT) method to overcome the sensors errors and improve the accuracy significantly.

The rest of this paper is organized as follows: Sect. 2 introduces details of the design, framework and the processing method. Section 3 presents the experiments and results. Section 4 concludes the paper.

2 Methods

The overall design and the execution framework of the system is shown in Fig. 1. The system consists of five modules: data acquisition, ZUPT, step length calculation, Kalman filter, and trajectory calculation.

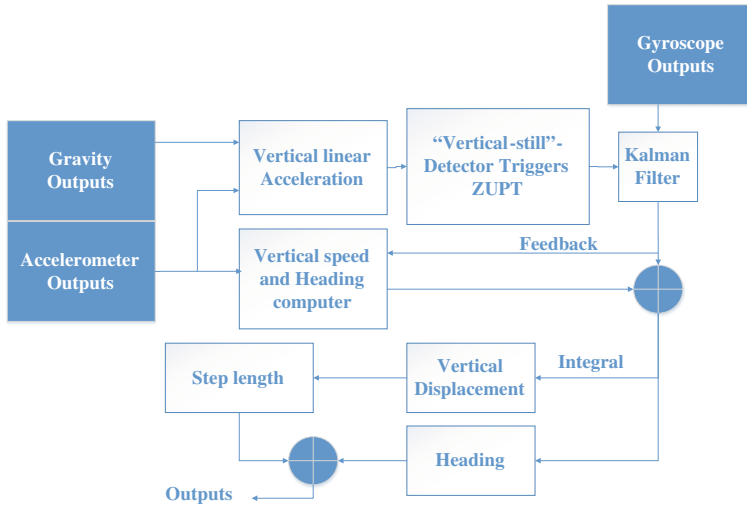


Fig. 1 System architecture

2.1 Data Acquisition

We designed an android APP to acquire data from accelerometers, gyroscopes and gravity sensor (virtual sensor filtered by acceleration) from the smart phone hold by the pedestrian. The sampling interval is 40 ms. We found that sampling interval faster than 40 ms doesn't improve experimental accuracy, only to increase the phone's power consumption. The sampling interval slower than 40 ms will reduce the experimental accuracy. The outputs of the accelerometers and gravity sensor are sent to ZUPT module to detect the zero vertical velocity moments. In addition, a Kalman filter based framework is used to fusion the outputs of the sensors and to estimate the non-linear error of the vertical velocity and heading that increases over time, so the acceleration information is also sent to Kalman filter together with the outputs of the gyroscopes.

2.2 ZUPT

ZUPT (zero velocity update) has been proved to be an effective method to control and eliminate data drift errors. ZUPT is triggered when the vertical velocity of trunk is zero. During one step (from one heel-touching-ground event to the next heel-touching-ground event), there are two events which vertical velocity of trunk will be zero, the first one is a heel-touching-ground event, which happens when the

heel just hits the ground and the trunk is in its lowest position during the entire step [18]. The second one is the stance, which occurs when the foot is flat on the ground and the trunk reaches its highest point. Meanwhile, while these two event happen, the linear acceleration of trunk meets its minimum and maximum points respectively. So we can extract these two zero vertical velocity moments of every step.

Gravity sensor is a virtual sensor obtained by filtering the acceleration sensor, its result is a three-dimensional vector of gravity acceleration. Taken straight up to a positive direction, according the vector dot product principle, the vertical linear acceleration a_z is computed as (1). To extract each of the peaks and valleys will get two zero vertical velocity moments in every steps.

$$a_z = -\frac{(\vec{a} - \vec{g}) \cdot \vec{g}}{|\vec{g}|} \tag{1}$$

where \vec{a} represents the result vector of accelerometer, \vec{g} is the acceleration of gravity. The detection of zero vertical velocity is shown in Fig. 2.

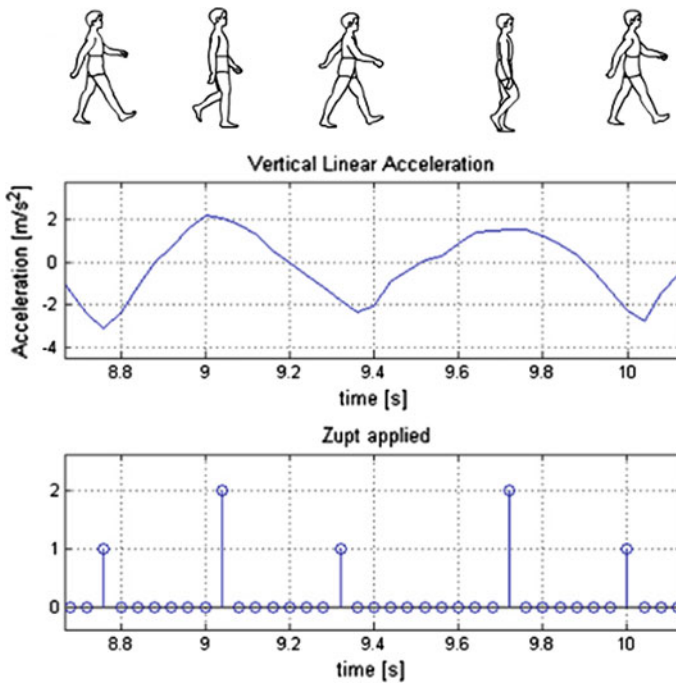


Fig. 2 Detection of zero vertical velocity

2.3 Kalman Filter Design

Navigation parameters used in the study are vertical velocity, vertical displacement, and attitude information. So we built a transition model $X(k)$ which consist of vertical velocity ($V_z(k)$), vertical displacement ($Z(k)$), and attitude information of the pedestrian ($Roll(k)$, $Pitch(k)$, $Yaw(k)$) respectively:

$$X(k) = [Z(k), V_z(k), Roll(k), Pitch(k), Yaw(k)] \tag{2}$$

We listed the equations of motion to solve the vertical velocity and vertical displacement of the transition model:

$$\begin{bmatrix} Z(k+1) \\ V_z(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} Z(k+1) \\ V_z(k) \end{bmatrix} + \begin{bmatrix} 0 & 0 & \frac{T_s^2}{2} \\ 0 & 0 & T_s \end{bmatrix} \otimes \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \tag{3}$$

where T_s represents the sampling interval, a_z , a_y , a_x represents x, y, z-axis acceleration respectively.

While time line moves to the moment where zero vertical velocity is detected by ZUPT, Kalman filter is triggered, updating vertical velocity to zero, and the previous non-zero vertical velocity feeds back to the Kalman filter as the offset to eliminate the accumulated error.

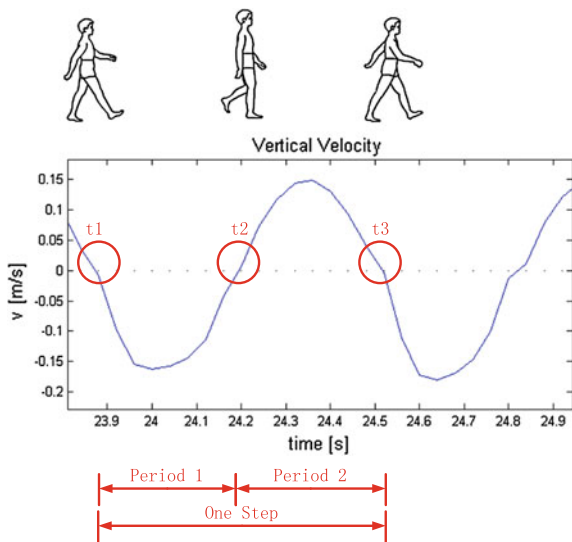
2.4 Step Length Calculation

The transition model $x(k)$ of the Kalman filter consists of vertical velocity, it is sin-like curve which crosses x-axis twice in one step, and the zero points are the moments with maximum and minimum vertical acceleration which detected by ZUPT. The vertical velocity of trunk in one step (from one heel-touching-ground event (t1) to the next heel-touching-ground event (t2) is shown in Fig. 3.

The whole one step can be divided into two periods according to zero points (t1, t2, t3). The period from the start point (t1) to the middle zero point (t2) represents the period from the first heel-touching-ground event to the stance event, that is, the period for the foot behind to catch the other foot which is in front. While the period from the middle zero point (t2) to the end point (t3) represents the period from the stance event to the second heel-touching-ground event, that is, the foot behind not only catches up the front foot (last period) but also opens up a new pace and it will become the front foot in next step.

At the beginning of walking, that is, the beginning of the first step, both feet are in stance state, waiting for opening up a new step. So there is no first period but the second period immediately, making vertical velocity to contain the period above the x-axis only. The final step only contains first period but no second one. The foot

Fig. 3 Vertical velocity of trunk in one step



behind catches the front foot and stops once it has caught it up and the whole walking stops but not to open up new pace. So the graph of final step only contains the part below the x-axis.

So the vertical displacement of trunk in one step can be obtained through integrating the second period of the vertical velocity graph, and then taking the absolute value. Assuming the length of legs L is known, we can obtain the step length ($D(k)$) through the Pythagorean Theorem by (4).

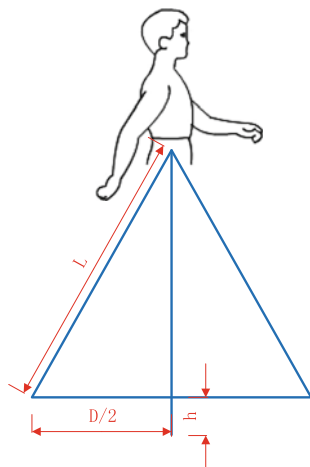
$$D(k) = 2 \sqrt{-2L \int_{t1}^{t2} V_z(k) - \left(\int_{t1}^{t2} V_z(k)\right)^2} \tag{4}$$

The method to obtain step length is also shown in Fig. 4. where L represents the length of leg, D is the step length, h is the vertical displacement of trunk which can be obtained by (5).

$$h(k) = - \int_{t1}^{t2} V_z(k) \tag{5}$$

So the process we calculate step length is actually the process of calculating how far the foot behind is to catch up the other foot in front. The pace opened up by the first step is to be chased by the second step. So the first step should not be counted in (first step has no second period). The final step is not to open up a new pace but to chase the pace opened up by last step. So the final step should be counted in.

Fig. 4 Vertical velocity of trunk in one step



2.5 Trajectory Calculation

Gyroscope measured angular velocity is converted to Euler angles through the quaternion in real time to obtain the heading angle of each sampling point. Because we have obtained length of every step in previous job, what to do next is to calculate coordinate of every step. The method is shown in (6).

$$\begin{bmatrix} x(k+1) \\ y(k+1) \end{bmatrix} = \begin{bmatrix} x(k) \\ y(k) \end{bmatrix} + D(k+1) \times \begin{bmatrix} \sin(\text{Yaw}(k+1)) \\ \cos(\text{Yaw}(k+1)) \end{bmatrix} \quad (6)$$

where $x(k)$ and $y(k)$ represent the coordinate of steps, $D(k)$ represents the step length of the k th step, $\text{Yaw}(k)$ represents the heading angle of k th step.

3 Experiment and Results

We use the Android smartphone as a carrier to acquire the phone built-in sensors data, to do real-time trajectory computing and to display the trajectory on screen. HTC Droid DNA is used in experiments, which is equipped with accelerometer BMA250, gyroscope InvenSense MPL Gyro, and virtual gravity sensor. The pedestrian in the experiment is a 24 years old male with leg length of 0.96 m, and he walks through a building with the smartphone held on hand static to trunk. The data are recorded at 25 Hz clock rate. All the experiments were conducted at the 6th floor of No.2 Research Building in Haiyun Park of Xiamen University. The experimental data are record in text format and can be found at GitHub.¹

¹All the experimental data can be found at: <https://github.com/ECG-XMU/SmartPhoneIndoorLocation-Dataset>.

3.1 Simulation of ZUPT and Kalman Filter

One experiment was conducted for detecting the two time of zero vertical velocity and to eliminate the drift error. The pedestrian in the experiment walked in a normal pace.

Using the data collected, the calculated vertical linear velocity and the moments ZUPT triggered were shown in Fig. 5.

We use +1 and -1 as the threshold of vertical linear acceleration to detect zero vertical velocity (the red line in the first graph of Fig. 5). As we can see in the second graph of Fig. 5, almost all the zero vertical velocity moments were detected (we use 1 to mark the heel-touch-ground event, and 2 to mark the stance event).

After obtaining zero vertical velocity moments, Kalman filter was applied to eliminate the drift error of sensors and to calculate the vertical velocity. We extract the calculated vertical velocity data of the experiment from beginning to about 22 s of time to show the power of ZUPT in Fig. 6.

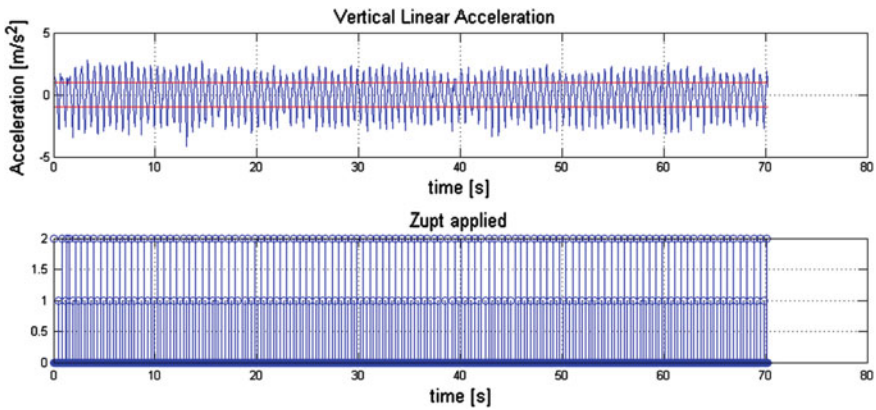
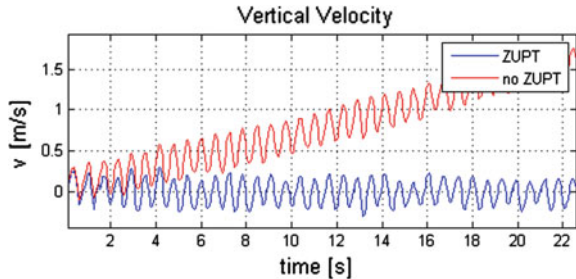


Fig. 5 The calculated vertical linear velocity and the moments ZUPT triggered

Fig. 6 Contrast between vertical velocity with ZUPT and without ZUPT



As shown in Fig. 6, the ZUPT limits the vertical velocity changing range around zero. It has successfully eliminated drift error of accelerator and greatly enhanced vertical velocity accuracy.

3.2 Simulation of Step Length

To verify the performance of the step length calculation, an experiment was carried out. The participant walked from the left side of the building to the right side of the building, then turned back and entered the room 605. The whole path contains four turnings including two clockwise turnings and two anticlockwise turnings. Figure 7 shows calculated vertical velocity, the vertical displacement (integral of vertical velocity) and the step length of the experiment.

As can be seen from data in Fig. 7, vertical displacement of each step fell roughly between 0.025 and 0.05 m, and length of each step fell about 0.55 m, which is consistent with the physiological parameters with a male walking at normal speed whose leg length is 0.96 m [19].

Four phases marked in Fig. 7 represent four turnings along the path. The first two turnings are anticlockwise turnings and the other two are clockwise turnings. As shown in Fig. 7, the step lengths computed while making turnings were significantly smaller than the straight ones, this is consistent with the observation from daily walking.

After the first experiment, another set of three experiments were conducted under same environment to compare the distance of tracks measured by the system and the tape. The three experiments are walking in rectangle, straight and free walking. The distance data collected are presented in Table 1.

Fig. 7 Contrast between vertical velocity with ZUPT and without ZUPT

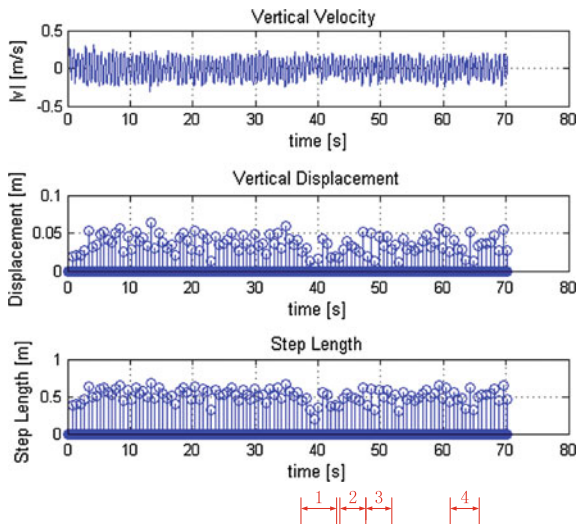


Table 1 The distance measured in given tracks

	Rectangle	Straight	Free walking
Real distance (m)	78	37.2	38.7
Distance calculated (m)	76.9485	36.7204	38.1859
Error rate (%)	1.35	1.29	1.33

where

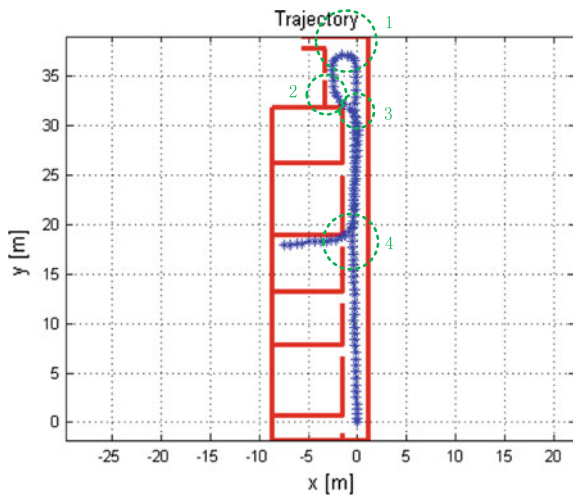
$$Error\ rate = \frac{|True\ distance - Distance\ calculated|}{True\ distance} \tag{7}$$

According to the results, the maximum error rate is 1.35 %, and the average error rate is 1.32 %. So the distances tracked using the system matched the real track within a small variant.

3.3 Experiment of Participant Navigation

In this experiment, the participant held the smartphone on hand relatively stable to his trunk, walking from one side of 6th floor of No.2 Research Build straightly to the other side of building along the corridor. Then, the participant turned left (1), made a big turn till he changed the direction thoroughly and then kept walking. Before reaching the wall, he turned left (2) slightly until backed to the path he has walking. Then he turned right (3) to walk straightly on the previous path but only in the opposite direction. Finally, when he arrived at the door of room 605, he turned

Fig. 8 Experiment of pedestrian navigation



right (4) and entered the room, walked by the wall till he reached the far end of the room. The paths are shown in Fig. 8.

In the experiment, the distance of whole real track is 65.3 m, and the distance of measured whole track is 64.3355 m, the error is about 1.5 %. It can be concluded that the system achieved a reasonable accuracy and the measured track matches the real track.

4 Conclusion

This paper presented an algorithm for indoor positioning using a hand-held smartphone. We developed a method to detect the two times of zero vertical velocity through acceleration and gravity, and to correct the vertical velocity through a Kalman filter. The Pythagorean Theorem is applied to calculate the step length by vertical velocity. The main contribution of this paper is that we obtained high precision results with common hand-held smartphone common on market. The experimental results showed that we successfully achieved indoor positioning without any infrastructural information sources and this result demonstrated that the algorithm could be used in smartphone indoor navigation. Further work will be carried out to evaluation the system with more participants and in different indoor environments. The implementation of a downloadable mobile application will also be undertaken.

Acknowledgment This work was supported by the Natural Science Foundation of China (NSFC, No. 61201196).

References

1. Ojeda L, Borenstein J (2007) Personal dead-reckoning system for GPS-denied environments. In: Proceedings of IEEE international workshop safety, security rescue robot, pp 1–6
2. Alvarez JC, Gonzalez RC, Alvarez D, Lopez AM, Rodriguez-Uria J (2007) Multisensor approach to walking distance estimation with foot inertial sensing. In: Proceedings of IEEE 29th annual international conference engineering in medicine and biology society, pp 5719–5722
3. Dippold M (2006) Personal dead reckoning with accelerometers. In: Presented at the 3rd international forum applied wearable computing, Bremen, Germany
4. Eric F (2005) Pedestrian tracking with shoe-mounted inertial sensors. *Proc IEEE Comput Graph Appl* 25:38–46
5. Jimenez AR, Seco F, Prieto C, Guevara J (2009) A comparison of Pedestrian dead-reckoning algorithms using a low-cost MEMS IMU. In: Proceedings on IEEE international symposium on intelligent signal processing, pp 37–42
6. RaúlFeliz EZ, García-Bermejo JG (2009) Pedestrian tracking using inertial sensors. *J Phys Agents* 3:35–42

7. Sagawa K, Inooka H, Satoh Y (2000) Non-restricted measurement of walking distance. In: Proceedings of IEEE international conference on systems, man, and cybernetics, vol 3, pp 1847–1852
8. Xiaoping Y, Bachmann ER, Moore H, Calusdian J (2007) Self-contained position tracking of human movement using small inertial/magnetic sensor modules. In: Proceedings of IEEE international conference on robotics and automation, pp 2526–2533
9. Alvarez D, Gonzalez RC, Lopez A, Alvarez JC (2006) Comparison of step length estimators from wearable accelerometer devices. In: Proceedings of IEEE 28th annual international conference of engineering in medicine and biology society, pp 5964–5967
10. Lei F, Antsaklis PJ, Montestruque LA, McMickell MB, Lemmon M, Yashan S, Hui F, Koutroulis I, Haenggi M, Min X, Xiaojuan X (2005) Design of a wireless assisted pedestrian dead reckoning system—the NavMote experience. *IEEE Trans Instrum Meas* 54(6):2342–2358
11. Shin SH, Park CG, Hong HS, Lee JM (2005) MEMS-based personal navigator equipped on the user's body. In: Presented at the ION GNSS 18th International Tech. Meeting Satellite Division, Long Beach, CA, USA, 13–16 Sep 2005
12. Weinberg H (2002) Using the ADXL202 in pedometer and personal navigation applications. *Appl. Notes An-602*, Analog Devices, Inc., Norwood, MA, USA
13. Renaudin V, Demeule V, Ortiz M (2013) Adaptive pedestrian displacement estimation with a smartphone. In: International conference on indoor positioning and indoor navigation, vol 12, pp 916–924
14. Kamisaka D, Muramatsu S, Iwamoto T et al (2011) Design and implementation of Pedestrian dead reckoning system on a mobile phone. *IEICE Trans Inf Syst* E94-D(6):1137–1145
15. Zhou B, Li Q, Mao Q, Tu W, Zhang X (2014) Activity sequence-based indoor pedestrian localization using smartphones. *IEEE Trans Hum Mach Syst* 1–13
16. Zheng L, Zhou W et al (2015) A foot-mounted sensor based 3D indoor positioning approach. In: IEEE twelfth international symposium on autonomous decentralized systems (ISADS), pp 145–150
17. Zheng X, Yang H, Tang W et al (2014) Indoor Pedestrian navigation with shoe-mounted inertial sensors. *Multimedia and Ubiquitous Engineering*. Springer, Berlin, pp 67–73
18. Lan K-C, Shih W-Y (2014) Using smart-phones and floor plans for indoor location tracking. *IEEE Trans Hum Mach Syst* 44(2):211–221
19. Hunter J, Marshall R, McNair P (2004) Interaction of step length and step rate during sprint running. *Med Sci Sports Exerc* 36:261–271

A Variational Bayesian Approach for Unsupervised Clustering

Mu-Song Chen, Hsuan-Fu Wang, Chi-Pan Hwang,
Tze-Yee Ho and Chan-Hsiang Hung

Abstract Gaussian Mixture Models are among the most statistically mature methods which are used to make statistical inferences as well as performing unsupervised clustering. Formally, a gaussian mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the data set. In this paper, a probabilistic clustering based on the finite mixture models of the data distribution is suggested. An important issue in the finite mixture model-based clustering approach is to select the number of mixture components of clusters. In this sense, we focus on statistical inference for finite mixture models and illustrate how the variational Bayesian approach can be used to determine a suitable number of components in the case of a mixture of Gaussian distributions.

Keywords Gaussian mixture model · Clustering · Variational Bayesian

M.-S. Chen (✉)

Department of Electrical Engineering, Da-Yeh University, Changhua, Taiwan

e-mail: chenms@mail.dyu.edu.tw

H.-F. Wang

Department of Electrical Engineering and Energy Technology, Chung Chou

University of Science and Technology, Changhua, Taiwan

e-mail: rex.wang.sige@gmail.com

C.-P. Hwang · C.-H. Hung

Department of Electronic Engineering, National Changhua University of Education,

Changhua, Taiwan

e-mail: cphwang@cc.ncue.edu.tw

T.-Y. Ho

Department of Electrical Engineering, Feng Chia University, Taichung, Taiwan

© Springer Science+Business Media Singapore 2016

J.C. Hung et al. (eds.), *Frontier Computing*, Lecture Notes

in Electrical Engineering 375, DOI 10.1007/978-981-10-0539-8_63

1 Introduction

Due to the wide availability of huge amounts of data and the imminent need for exploring large data sets and for transforming such data into useful knowledge, data mining [1] has attracted a great deal of attention in the scientific researches and industry applications. Usually, data mining can be viewed as a result of natural evolution of information methodology. Clustering is one of the data mining techniques, which is used to classifier data elements into related groups without advance knowledge of group definitions and to discover the hidden structure of given samples. The Gaussian mixture model (GMM) [2] is a powerful method for data clustering. The mixture density is made up of a linear combination of multiple Gaussian component densities where each component corresponds to one cluster. Given a data set X with a sequence of independent, identically distributed samples $\{\mathbf{x}_n\}_{n=1,\dots,N}$, the goal is to model the probability density as a weighted mixture of a finite number of Gaussian components.

The complete generative model is specified by K Gaussian components, a set of K component probabilities (weights), and another set of latent variables z_{nk} , $n = 1, \dots, N$, $k = 1, \dots, K$, indicating which of these components was used to generate each sample \mathbf{x}_n . Basically, the GMM parameters are estimated from sample data using the iterative Expectation-Maximization (EM) algorithm [3] or Maximum a Posteriori (MAP) estimation [4] from a well-trained prior model. The difficulty appears because in general all those model variables are jointly dependent. As there are K^N different ways to assign N samples to K mixture components, the computation becomes tedious and impractical even for a medium size of N and K . Another obvious problem is the choice of correct number of components. Usually, deciding a proper number of components is often an ad hoc decision based on prior knowledge, assumptions, and practical experience. Certainly, an incorrect choice will invalidate the clustering process. An empirical way to find the best number of clusters is to try with different number of clusters or components and measure their validity.

Recently, variational Bayesian (VB) methods [5, 6] have gained popularity in the machine learning literature and have also been used to estimate the parameters of finite mixture models. The variational Bayesian method aims to construct a tight lower bound on the data marginal likelihood and then seeks to optimize this bound using an iterative scheme [7, 8]. Using variational methods, the order of mixture model can also be estimated. The VB approach can be set up for automatic complexity determination. If we assign prior parameter distributions such that they tend to favor a model with few mixture components, the learning procedure can in effect disregard some of the initially designated components. This can be viewed as a convenient automatic Occam's razor [9]. This is particularly useful because in practical applications it is usually impossible to know a priori how many model components are actually needed to accurately describe the data.

This rest of the paper is organized as follows. Section 2 introduces the Gaussian Mixture Model and the Expectation-Maximization algorithm to identify the parameters for the mixed model. Section 3 presents the variational Bayesian method

for the GMM, followed by simulation experiments to verify the validity of the proposed method in Sect. 4. Finally, conclusions are drawn in the last section.

2 Gaussian Mixture Model, GMM

A gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. The GMM is often used for data clustering. The posterior probabilities for each sample indicate that each data point has some probability of belonging to each cluster. Clusters are assigned by selecting the component that maximizes the posterior probability. Usually, a GMM can be expressed as a linear superposition of K Gaussians in the form as

$$p(\mathbf{x}_n|\Theta) = \sum_{k=1}^K \pi_k N(\mathbf{x}_n|\theta_k) \quad (1)$$

The Gaussian density $N(\mathbf{x}_n|\theta_k)$ is called a component of the mixture associated with a parameter set $\theta_k = \{\boldsymbol{\mu}_k, \Sigma_k\}$ where $\boldsymbol{\mu}_k$ and Σ_k are mean vector and covariance, respectively. π_k is the prior distribution (or simply the mixing coefficient) of the sample data \mathbf{x}_n from a data set $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$. π_k is also subject to the following constraints, i.e. $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^K \pi_k = 1$. Moreover, Θ contains the ensemble parameters of π_k and θ_k , e.g. $\Theta = \{\theta_k, \pi_k\}_{k=1, \dots, K}$.

By using a sufficient number of Gaussians and by adjusting their corresponding parameters θ_k as well as the mixing coefficients π_k , almost any continuous density function can be approximated to arbitrary accuracy. If \mathbf{x}_n is drawn independently from the distribution, then the standard GMM assumes that the density function at an observation \mathbf{x}_n is given by Eq. (1). In this sense, the joint conditional density of the data set X can be written as

$$p(X|\Theta) = \prod_{n=1}^N p(\mathbf{x}_n|\Theta) = \prod_{n=1}^N \left(\sum_{k=1}^K \pi_k N(\mathbf{x}_n|\theta_k) \right) \quad (2)$$

Given the joint conditional density, the log-likelihood function of the GMM is given by

$$\ln p(X|\Theta) = \sum_{n=1}^N \ln \left(\sum_{k=1}^K \pi_k N(\mathbf{x}_n|\theta_k) \right) \quad (3)$$

In order to maximize the log-likelihood function, the parameter set Θ needs to be properly determined. Usually, the maximum likelihood (ML) estimate [10], defined as $\Theta_{\text{ML}} = \arg \max(\ln p(X|\Theta))$, cannot be found analytically. The same is true for the

maximum a posteriori probability (MAP) estimate $\Theta_{\text{MAP}} = \arg \max(\ln p(X|\Theta) + \ln p(\Theta))$, given some prior $p(\Theta)$. In [11], the well-known expectation–maximization (EM) algorithm is used to approximate the maximum likelihood. Setting the derivatives of $\ln p(X|\Theta)$ with respect to the means μ_k and Σ_k , we obtain

$$\mu_k = \frac{\sum_{n=1}^N z_{nk} \mathbf{x}_n}{\sum_{n=1}^N z_{nk}} \text{ and } \Sigma_k = \frac{1}{\sum_{n=1}^N z_{nk}} \sum_{n=1}^N z_{nk} (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T \quad (4)$$

where z_{nk} denotes the homogeneity of samples within the class and is used to assign individuals \mathbf{x}_n to the class Ω_k . In other words, z_{nk} also denotes the probability of a sample measurement vector \mathbf{x}_n belonging to the k th component and

$$z_{nk} = \frac{\pi_k N(\mathbf{x}_n | \theta_k)}{\sum_{j=1}^J \pi_j N(\mathbf{x}_n | \theta_j)} \quad (5)$$

Finally, the maximization of Eq. (3) with respect to the mixing coefficients π_k is similar with the constraint that the sum of π_k is equal to one. Therefore,

$$\pi_k = \frac{1}{N} \sum_{n=1}^N z_{nk} \quad (6)$$

In summary, the EM algorithm can be separated into two steps.

- (a) E-step: Computes the conditional expectation of the log-likelihood function in Eq. (3). Since the log-likelihood is linear with respect to the missing z_{nk} , the conditional expectation is calculated by Eq. (5).
- (b) M-step: The prior, the sample mean vector, and the sample covariance matrix of each component are estimated from Eqs. (6) and (4), respectively.

The E- and M-steps alternate until the conditional expectation of the log-likelihood function reaches a local maximum.

From the aforementioned discussions, a critical issue in mixture modelling is the selection of the number of components. With too many components the model may over-fit the data, while a model with too few components may not be flexible enough to approximate the true underlying model. It is shown that the variational approximation method leads to an automatic choice of model complexity. This issue is addressed in the next section.

3 The Variational Bayesian (VB) Approach

Intuitively, a convenient way to control complexity in mixture models is by adjusting the values of the mixing coefficients π_k of Eq. (1). A component is thus removed if its mixing coefficient is nearly zero. Consequently, it is possible to consider a mixture model with a large number of components and maximize a suitable objective function

with respect to the mixing coefficients. The competition between components suggests a suitable approach for addressing the model selection problem. The variational approximation does provide a feasible approach to the estimation of the aforementioned model. The VB learning aims to maximize the lower bound of data marginal likelihood and therefore brings the approximate posterior as close as possible to the true posterior. The approach consists of converting the complex inferring problem into a set of simpler calculations, characterized by decoupling the degrees of freedom in the original problem. This decoupling is achieved at the cost of including additional parameters that must fit the given model and data.

Consider a probabilistic model with the observed variables by X and all of the hidden variables by Z . The joint distribution $p(X, Z | \Theta)$ is governed by the parameter set Θ . The goal is to maximize the likelihood function that is given by $p(X | \Theta) = \sum_Z p(X, Z | \Theta)$. It is straightforward to show that the log-likelihood of $p(X | \Theta)$ can be written as

$$\ln p(X | \Theta) = L(q, \Theta) + \text{KL}(q || p) \tag{7}$$

with

$$L(q, \Theta) = \int q(Z) \ln \frac{p(X, Z)}{q(Z)} dZ \tag{8}$$

and

$$\text{KL}(q || p) = - \int q(Z) \ln \frac{p(Z | X)}{q(Z)} dZ \tag{9}$$

where $q(Z)$ is any probability density function, which provides a close approximation to the true joint conditional density. $\text{KL}(q || p)$ is the Kullback-Leibler (KL) divergence between the true joint conditional density $p(Z | X)$ and the approximating density $q(Z)$. It should be noted that that $L(q, \Theta)$ contains the joint distribution of X and Z while $\text{KL}(q || p)$ contains the conditional distribution of Z given X . Because of the positivity of the KL divergence, maximizing the lower bound of Eq. (7) w.r.t $q(Z)$ corresponds to minimizing the KL divergence.

Suppose the elements of Z are partitioned into disjoint groups. Then the q distribution factorizes with respect to these groups as

$$q(Z) = \prod_{i=1}^K q_i(z_i) \tag{10}$$

Thus, we wish to find the $q(Z)$ of the form of Eq. (10) that maximizes the lower bound $L(q, \Theta)$. Using Eq. (10)

$$\begin{aligned}
 L(q, \Theta) &= \int \prod_i q_i \left(\ln p(X, Z) - \sum_i \ln q_i \right) dZ \\
 &= -\text{KL}(q_j \| \tilde{p}) - \sum_{i \neq j} \int q_i \ln q_i dz_i
 \end{aligned}
 \tag{11}$$

Clearly the bound in Eq. (11) is maximized when the KL divergence $\text{KL}(q_j \| \tilde{p})$ becomes zero, which is the case for $q_j(z_j) = \tilde{p}(X, z_j)$. In other words, the expression for the optimal distribution $q_j^*(z_j)$ is

$$\ln q_j^*(z_j) = \langle \ln p(X, Z) \rangle_{i \neq j} + \text{const}
 \tag{12}$$

In summary, the VB EM algorithm is given in steps:

- (a) VB E-Step: Evaluate $q(Z)$ to maximize $L(q, \Theta)$.
- (b) VB M-Step: Find $\Theta_{\text{VB}} = \arg \max L(q, \Theta)$.

In this study, we concentrate on the development of clustering problems based on Gaussian mixture models, which are fitted with the use of the recently developed variational Bayesian approach. Furthermore, the component responsible for generating the observed data point is discussed, which in turn decides the number of clusters in question. In this case, a prior on the mixing weights are treated as parameters, instead of the random variable. In this Bayesian GMM, Gaussian and Wishart priors are assumed for $\boldsymbol{\pi}$ and \mathbf{T} respectively. Thus,

$$p(\boldsymbol{\mu}) = \prod_{k=1}^K N(\boldsymbol{\mu}_k | \beta \mathbf{I})
 \tag{13}$$

$$p(\mathbf{T}) = \prod_{k=1}^K W(\mathbf{T}_k | \nu, \mathbf{V})
 \tag{14}$$

where the scalar ν denotes the degrees of freedom, \mathbf{V} is the scale matrix, and the expected value of \mathbf{T}_k is $\langle \mathbf{T}_k \rangle = \nu \mathbf{V}^{-1}$. It should be noted that $\mathbf{T} = \{\mathbf{T}_k\}$ is the precision (i.e., inverse covariance) matrix of the mixture component. To achieve the optimal number of components, we can maximize the marginal likelihood $p(X, \boldsymbol{\pi})$ obtained by integrating out the latent variables $\hat{Z} = \{Z, \boldsymbol{\mu}, \mathbf{T}\}$, i.e.

$$p(X, \boldsymbol{\pi}) = \int p(X, \boldsymbol{\pi}, \hat{Z}) d\hat{Z}
 \tag{15}$$

with respect to the mixing weights $\boldsymbol{\pi}$ that are treated as parameters. The variational approximation suggests the maximization of a lower bound of the logarithmic marginal likelihood

$$L(q, \boldsymbol{\pi}) = \int q(\hat{Z}) \ln \frac{p(X, \boldsymbol{\pi}, \hat{Z})}{q(\hat{Z})} d\hat{Z} \leq p(X, \boldsymbol{\pi}) \quad (16)$$

where $q(\hat{Z})$ is an arbitrary distribution approximating the posterior $p(\hat{Z}|X)$. The maximization of L is performed in an iterative way using the variational EM algorithm. At each iteration, two steps take place. Therefore, maximization of the bound with respect to q is taken and maximization of the bound with respect to $\boldsymbol{\pi}$ is then followed. A notable property is that during maximization of L , if some of the components fall in the same region in the data space, then there is strong tendency in the model to eliminate the redundant components. Consequently, the competition between mixture components suggests an approach for addressing the model selection problem.

To implement the maximization with respect to q , the mean-field approximation has been adopted that assumes q to be a product of the form

$$q(\hat{Z}) = q_z(Z)q_\mu(\boldsymbol{\mu})q_T(\mathbf{T}) \quad (17)$$

After performing the necessary calculations in Eq. (12), the results are the following set of densities

$$q_z(Z) = \prod_{n=1}^N \prod_{k=1}^K r_{nk}^{z_{nk}} \quad (18)$$

$$q_\mu(\boldsymbol{\mu}) = \prod_{k=1}^K N(\boldsymbol{\mu}_k | \mathbf{m}_k, \mathbf{S}_k) \quad (19)$$

$$q_T(\mathbf{T}) = \prod_{k=1}^K W(\mathbf{T}_k | \eta_k, \mathbf{U}_k) \quad (20)$$

where the parameters of the densities can be computed as

$$r_{nk} = \frac{\hat{r}_{nk}}{\sum_{j=1}^K \hat{r}_{nj}} \quad (21)$$

$$\hat{r}_{nk} = \pi_k \exp\left(\frac{1}{2} \ln |\mathbf{T}_k| - \frac{1}{2} \text{tr}\left\{\langle \mathbf{T}_k \rangle (\mathbf{x}_n \mathbf{x}_n^T - \mathbf{x}_n \langle \mathbf{u}_k \rangle^T - \langle \mathbf{u}_k \rangle \mathbf{x}_n^T + \langle \mathbf{u}_k \mathbf{u}_k^T \rangle)\right\}\right) \quad (22)$$

$$\mathbf{m}_k = \mathbf{S}_k^{-1} \langle \mathbf{T}_k \rangle \sum_{n=1}^N \langle z_{nk} \rangle \mathbf{x}_n \quad (23)$$

$$\mathbf{S}_k = \beta \mathbf{I} + \langle \mathbf{T}_k \rangle \sum_{n=1}^N \langle z_{nk} \rangle \quad (24)$$

$$\eta_k = v + \sum_{n=1}^N \langle z_{nk} \rangle \tag{25}$$

$$\mathbf{U}_k = \mathbf{V} + \sum_{n=1}^N \langle z_{nk} \rangle (\mathbf{x}_n \mathbf{x}_n^T - \mathbf{x}_n \langle \mathbf{u}_k \rangle^T - \langle \mathbf{u}_k \rangle \mathbf{x}_n^T + \langle \mathbf{u}_k \mathbf{u}_k^T \rangle) \tag{25}$$

After the maximization of L with respect to q , the second step of each iteration of the training method requires maximization of L with respect to $\boldsymbol{\pi}$, leading to the following simple update equation for the variational M-step

$$\pi_k = \frac{\sum_{n=1}^N r_{nk}}{\sum_{j=1}^K \sum_{n=1}^N r_{nj}} \tag{26}$$

The above variational EM update equations are applied iteratively and converge to a local maximum of the variational bound. During the optimization some of the mixing coefficients converge to zero thus the corresponding components are eliminated from the mixture. In this way complexity control can be achieved.

4 Simulation Results

To verify the VB approach for clustering task, a toy example is evaluated. The given set contains three Gaussians to generate 1,000 samples, without knowing their class labels. As shown in Fig. 1, two of the clusters are partially overlapped.

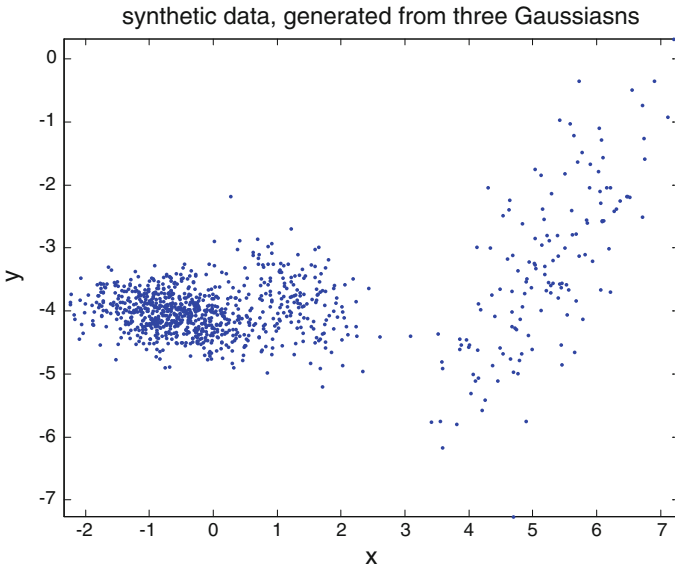


Fig. 1 An illustrative example for unsupervised clustering. Three Gaussians are used for generating 1000 samples

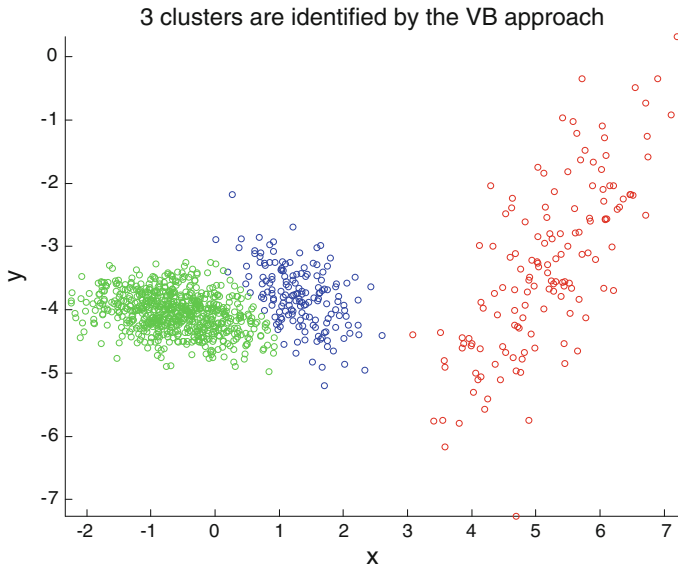


Fig. 2 Three clusters with different colors are identified for unsupervised classification

The VB approach is then taken. It turns out that if one starts with a large number of components, superfluous components are removed as the method converges to a solution, thereby leading to an automatic choice of model complexity. Simulation results in Fig. 2 also suggest that the VB method is able to recover the correct number of components. Finally, the log-likelihood function in Fig. 3 can converge in a finite number of iterations in this experiment.

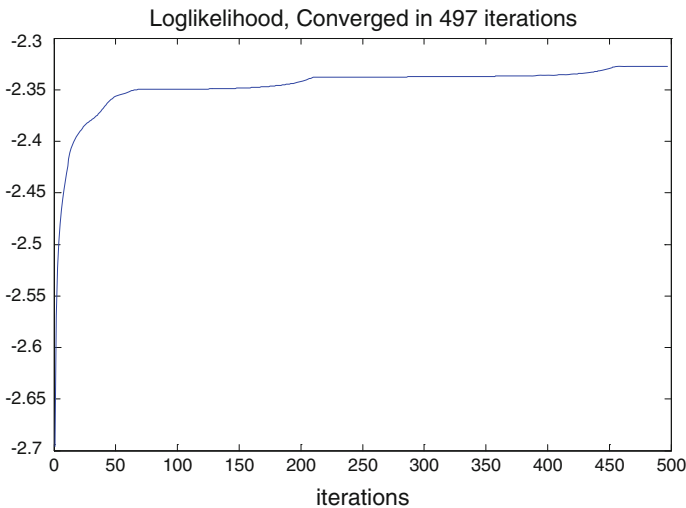


Fig. 3 The convergence of the log-likelihood function

5 Conclusion

One of the key issues in the application of mixture models is the determination of a suitable number of mixture components. Conventional approaches based on cross-validation are computationally expensive. In this paper, we applied the variational techniques to address this issue. Simulation results indicate that this approach is able to recover the appropriate number of components in a synthetic data problem and that it also possibly presents a useful and practical methodology to density estimation in real world data sets.

Acknowledgment This research was supported by the Ministry of Science and Technology under contract numbers MOST 103-2221-E-212-011 and MOST 104-2221-E-212-011.

References

1. Hand D, Mannila H, Smyth P (2001) Principles of data mining. MIT Press, Cambridge
2. Yu Guoshen (2012) Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. *IEEE Trans Image Process* 21(5):2481–2499
3. Yin J, Zhang Y, Gao L (2012) Accelerating expectation-maximization algorithms with frequent updates. In: Proceedings of the IEEE international conference on cluster computing
4. Murphy KP (2012) Machine learning: a probabilistic perspective. MIT Press, Cambridge, pp 151–152
5. Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin
6. Fox C, Roberts S (2012) A tutorial on variational Bayes. *Artif Intell Rev* 38(2):85–95
7. Takekawa Takashi, Fukai Tomoki (2009) A novel view of the variational Bayesian clustering. *Neuro Comput* 72:3366–3369
8. Attias H (2000) A variational Bayesian framework for graphical models. In: Leen T et al (eds) Advances in neural information processing systems, vol 12. MIT Press, Cambridge
9. Mackay et al (2003) Model comparison and Occam’s Razor. Cambridge University Press, Cambridge
10. Sijbers J, den Dekker AJ (2004) Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magn Reson Med* 51(3):586–594
11. Roche A, Ribes D, Bach-Cuadra M, Kruger G (2011) On the convergence of EM-like algorithms for image segmentation using Markov random fields. *Med Image Anal* 15:830–839

Virtualized Multimedia Environment for Shoulder Pain Rehabilitation

Chih-Chen Chen, Hsuan-Fu Wang, Shih-Chuan Wang,
Chih-Hong Chou, Heng-Chih Hsiao and Yu-Luen Chen

Abstract Whether a person's upper limbs healthy or not will seriously impact his/her daily life. For those who have healthy, normal function of upper limbs or those who suffered from impairment but in the process of rehabilitation training, it is very important to remain appropriate, reasonable and moderate exercise to keep upper limbs functioning normally or in gradual recovery. This study use Kinect somatosensory device with Unity software to develop 3D situational games such as: "Shoulder finger ladder" and "Single curved shoulder". The collected data from this training process can be uploaded via the internet to the cloud or server for participants to do self-inspection or it can be a reference for medical staffs to assess training effectiveness for those with impairments and planning in rehabilitation courses. In order to have more effective ways, researchers have imported games and virtual reality training effect to help those participants to train their upper limbs in a relaxed and extricated environment. In Shoulder finger ladder and Single curved shoulder training activities, the results of 8 subjects with normal upper limbs function represented that the system has good stability and reproducibility. And it

C.-C. Chen (✉) · H.-C. Hsiao
Department of Management Information System, Hwa Hsia University of Technology,
Taipei, Taiwan
e-mail: weib120@gmail.com

H.-F. Wang
Department of Electrical Engineering and Energy Technology,
Chun Chou University of Science and Technology, Yuanlin, Taiwan
e-mail: rex.wang.sige@gmail.com

S.-C. Wang
Computer and Networking Center, Hwa Hsia University of Technology,
Taipei, Taiwan

C.-H. Chou
Department of Electronic Engineering, National Taipei University of Technology,
Taipei, Taiwan

Y.-L. Chen
Department of Digital Technology Design, National Taipei University of Education,
Taipei, Taiwan
e-mail: allen001212@gmail.com

also showed that motion of dominant side is more flexible than the non-dominant side. Flexibility and responsiveness in the elders are slightly behind the young. Another six weeks of training were held for subjects with frozen shoulder combined with Shoulder finger ladder and Single curved shoulder games. It showed that on the 3rd week, the average performances were stable. The T-test average score from 1–2 week and 3–4 weeks/5–6 weeks showed significant difference ($P < 0.05$).

Keywords Kinect somatosensory device · Rehabilitation · Virtual reality

1 Introduction

It is very important to have healthy upper extremity function. Therefore, moderate exercise or training and appropriate maintenance are critical. And for the impaired limbs, it is even more crucial to do reasonable rehabilitation training to restore its normal daily function.

Generally speaking, upper limbs dysfunctional patients need to be trained repeatedly through rehabilitation equipments in order to recover. Clinically, Exercise skate of arm, Exercise skate of hand, Vertical tower, Incline board, Stacking cones, Cura motion exercise are some of the most frequent use equipments for it. There are many therapeutic ways such as, mechanical arms (passive or positive patient training through mechanical structures) [1–9], video games (follow the instructions on the screen to move mechanical arms to help neural rehabilitation) [10], and virtual reality (integrate and improve sound, video, graphics and text to make users feel that they are experiencing it for real) ... etc. [11–14].

There was a positive meaning for the researchers to introduce the concepts of virtual reality to train upper limbs or change the traditional rehabilitation due to the fact that it increases the enthusiasm and repeatability of the participants. The main goal of this research was to train upper limbs. In order to facilitate the participants to use in daily life, Microsoft Kinect somatosensory devices (for Windows) was applied as 3D human motion capture system. It could detect human skeleton coordinates such as palms, wrists and both side of shoulders to develop the Unity games. Participants could be trained through those games and scenes on the screen without actual touching the real entity. This research was mission-oriented training which applied Kinect somatosensory devices software development with Unity 3D games to enhance the training effects. It was a very convenient and safe way to do that participants only need to touch assigned objects through upper extremity.

Unity 3D is a low price, powerful and intuitive game engine which is applied widely in current industry. Even though it can be used to develop games, it would not support somatosensory part. Therefore, scenarios are played through Kinect. It is designed to detect human skeleton information which is the key point to develop somatosensory games. The signals are captured and transmitted to Unity 3D through Microsoft SDK (Kinect for Windows SDK) or Open NI(open nature interaction) to drive the actions of the game characters.

2 Materials and Methods

The basic structures of this study, as shown in this study Fig. 1., can be generally divided into 3 parts—hardware interface, development interface and application interface. In (1) hardware interface part: It used Microsoft’s product, hardware of Kinect for Windows to connect to the computer; the advantage of Kinect is that its skeleton recognition technology can be directly used to determine the actions while other relevant information of actions in the general attitude captures Technology need to be captured by having physical installation of many sensing elements and cables. (2) development interface and application program parts: It is mainly composed by the installation of drivers of Kinect for Windows SDK and Unity3D software. Common programming language C# between these two was used to write a program and the game application was produced by the compilation and function calls to Kinect SDK through Unity3D. It could be used to measure data while the game was running, as well as recording 3D coordinates of the skeleton of the participants during the training process.

Shown in Fig. 2 was the overall schematic diagram of this study. The human skeleton in actions was displayed on the PC by the use of Windows SDK, Unity 3D software tool, and Kinect’s sensing device. Game activities were built by adopting Unity3D software; the reason was that Kinect was a device for sensing the human skeleton data; the 3D coordinate data that Kinect captured must be able to be projected to the corresponding Unity3D’s virtual scenes on the PC screen. The

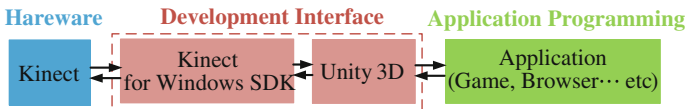


Fig. 1 Basic structures

Fig. 2 The overall scheme

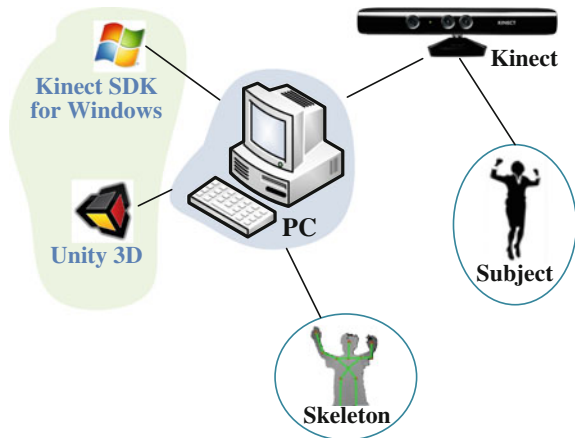
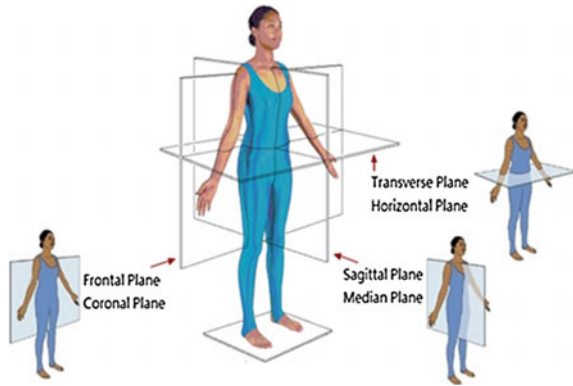


Fig. 3 Orientation of the human body

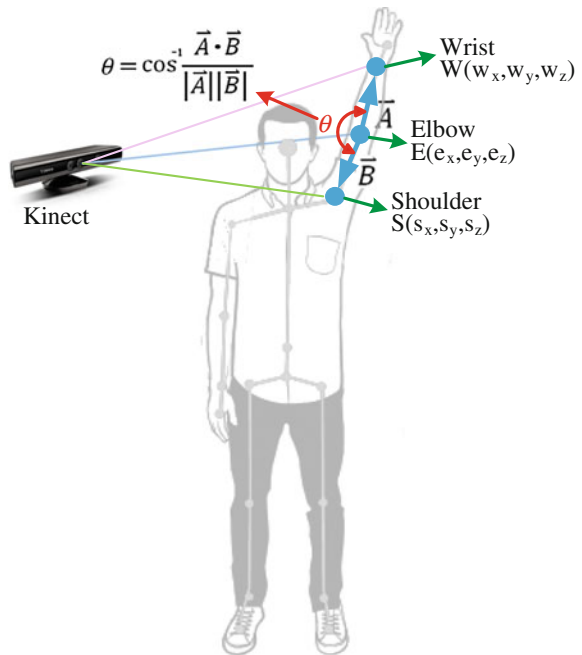


virtual scenes could also be used to construct and plan game scenes or express different collision effects of design.

Shown in Fig. 3 was the Orientation of the Human Body, which can be divided into three sport planes for reference—Sagittal Plane, Frontal Plane and Transverse plane.

To capture human skeleton coordinates through importing Kinect device, the angles could be surmised by the bones that connect to those joints and the angles differences in movements, which helped medical workers understand the accuracy of poses and actions changes for patients in the process of training activities. Shown in Fig. 4 was the collected data of Shoulder, Elbow, and Wrist joint coordinates, which were $S(s_x, s_y, s_z)$, $E(e_x, e_y, e_z)$, $W(w_x, w_y, w_z)$

Fig. 4 Computing angle of elbow



wherein

$$\vec{A} = \vec{ES}, \quad \vec{B} = \vec{EW}$$

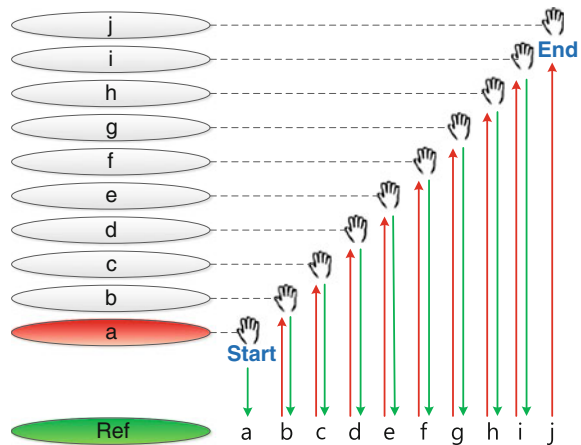
The elbow angle formula was

$$\theta = \cos^{-1} \frac{\vec{A} \cdot \vec{B}}{|\vec{A}| |\vec{B}|}$$

Shown in Fig. 5 was the training activities in the Frontal Plane—the design of “shoulder finger ladder”, which can be used as training for upper extremity lifting or measurement purposes. The green elliptical area at the bottom is Reference(Ref), the remaining elliptical areas were reminder for participants to touch virtual position (represented in letter from a to j) in sequence (height) by lifting their upper limbs, where the red ellipse area (figure marked as a) was used for participants to touch according to the Ref (the very bottom of the green elliptical area) upon completion of the designated order of touch (shown as a step) when virtually touching by hands, at some point to the height where it should be. At the same time, the red elliptical area would automatically move on the screen on one order to show the completion of one round, as soon as the participants completed a–j sequential order (or reaching individual’s maximum operating limit). The actual orders (Fig. 5. shown is 10 orders) could be adjusted based on design needs.

Shown in Fig. 6 was the training activities in the Frontal Plane—the planning design of the Single curved shoulder. Subject sequentially touched the objects virtually from the left side of the hand toward the right, complete in clockwise and then in the counterclockwise direction to the original starting point. The main training of this activity included the shoulder, elbow, forearm, and it was also very helpful for hand-eye coordination and reaction cultivation.

Fig. 5 Design of shoulder finger ladder motion trajectory (take *right hand* as an example)



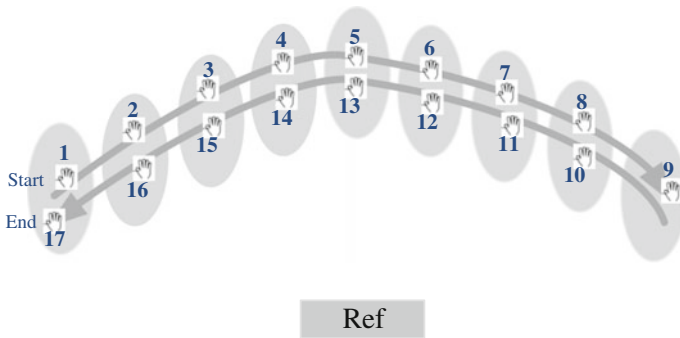


Fig. 6 Design of single curved shoulder motion trajectory (take right hand as an example)

To confirm the reproducibility and stability of the system development, it was very helpful to understand whether this system was stable enough for clinical training activities by assessing and analyzing the test results of participants with different ages and body types or same participant with different test time after taking the tests [15, 16].

3 Results

Eight subjects (age range were 21–30, 31–40, 41–50, 51–60, 2 subjects in each group) with normal upper limbs function were chosen to do “Shoulder finger ladder” and “Single curved shoulder” in 6 weeks (twice a week, 3 rounds a time) to test reproducibility and stability.

Figure 7 was the performance of subject A-1 (left handed) in 6 weeks. It was observed that the average using time of left hand was less than the right and it reached stability (21.1 s) in first trial of the third week (wk3-1). However, it reached stability in the fifth week (~23.2 s) in right hand. The total average using time of

Fig. 7 Subject A-1 shoulder finger ladder test performance in six weeks

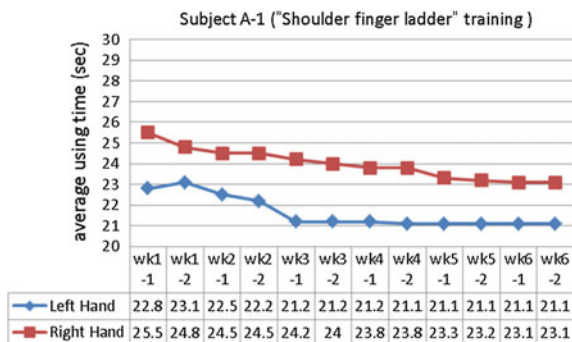
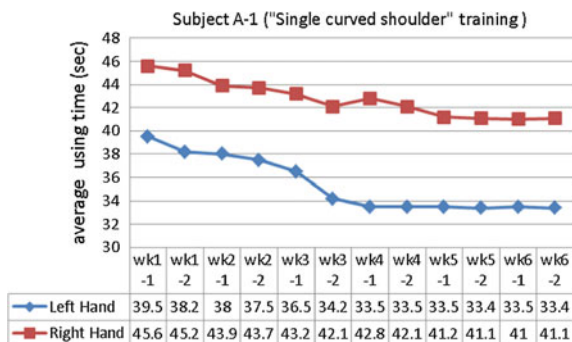


Fig. 8 The performance of subject A-1 (left handed) in 6 weeks



left hand was 21.64 s, standard deviation was 0.77. The total average using time of right hand was 23.98 s, standard deviation was 0.75.

Shown in Fig. 8 was the performance of subject A-1 (left handed) in 6 weeks. Left (right) hand average using time reached stability ~33.5 s (~41.1 s) in the fourth (fifth) week. The total average using time of left (right) hand was 35.39 (42.75) s, standard deviation was 2.35 (1.60). The results also indicated that the movement of the dominant side is more flexible than the non-dominant side.

Shown in Table 1 was the performance record of eight (four age groups) subjects with normal upper limbs function. The results indicated that the dominant side of average using time of the subjects in the same age group were close in Shoulder finger ladder and Single curved shoulder which showed good stability of the system. The average using time of the subjects' both hands indicated that the movement of the dominant side is more flexible than the other side. Moreover, the elders in terms of flexibility and ability to response were slightly inferior to the youngs.

Table 1 The performance record of eight subjects with normal upper limbs function

Year group	Subject/sex dominant side	Average using time (s)			
		Shoulder finger ladder		Single curved shoulder	
		Left hand	Right hand	Left hand	Right hand
21–30	A-1/σ/left hand	21.64 ± 0.77	23.98 ± 0.75	35.39 ± 2.35	42.75 ± 2.60
	A-2/♀/right hand	23.89 ± 0.78	22.38 ± 0.73	43.02 ± 2.31	35.12 ± 2.15
31–40	B-1/♀/left hand	23.35 ± 0.78	25.12 ± 0.86	38.12 ± 2.18	43.98 ± 2.71
	B-2/σ/right hand	25.38 ± 0.92	23.12 ± 0.75	44.57 ± 2.82	37.35 ± 2.32
41–50	C-1/σ/left hand	25.10 ± 0.83	28.58 ± 0.92	40.21 ± 2.56	48.11 ± 2.62
	C-2/♀/right hand	29.11 ± 0.96	25.32 ± 0.91	50.05 ± 2.78	41.35 ± 2.72
51–60	D-1/σ/left hand	27.02 ± 1.02	32.88 ± 1.56	45.15 ± 3.32	49.38 ± 2.95
	D-2/♀/right hand	32.15 ± 1.76	26.95 ± 0.98	51.15 ± 2.86	43.58 ± 3.27

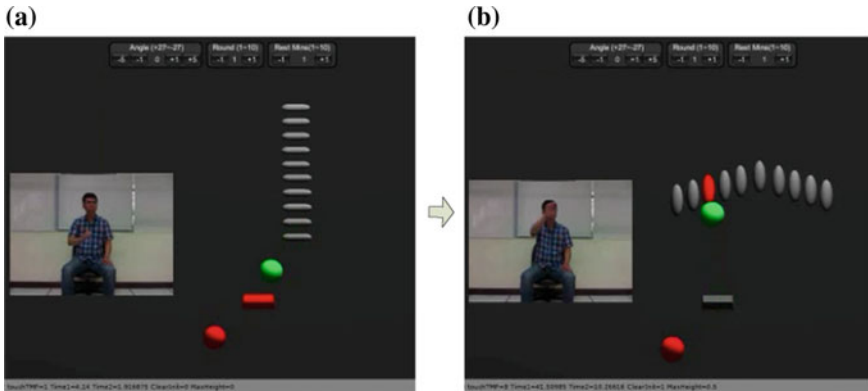


Fig. 9 The training screen of the subjects with impaired upper limbs. **a** Shoulder finger ladder. **b** Single curved shoulder

Shown in Fig. 9 was the training screen of the subjects with impaired upper limbs. Subjects performed Shoulder finger ladder test first to determine the angle movement range of the shoulders. And then, Single curved shoulder test was performed based on the test results of shoulder finger ladder. The total training period was 6 weeks (twice a week), three rounds in each practice, three minutes rest between each round. Subjects would go through two tests after complete one practice.

Shown in Fig. 10 was the test performance of subject P with frozen shoulder in clinical trial. In the beginning of Shoulder finger ladder test, subject P reached the fifth ladder, the total using time was 37.23 s. Followed by Single curved shoulder, the time was 75.38 s. In the following 6 weeks tests, the test results were based on the comparison of the average using time of the previous two tests. In the fourth week, it reached stability: shoulder finger ladder was 27.64 ± 0.493 s and the Single curved shoulder was 47.24 ± 0.461 s.

Shown in Table 2 was the performance of subject P during the entire test period of Shoulder finger ladder (average using time). It was presented in bi-week interval.

Fig. 10 The test performance of subject P with frozen shoulder in clinical trial

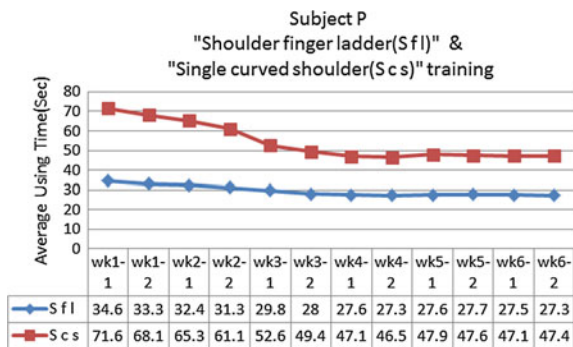


Table 2 The performance of subject P during the entire test period of shoulder finger ladder

		Week 1–2	Week 3–4	Week 5–6
Average using time (s)	Average 1	34.55	29.78	27.56
	Average 2	33.25	28.01	27.73
	Average 3	32.43	28.58	27.46
	Average 4	31.25	27.26	27.26

Table 3 The results of statistical test (pairwise) (shoulder finger ladder)

	Week 1–2/week 3–4	Week 3–4/week 5–6	Week 1–2/week 5–6
Average value	32.87/28.41	28.41/27.50	32.87/27.50
Standard deviation	1.92/1.12	1.12/0.03	1.92/0.03
P(T ≤ t)	0.0022	0.1449	0.0003
Significant difference	Significant (<0.05)	No significant	Significant (<0.05)

To assess the average using time performance of subject P in the six week test trial, the benchmark was the average score of tests in four times biweekly. The statistical test was performed pairwise. The results were shown in Table 3. T test were 0.0022, 0.1449, 0.0003 in bi-week intervals. The first and second week was adjustment stage for subject P; therefore, the performance was not ideal. There was significant improvement in the third-fourth and the fifth-sixth week. Moreover, the T-test results of 1–2 week/3–4 week and 1–2 week/5–6 week were significant difference. The average performance reached stability after the third week and there was no significant difference of the T-test result in 3–4 week/5–6 week. The results have indicated that this system was effective on the training of subject P.

4 Discussion

Shoulder joint has the largest range, the most complicated action forms and is the most frequently use part in physical activities which results in the higher frequency of injury. People should maintain it in their daily life. It may lead to chronic degradation if don't take care of it well. This study use Kinect somatosensory device with Unity software to develop 3D situational games for upper extremity training activities. Taking games as training methods help to improve concentration, interests of participation and temporarily forget about their body discomfort. In this study, the equipments are inexpensive, easy to obtain, and the system is easy to install. People can do simple self-training both at home or in the office. We will continue to recruit the cases with impaired upper limbs function to do related research and develop suitable rehabilitation training games. In order to have more effective ways, researchers have imported games and virtual reality training effect to

help those participants to train their upper limbs in a relaxed and extricated environment.

Acknowledgment This work was supported by the Ministry of Science and Technology, TAIWAN, 103-2221-E-146-002-, and 103-2221-E-152-002-.

References

1. Krebs HI, Hogon N, Aisen ML, Volpe BT (1998) Robot-aided neuro rehabilitation. *IEEE Trans Rehabil Eng* 6(1):75–87
2. Fasoli SE et al (2003) Effects of robotic therapy on motor impairment and recovery in chronic stroke. *Arch Phys Med Rehabil* 84(4):477–482
3. Cozens JA (1999) Robotic assistance of an active upper limb exercise in neurologically impaired patients. *IEEE Trans Rehabil Eng* 7(2):254–256
4. Coote S et al (2008) The effect of the GENTLE/s robot-mediated therapy system on arm function after stroke. *Clin Rehabil* 22(5):395–405
5. Lo AC et al (2010) Robot-assisted therapy for long-term upper-limb impairment after stroke. *N Engl J Med* 362(19):1772–1783
6. Frisoli A et al (2012) A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks. *IEEE Trans Syst Man Cybern Part C: Appl Rev* 42(6):1169–1179
7. Cozens JA (1999) Robotic assistance of an active upper limb exercise in neurologically impaired patients. *IEEE Trans Rehabil Eng* 7(2):254–256
8. Coote S et al (2008) The effect of the GENTLE/s robot-mediated therapy system on arm function after stroke. *Clin Rehabil* 22(5):395–405
9. Lo AC et al (2010) Robot-assisted therapy for long-term upper-limb impairment after stroke. *N Engl J Med* 362(19):1772–1783
10. Szturm T, Peters JF, Otto C, Kapadia N, Desai A (2008) Task-specific rehabilitation of finger-hand function using interactive computer gaming. *Arch Phys Med Rehabil* 89
11. Chen C-C, Chen W-L, Chen B-N, Shih Y-Y, Laie J-S, Chen Y-L (2014) Low-cost computer mouse for the elderly or disabled in Taiwan. *Technol Health Care* 22:137–145
12. Chen C-C, Hong D-J, Chen S-C, Shih Y-Y, Chen Y-L (2013) Study of multimedia technology in posture training for the elderly. In: 7th international conference on bioinformatics and biomedical engineering (iCBBE)
13. Chen C-C, Chang K-T, Chou C-H, Chen Y-L (2014) Virtual reality rehabilitation system for shoulder movement disorders. In: The annual conference on engineering and technology (ACEAT)
14. Ciou SH, Chou CH, Hwang YS, Chen CC, Chen SC, Chen YL (2012) Development of an interactive game based assessment and training system—for the stroke. In: 6th international conference on bioinformatics and biomedical engineering (iCBBE)
15. Cook AM, Hussey SM (2007) Assistive technologies, 2nd edn. Principles and practice
16. American Association on Mental Retardation (2007) Mental retardation: definition, classification, and systems of supports, 10th edn

Multimedia Technology with Tracking Function for Hand Rehabilitation

Ying-Ying Shih, Yen-Chen Li, Chih-Chen Chen, Hsuan-Fu Wang,
Shih-Wei Chou, Sung-Pin Hsu and Yu-Luen Chen

Abstract In the modern busy work environment, the limb functional damage caused by career injury, sport injury, accident, and illness emerge in endlessly, which makes the rehabilitation training becoming one of the important projects in medical service. The upper limbs are the most frequently used part in the daily activities, so it is important to maintain the sound function of upper limbs; the function of upper limbs could perform normally by tempering the use or exercise of upper limbs, and making proper maintenance. The limb function injured person shall need proper reasonable rehabilitation training, in the expectation of recovering the normal function, to avoid influencing the future daily life. This paper takes the upper limbs rehabilitation training system development as the research topic, with the hand gliding cart of barcode scanner with wireless transmission function, aiming at the several barcode arrangement types, via the participator push the hand

Y.-Y. Shih · Y.-C. Li
Department of Physical Medicine and Rehabilitation,
Chang Gung Memorial Hospital, Taipei, Taiwan

C.-C. Chen (✉)
Department of Management Information System,
Hwa Hsia University of Technology, Taipei, Taiwan
e-mail: weib120@gmail.com

H.-F. Wang
Department of Electrical Engineering and Energy Technology,
Chun Chou University of Science and Technology, Yuanlin, Taiwan
e-mail: rex.wang.sige@gmail.com

S.-W. Chou
Department of Mathematics, National Central University,
Jhongli City, Taiwan

S.-P. Hsu
Department of Digital Computer Science College of Science,
National Taipei University of Education, Taipei, Taiwan

Y.-L. Chen
Department of Digital Technology Design, National Taipei University
of Education, Taipei, Taiwan
e-mail: allen001212@gmail.com

gliding cart to detect the move location, motion time, and whether within the preset track in the scanning way, to record the quantifying information. Through the analysis and integration of software program in the host, its result could be provided for the medical personnel as the key reference of training effect of the participator.

Keywords Wireless transmission · Hand gliding cart · Barcode · Scanner · Rehabilitation

1 Introduction

In the sport, the muscle may be strained for the hyperextension or sudden twist, or be sprained for the sudden wrap or pull of ligament and tissue around the joint, which are all the common sport injury. The rehabilitation method of physiotherapy after sport injury is to accelerate the injure healing and body function recovery, to prevent the joint stiffness and muscle atrophy, with the expectation of body maintaining the good function and status; the adopted methods include (1) passive, active or resistant medical treatment method of action, (2) hand power therapy of massage and joint motion, (3) traction treatment applying the acting force and counter-acting force theory in the applied mechanics, to relieve the spasm through traction, and reduce the neurothlipsis. The sport injury may have negative influence, including the interference to the fitness, study and work of the injured, shortening the sportive life, and even causing the physical disability or death.

This research integrates the hand gliding cart and barcode scanner, matching with different barcodes in the permutation and combination. The patients shall do the upper limbs rehabilitation motion according to the instruction, including the transverse and longitudinal linear movement, polygonal zone linear movement, \bigcirc and ∞ curves with curvature change, which all need the complicated action combination of many parts of upper limbs, such as hand, wrist, elbow, and shoulder. The barcode information acquired by scanning in the action process shall be saved for follow-up statistical analysis, whose result would show the position of patients in the action process, track change condition and using time. The established quantitative index could be the reference for the terapeuta planning the relevant training courses, besides to let the patients understand whether their upper limbs movements realize the preset time and track route.

2 Materials and Methods

The purpose of this study is to assist patients with impaired upper limb function rehabilitation. Patients with injured hand handshake cart, then follow the path specified for the passage of the pulley, upper extremity training activities carried out.

2.1 Literature Review

Cloud Y.-S. Hwang etc. study point out, a simple approach to rehabilitation is taken to help patients recover upper limb function. The equipment used comprises a panel of magnetic sensor components, a computer server, and a hand gliding cart. When the patient moves the hand-gliding cart across on the panel, the screen will display automatically the route taken by the gliding cart, and the route spots during the course of the movement will be recorded for quantification. This auxiliary device, developed in this study for rehabilitation training, can actually help patients stretch their impaired upper limbs, prevent continual upper limb atrophy, and reduce the negative impact from the disability on daily life and activities [1].

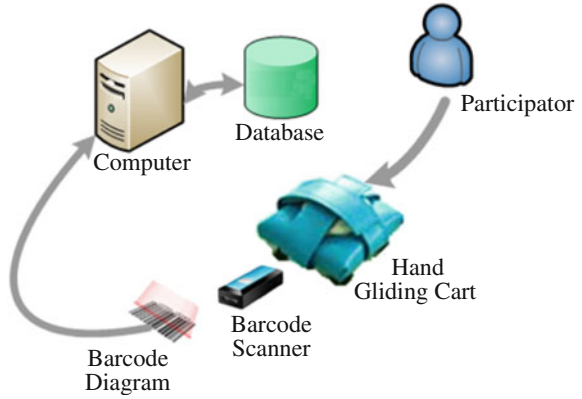
C.-C. Chen etc. study point out, this research involves developing a hand-skate training system for the upper extremity rehabilitation. Included in this system are a hand-skating board, radio frequency identification (RFID) reader, and a computer. The hand-skating board is a platform with multiple RFID-tags and a RFID reader under the hand-skate for the patients to operate “∞” figure. The flexion, extension, and coordination of hand muscle will be achieved after repeatedly performing this training. Meanwhile, in addition to multiple training items of corresponding figures, this system also provides functions for building a patient database to offer clinicians scientific data for long-term monitoring the rehabilitation effects [2].

The research of Mr. Liu is to put the Wii Remote in the place 1 m above the center of test desk, and place the infrared emitter in the center above the hand gliding cart, while the movement track to be followed by the testee is shown in the computer screen. The testee shall use the hand gliding cart to move on the table and imitate the track, to achieve the purpose of practically stretching the upper limbs of the tested [3].

2.2 System Planning

Figure 1 shows the overall schematic diagram of the system, constituted by the hand gliding cart, barcode scanner (placed below the hand gliding cart), barcode diagram, personal computer and database; using the combination of hand gliding cart and barcode scanner, matching with different barcode arrangements, detecting the action conditions as hand gliding cart movement position, action time, and whether in the preset track through scanning, and making quantitative record, then after the analysis and integration of software in the host, its result could be the reference of participator rehabilitation condition for the medical rehabilitation personnel.

Fig. 1 Overall schematic diagram



2.2.1 1D Barcode

This Research adopts the Code 39 barcode, and makes it into the bcoCode mode showing in Fig. 2a. The actual placing principle of barcode on the panel is as shown in Fig. 2b, that the area of each barcode is 1 cm * 1 cm, and the space between each layer is 0.3 cm. It shall permute and combine the barcodes within the scope of 100 cm * 80 cm, to form the linear action (longitudinal, transverse) barcode shown in Fig. 2c, or ∞ action barcode shown in Fig. 2d is one such emerging paradigm which makes use of contemporary virtual machine technology.

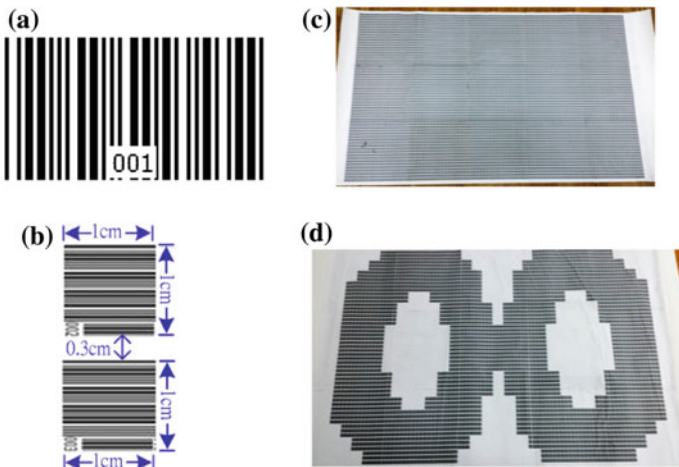


Fig. 2 Code 39 barcode **a** bcoCode mode, **b** barcode on the panel, **c** linear action barcode and **d** ∞ action barcode

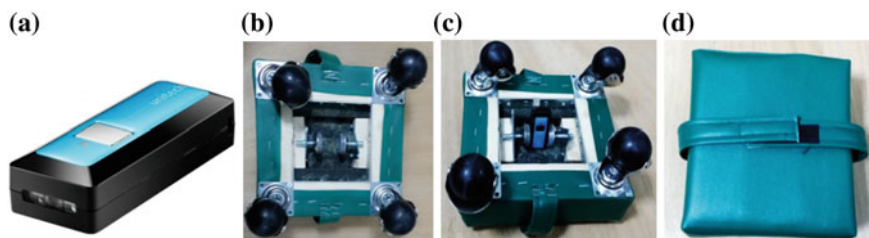


Fig. 3 Scanner and hand gliding cart. **a** Barcode Scanner. **b** Cart bottom. **c** Installation barcode scanner. **d** Cart front

2.2.2 Barcode Scanner and Hand Gliding Cart

Adopting the sensing element of Visible Red LED (wavelength: 625 nm) Unitech MS910 Wireless Barcode Scanner (as shown in Fig. 3a) for the barcode scanning, its scanning distance is 30–185 mm, and scanning rate is 240 scans/second. This Scanner after setting could scan the barcode constantly, and its information data is transmitted to the computer through the Bluetooth wireless transmission method. Figure 3b–d show the bottom of self-designed hand gliding cart, which are the material photo of the cart bottom with barcode scanner and the cart front.

2.2.3 Software Development

Under the Microsoft Visual Studio 2010 environment, the program is designed in the Visual Basic language, and the Access 2010 is the basic data storage of medical personnel and patients. It shall save the patients barcode data and time generated in the rehabilitation movement, for the convenience of data establishment and future inquiry reference of rehabilitation condition.

2.3 *Experimental Methods*

The plan of the test is divided into the person with normal upper limb function and the patient with upper limb dysfunction for verification. It shall ask the participator to glide the hand gliding cart on the panel by hands, and the system shall record the movement track coordinate and make statistical analysis, for the reference of clinical personnel.

2.3.1 Experimental Design

In the normal person part, it includes 10 times of trainings and tests on the device of this research by three persons with normal upper limb function, to compare the

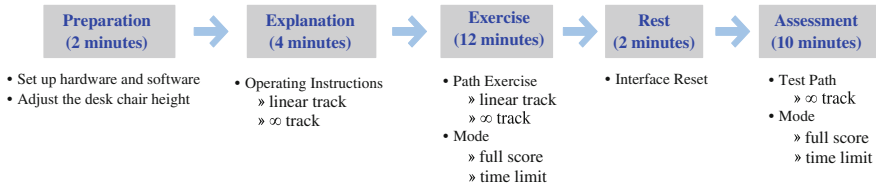


Fig. 4 Experimental process planning

change after training, and judge whether there is the same assessment effect when using this system aiming at the strong hand, to ensure the system stability and reproduction [4, 5].

In the upper limb dysfunction part, it is drafted to collect the data of one patient with upper limb dysfunction, for the training and test twice a week for four weeks.

2.3.2 Experimental Procedures

Figure 4 shows the test flow plan of once training, divided into five phases of preparation, explanation, exercise, rest and assessment, totally about 30 min. The training action is divided into the linear (longitudinal, transverse) track and ∞ track. The assessment takes the ∞ track as baseline, divided into the “full score” and “time limit” two modes; in which, the “full score” mode is to do the movement repeatedly if there does not read the barcode, until reading all the barcodes (namely: zero omission). While the “time limit” mode is to move the hand gliding cart in proper speed within the limited time (30 s), to record the barcode omission times.

3 Results

The participator shall move the hand gliding cart according to the appointed track, and the barcode scanner shall record and transmit to the computer server through Bluetooth wireless transmission, whose movement track would show in the computer screen, and the computer shall trace the relative position data of barcode diagram track. Figure 5a shows when the participator does the transverse track training, the computer screen shall display the track route in real time, to let the participator understand the action condition of operating the hand gliding cart, as the reference of amending the hand gliding cart movement direction. Figure 5b shows the track route picture displayed in the screen, in which the diamond shows the color representing the barcode scanner reading the corresponding barcode, which also means the change condition of hand gliding cart moving track route.

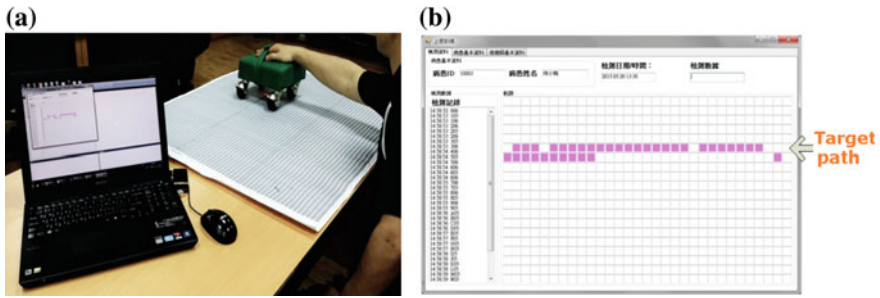


Fig. 5 The transverse track training. **a** Track route displayed in the screen. **b** Track route displayed in the screen

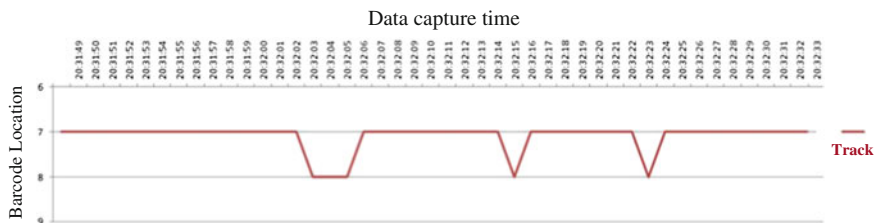


Fig. 6 The corresponding track diagram

Figure 6 shows the corresponding track diagram described according to the detection data stored in the computer, which could be the reference of all previous training results.

3.1 Test of the Persons with Normal Upper Limb Function

In order to verify the stability and reproduction of the system, by three persons with normal upper limb function of the system testing. Table 1 shows the basic information of the subject.

Table 2 shows the average value of omission times for 10 continuous tests of transverse, longitudinal straight-line and ∞ track movements with the dominant

Table 1 The basic information of the subject

Participants	Mr. A	Mr. B	Ms. C
Sex	♂	♂	♀
Age	42	56	35
Height (cm)	170	168	159
Weight (kg)	80	72	58
Dominant side	Right	Left	Right

Table 2 The average value of omission times (persons with normal upper limb function)

Participants	Transverse straight-line	Longitudinal straight-line	∞
Mr. A	2.7	2.1	7.1
Mr. B	2.5	1.8	6.7
Ms. C	2.8	2.3	7.6

Table 3 The full score of ∞ action test (persons with normal upper limb function)

Class	Average usage time (s)	
	Left	Right
Mr. A (dominant side: right)	201	160
Mr. B (dominant side: left)	140	173
Ms. C (dominant side: right)	164	133

P.S. Characters have shading: the better performers

side; in which, in the “longitudinal straight-line action”, the arm makes the stretch in back and forth direction, whose omission times are the lowest in the three movement trainings. While in the “transverse straight-line action”, the arm needs to do the bending and extending, so the omission times are a little more. And in the “ ∞ action”, the arm and hand wrist need to do some coordination, thus the omission times are the most.

Table 3 shows the full score test of doing “ ∞ action”, which indicates the average usage time performing well in the Dominant side.

3.2 Test of the Persons with Impaired Upper Limb Function

There is one patient with right hand elbow (tennis elbow), hand wrist (wrist tunnel syndrome) functional damage caused by sport injury, to do the training and verification by the training device developed by this research. Table 4 shows the basic information of the subject.

The training period of Mr. D is four weeks (twice/week, 3 rounds/time). Before the training, he stretches the body, muscle and bone by the “transverse straight-line” and “longitudinal straight-line” movement. After the line track is stable, he enters

Table 4 The basic information of the subject

Participants	Mr. D
Sex	σ
Age	38
Height (cm)	175
Weight (kg)	78
Dominant side	Right
Symptom	Right hand elbow (tennis elbow), right hand wrist (wrist tunnel syndrome)

Table 5 The performance of Mr. D during the training period

Week and performance	Usage time of full score (s)		Barcode omission times (limit time: 30 s)	
	Training			
	Before	After	Before	After
Wk-1	683.6	531.6	85.7	76.3
Wk-2	589.2	429.5	70.3	58.7
Wk-3	379.8	203.2	45.2	31.6
Wk-4	205.6	187.5	38.7	20.3

the “∞ action” training and test (usage time of full score, and barcode omission times within the limited time (30 s)). The performance of Mr. D during the training period is as shown in Table 5, which indicates his performance improves stably, which means the upper limb function is indeed improved.

4 Discussion

Utilizing the light barcode reader with Bluetooth wireless transmission function, to develop the low-cost and portable upper limb training device (hand gliding cart), it could automatically capture the barcode information in the rehabilitation training process, which is highly effective in the aspects of data recording correctness and manpower saving.

From the initial linear movement stretch gradually changing to the advanced ∞ curve movement, it could detect the stability, flexibility and maximum angle in the movement of hand activity.

The training track route could be programmed, so the conductor could make adjustment for the training track according to the condition of trainee, which conforms to the advantage of easy to use and flexible change, and is worth of practical promotion for application. It contributes to improve the personal life quality of patients, and reduce the social cost for care.

Acknowledgment This work was supported by the Ministry of Science and Technology, TAIWAN, 104-2221-E-146-011.

References

1. Hwang Y-S, Chen S-C, Chen C-C, Chen W-L, Shih Y-Y, Chen Y-L (2013) Development of digitized apparatus for upper limb rehabilitation training. *Technol Health Care* 21:571–579
2. Chen C-C, Lin J-C, Chou C-H, Shih Y-Y, Chen Y-L (2010) Digital skating board with RFID technique for upper extremity rehabilitation. In: *Second international conference, ICMB2010*, pp 209–214

3. Liu C-T (2011) Infrared detection of digital train system for upper extremity. National Taipei University of Education Master Thesis
4. American Association on Mental Retardation (2007) Mental retardation: definition, classification, and systems of supports, 10th edn
5. Cook AM, Hussey SM (2007) Assistive technologies, 2nd edn. Principles and practice

LBS with University Campus Navigation System

Jiun-Ting Chen and Ya-Chen Chang

Abstract University campus may be very large or it may have many campuses. Every year lots of new students come in the university. Students need to buy books, stationeries and need to find something to eat. In those huge campus, it's very difficult to find where to buy they needs. It creates problem to the new comers and visitors to reach easily and timely. Location-based services (LBS) now are very popular marketing tools. In this paper, we discuss the development a mobile application that delivers personalized campus maps for universities. Everyone can build own list to save that's needed stores. In this paper, we discuss the development a mobile application it will combine campus maps and LBS with stores around universities. It will save problems for visitors, new comers and also new faculty; staff may found some stores they never find. For those stores, revenues will increase and profits will improve. If this feature is integrated with Google Maps, it will be very helpful both for existing and new comers of University campus. There will be an administrator who will update event information on server.

Keywords Mobile application · Campus navigation · LBS · University

1 Introduction

During the last few years, the development of mobile devices has gained significant progress with respect to memory capabilities, advanced processing power and higher transfer rates to name only a few performance parameters. Nowadays android mobile becomes the most popular in the smart phone market because android is an open source [1] mobile Operating System based on Linux with java support and it comes under free and open source software licenses.

J.-T. Chen (✉) · Y.-C. Chang
Digital Multimedia Arts, Shih Hsin University, Taipei, Taiwan
e-mail: andy@mail.shu.edu.tw

Y.-C. Chang
e-mail: jon121209@gmail.com

Location-based services (LBS) provide personalized services to the mobile clients according to their current location [2]. Geographical Information System (GIS) is the core part of LBS to provide all the valuable features of LBS [3]. People can track own location and also navigate from one location to another location very easily. There are lots of technology to track location like Cell Identification, Global Positioning System (GPS), Various Radiolocation systems, Accelerometers and Electronic Compass etc. [4]. GPS gives much higher accuracy of latitude and longitude compare to other techniques but it works only in outdoors, not in indoor. The Location Tracking techniques can be worked with all today's market cell phones with networks such as GSM (Global system for Mobile Communication), GPRS (General Packet Radio Service) and CDMA (Code Division Multiple Access).

There are many applications and commercial devices that provide driving directions and navigation such as Waze [Waze Navigator], Google Navigation [Google Maps], in-car navigation, Magellan navigation devices [Magellan Smart GPS], and Garmin navigation devices [Garmin Navigation] [5]. This navigation became easier with the help of Google Maps on GPS enabled android devices. GPS applications allow users to find a destination based on their current location. So, location searching becomes a new trend with the combination of Google Maps and GPS. It provides lots of additional features [6] like displaying congested route, smart driving decisions and improve driving safety and reduces time and energy while going to an unknown places.

2 Case Studies

The focus of this study was to investigate the usefulness of context-aware maps that utilize contextual information from new university students; helping them navigate around the campus, locate buildings that are relevant to them easily, and to become familiar with the university environment quickly. University orientation is the key event for delivering information to the new incoming students about the university services and environment that are relevant to them.

Based on two University administered student orientation surveys, a questionnaire and a focus group with university engagement staff, it was evident that new students often feel lost. It is not easy for them to become familiar with the university buildings and resources, in particular the location of important buildings such as the library, the administration center or their faculty building. Not surprisingly, students found it difficult to find buildings that were relevant to them, and it typically took them a long time to navigate within the university campus. At the subsequent focus group discussion, it was decided that a mobile application would have the potential to address these problems, as well as providing an opportunity to explore how different contexts could be utilized and combined to help a student settle into their university life. For this case study, we aimed to test the hypotheses that the personal and environmental context of a user has an effect on the usefulness of the map in a mobile application for new university students.

3 Definitions

People use Google maps for navigation purpose to drive or to walk on an unknown location or to make trip plan. But its capability is limited on street. It is not prominent for our university campus. Location based services is very helpful for navigation purpose.

3.1 *Android*

Android [3] is a combination of three components:

- Free and open source operating system for mobiles.
- Open source development platform for creating mobile applications.
- Devices, particularly mobile phones that run Android operating system and the applications created for it.

Android consists of mainly five layers these are Application Layer, Application Framework Layer, Libraries Layer, Android Runtime Layer, and Linux Kernel. Application Layer consists of lots of user specific applications like Short Message Service (SMS), email, phonebook, browser, map etc. Application Framework Layer consists of the programs that manage the phone's basic functions like resource allocation, telephone applications, switching between processes or programs and keeping track of the phone's physical location. Libraries Layer consists of the native libraries of Android. These libraries are shared to all programs. Android Runtime Layer includes Dalvik Virtual Machine (DVM) and set of core libraries of Java. Every application gets its own instance of DVM. DVM is programmed in such a way that multiple instance of DVM can run very efficiently in same time. Linux Kernel consists of Android's memory management programs, security settings, power management software and several drivers for hardware, file-system access, networking, inters process communication.

3.2 *Global Positioning System*

GPS is one kind of very popular navigation system. It helps to track user location with the Latitude, Longitude and altitude of device. The system consist of networks of 24 satellites in six different 12 h orbital paths spaced so that at least five are in view from every point on the globe and their ground stations [4]. So it gives more accurate value of Latitude and Longitude of a position. It updates the location of

device after every 5 s. Therefore its response time is slow, which is the major drawback of GPS system. Another drawback is that it works only in outdoors but not in indoor.

3.3 HTML, CSS, Java Script

HTML stands for Hyper Text Markup language. It's basically a language to design a static web page. It consists of some set of tags like <html>, <head>, <body>, <a>, <table>, <frame>, <frameset> etc. for design purpose.

CSS stands for Cascading Style Sheet; this is a part of Dynamic HTML (DHTML) by which different property of HTML can be accessed for beautification purpose, which is not possible with normal HTML.

Java Script is a scripting language invented by Netscape Navigator mainly used to add validation of input fields means to check the input given by user is correct or not. We can create new object and also can use system defined object. It also used for event handling purpose like onclick, onmouseover, onmouseout, onsubmit, onreset, onload etc.

3.4 JSP, Servlet

JSP (Java Server Page) is an example of server side programming technique that gets executed by web container environment of web server machine by which client's request can be traced and proper response can be generated. All the JSP code which is based on Java is converted to a Servlet code internally.

The Servlet is also a Java .class file used to extend the capabilities of server. Servlet is used mainly for:

- Process or store data that was submitted from an HTML form.
- Provide dynamic content such as the result of a database query.
- Manage the state information that does not exist in the stateless http protocol.

3.5 Apache Tomcat

Tomcat is an open source web server and Servlet container developed by the Apache Software Foundation (ASF). Tomcat implements the Java Servlet and the JSP specifications from Sun Microsystems, and provides a pure Java HTTP web server environment for Java code to run in. In the simplest configuration, Tomcat runs in a single operating system process. The process runs a Java Virtual Machine (JVM).

4 Application Design

This application serves as a digital campus map, providing information on university buildings, services, transportation, food courts and social places as students navigate through the university campus. And students also can find anything they want to buy on this application.

4.1 Standard Versus Personalized Map

In this application we created two mode to compare: (1) Standard Map and (2) Personalized Map. These two versions served the same purpose of helping students navigate through the campus and become familiar with the university environment.

- ‘Standard Map’ provides a digital map with all the campus buildings on their mobile devices. It also shows a user’s current location on the map as they navigate around the campus (see Fig. 1 left).

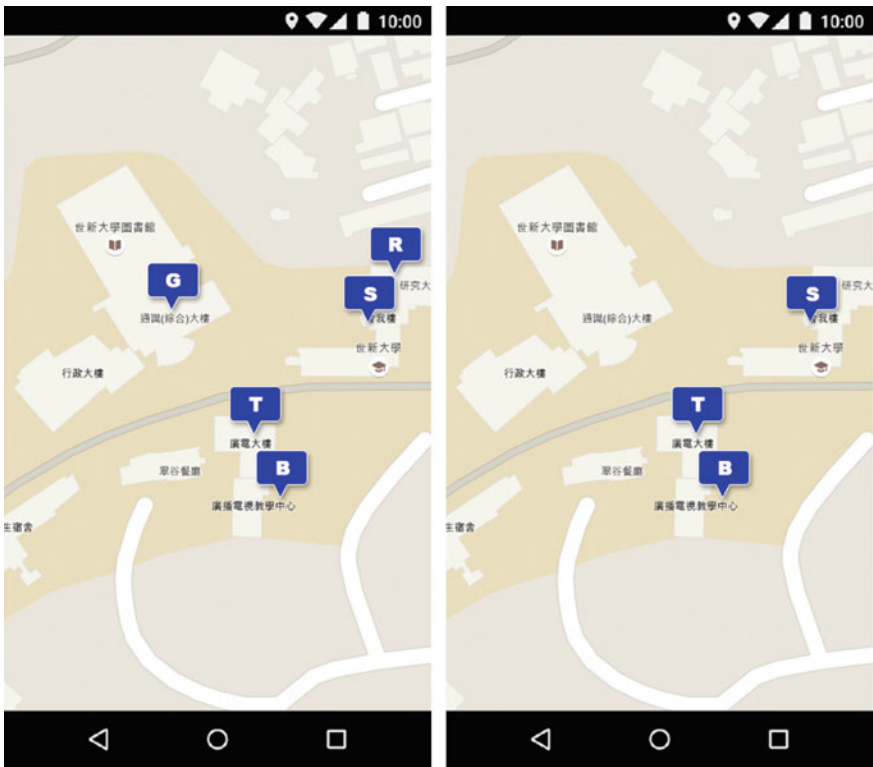


Fig. 1 Standard map (left), personalized map (right)

- ‘Personalized Map’ has an additional feature on top of the Standard Map. It filters the buildings on the map utilizing a student’s profile information, such as their faculty (e.g. business, education, etc.), type of student (i.e. domestic or international), and type of buildings (e.g. social, travel, eat, see or shop) that they are interested in. As a result, the mobile application displays a personalized campus map of the places that are important and relevant to the students (see Fig. 1 right).

4.2 Components

The application consists of three components: a map component, filter component, and assistance component.

The map component contains a digital map that displays a user’s current location and the buildings within the university campus. A user’s location is represented as a blue dot, and buildings are represented as square pins on the map. Users are able to scroll, zoom, center and rotate the map based on the device orientation. When launched, the application automatically displays the nearest university campus to the user’s location.

The filter component provides information on each building located on the map, the name of the building, and the services within the building. Moreover, users are able to change what they want to see on the map by changing the filter options. The filter options in Standard Map and Personalized Map are designed differently. Standard Map provides the option of buildings, places of interest, bus stops, Wireless Fidelity (Wi-Fi) Access Points, and Automated Teller Machine. The filter options in Personalized Map are faculty buildings, essential services (e.g. libraries and information desk), social places (e.g. guild bar and theatres), place for travel (e.g. train stations and bus stops), places to eat (e.g. food courts and cafes), places to see, and places to shop (e.g. bookshop and shopping center).

The assistance component is a feature that provides recommendations based on a user’s personal and university environment information. For personal context, a student’s faculty and type (e.g. domestic or international) are used to pinpoint their faculty buildings and the essential services that are relevant to them. For environmental context, a building’s type, services provided and location are used. Additionally, this component removes buildings that are not important to the students from the map. Student’s faculty buildings are marked as red and essential services are marked as green. Ultimately this data would be gathered automatically from university services, however for the prototype the researcher entered it.

5 Conclusions and Future Work

This pilot study describes how context-awareness has an impact on the usefulness of a campus map mobile application that utilizes personal and environmental contexts, and reports on the results from a university orientation case study. Results show that integrating personal and environmental contexts on digital maps can improve map usefulness and navigation efficiency. Future work will look into expanding the application with personalized campus tours, adding extra sensor data such as Wi-Fi hotspots or Bluetooth connectivity, as well as testing the application in indoor environments. We will also experiment the application in different types of locations, such as hospitals and shopping malls.

References

1. Bhattacharya S, Panbu MB (2013) Design and development of mobile campus, an android based mobile application for university campus tour guide. *Int J Innovative Technol Exploring Eng (IJITEE)* 2(3)
2. Singhal M, Shukla A (2012) Implementation of location based services in android using GPS and web services. *Int J Comput Sci Issues (IJCSI)* 9(1) (no 2)
3. Rani CR, Kumar AP, Adarsh D, Mohan KK, Kiran KV (2012) Location based services in android. *Int J Adv Eng Technol* 3(1):209–220
4. Bhatia S Hilal S (2013) A new approach for location based tracking. *Int J Comput Sci Issues (IJCSI)* 10(3) (no 1)
5. Cardei M, Jones B, Raviv D (2013) A pattern for context-aware navigation. In: 20th conference on pattern languages of programs (PLoP), PLoP'13
6. Cardei M, Zankina I, Cardei I, Raviv D (2013) Campus assistant application on an android platform. In: 2013 Proceedings of IEEE southeastcon. pp 1–6

An Efficient Energy Deployment Scheme of Sensor Node

Cheng-Chih Yang, Hsuan-Fu Wang and Yung-Fa Huang

Abstract This text provides an efficient energy scheme of wireless sensor node dynamic deployment. A single sensor node communication model is defined initial. Sensors are co-working with their neighbors in their transmitting range. Suitable neighbor node number and sensing radius make sensors field more efficient. A value CAPR is defined in this text for power efficiency discriminating. A self-regulated mechanism is proposed here that sensor can adjust its radius of sensing range for high efficient energy working.

Keywords Wireless sensor node · Coverage area · Neighbor node numbers · CAPR · Self-regulated mechanism

1 Introduction

Wireless Sensor Network (WSN) has been proposed for many applications in which the sensors nodes are capable of sensing, computation, and communication [1]. In WSN, sensors deployment builds the network topology. The topology of WSN will decide the performance of networks. Although most of the current WSNs consist of static sensor nodes, there are many applications where mobile nodes are involved [2]. Moreover, due the varying environment and the vulnerable of sensor nodes, the stationary deployment would suffer performance degradation in a long

C.-C. Yang
Department of Electrical and Information Technology Engineering,
Nankai University of Technology, Caotun, Taiwan

H.-F. Wang (✉)
Department of Electrical Engineering and Energy Technology,
Chung Chou University of Science and Technology, Changhua, Taiwan
e-mail: rex_wang@dragon.ccut.edu.tw

Y.-F. Huang
Department of Information and Communication Engineering,
Chaoyang University of Technology, Taichung, Taiwan

period. Therefore, the dynamic deployment with mobile nodes can adaptively optimize the network topology for WSNs [2–4]. Integrating multiple nodes deployment, the mainly focuses is how to deploy the sensors reasonably to guarantee a highly-effective on coverage for a Range of Interest (ROI), that is, how to construct a biggest coverage using minimum power consumption is the main goal in this paper. Many methods have been developed on the sensors placement and then to improve the coverage rate, such as force field based methods and virtual force algorithms (VFA) [5]. Recently, a Hybrid Virtual Force Algorithms (HVFA) has been proposed to improve the performance of coverage rate, connectivity and energy-efficient [6]. But power consumption is high for node moving. Further, we focus the selection of a sensor node's radius of transmit-receive communication according sensor node's local density. The rest of this paper is organized as follows. Section 2 describes wireless sensor node communication model and coverage area estimating. Section 3 discusses sensor field power consumption and power efficiency CAPR. Section 4 described a self-regulated mechanism. Section 5 provides some conclusions.

2 Wireless Sensor Node Communication Model

2.1 Single Sensor Node Communication Model

A wireless sensor node is working with sensing and communication jobs simultaneous in WSN. Each sensor node sends its beacon through transmitting ability. It also can detect an event occurring within its sensing range then delving this message to its neighbors within communication range. Meanwhile, each sensor node can receive (listen) any information from other sensor node within sensing range. Figure 1 depicts a single sensor node's sensing and communication working scope with sensing radius r_s and communication radius r_c . Transmitting distance r_c need larger than sensing distance r_s for information forward.

Fig. 1 Single sensor node communication model

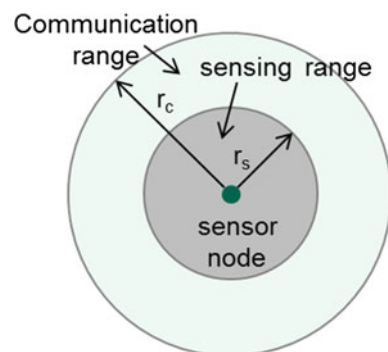


Fig. 2 Two closed sensors and their coverage area

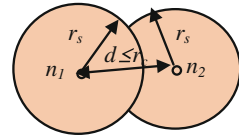
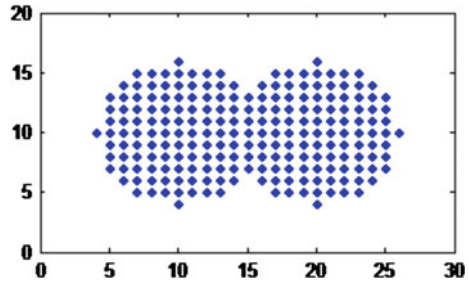


Fig. 3 Two nodes' coverage area calculation CA = 217 m²



2.2 Multiple Sensor Coverage Area Estimation

In real case, we need multiple sensor for sensing scope maximization in communication range. In Fig. 2, two sensor nodes n_1, n_2 are separated a distance $d \leq r_c$. Sensing range are overlapping each other. We call the total sensing area as coverage area CA of nodes n_1, n_2 . The larger CA in ROI, the larger scope they inspect.

A critical problem, how estimate the coverage area. In this paper, we calculate all integral position points, then accumulate total point numbers (grid-based calculated). An example in Fig. 3, we put two nodes at (10, 10), (20, 10), and let $r_s = 6$ m. The coverage area is calculated as CA = 217 m².

3 Sensor Field Power Consumption and Power Efficiency

3.1 Sensor Field Power Consumption

In front section, we had mentioned that a sensor node need join some another neighbor nodes in its communication range to achieve a maximum coverage area. In this section, we concern about how many neighbor nodes are need in a communication range and how much power energy are consumed. In a wireless sensor

field, transmitting power P_T and sensing power P_S are two main parts of power consumptions P_C . They can be expressed as:

$$P_c = P_T + P_s \tag{1}$$

$$P_T = k_T r_c^2 \tag{2}$$

$$P_S = k_S r_s^2 \tag{3}$$

Let $r_c = r_s = 1$ m, we simulate $P_T = 10$ mW, $P_S = 3$ mW. Then the parameters can be calculated as $k_T = 10 \times 10^{-3}$, $k_S = 3 \times 10^{-3}$. Consider multiple nodes, we consider 3 various neighbor node numbers deployment cases as below.

Case 1: Deployment with neighbor node numbers = 3, $r_c = r_s = 10$ m, in full coverage state that shown as Fig. 4. In this case, the total power consumption

Fig. 4 Sensor deployment with neighbor node numbers = 3, $r_s = r_c = 10$ m

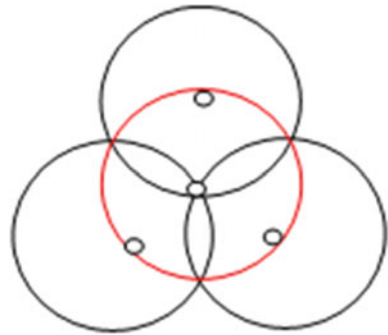


Fig. 5 Sensor deployment with neighbor node numbers = 4, $r_s = 0.8r_c$

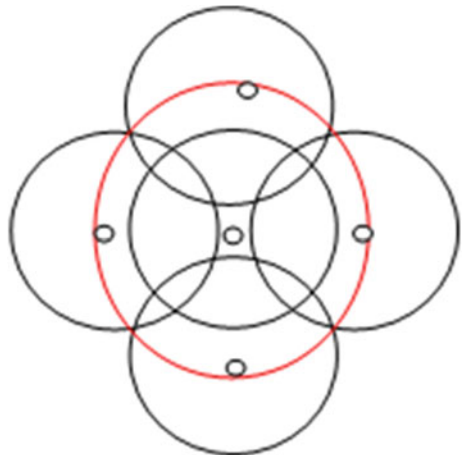
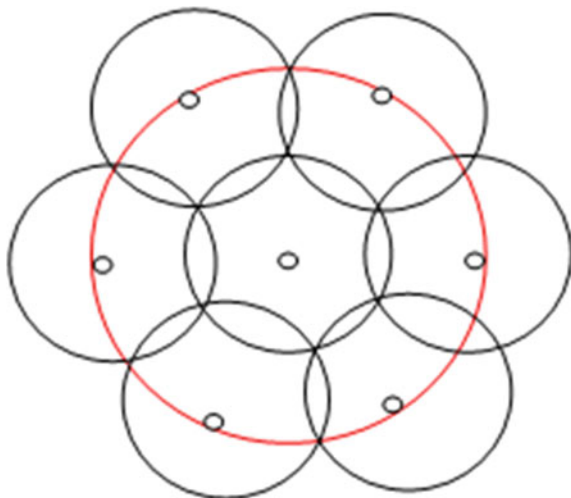


Fig. 6 Sensor deployment with neighbor node numbers = 6, $r_s = 0.6r_c$



$$P_C = P_T + 4P_S = 10 \times 10^{-3} \times 10^2 + 4 \times 3 \times 10^{-3} \times 10^2 = 2.2 \text{ W.}$$

Case 2: Deployment with node numbers = 4, $r_c = 10 \text{ m}$, $r_s = 0.8r_c = 8 \text{ m}$, in full coverage state that shown as Fig. 5. In this case, the total power consumption

$$P_C = P_T + 5P_S = 1 + 5 \times 3 \times 10^{-3} \times 8^2 = 1.96 \text{ W}$$

Case 3: Deployment with $n_e = 6$, $r_c = 10 \text{ m}$, $r_s = 0.6r_c = 6 \text{ m}$, in full coverage state that shown as Fig. 6. In this case, the total power consumption

$$P_C = P_T + 7P_S = 1 + 7 \times 3 \times 10^{-3} \times 6^2 = 1.76 \text{ W}$$

We summarize the results of three cases into Table 1.

Table 1 Summary of three cases in full coverage state

Case	Neighbor numbers	r_c (m)	r_s (m)	P_C (W)
1	3	10	10	2.2
2	4	10	8	1.96
3	6	10	6	1.76

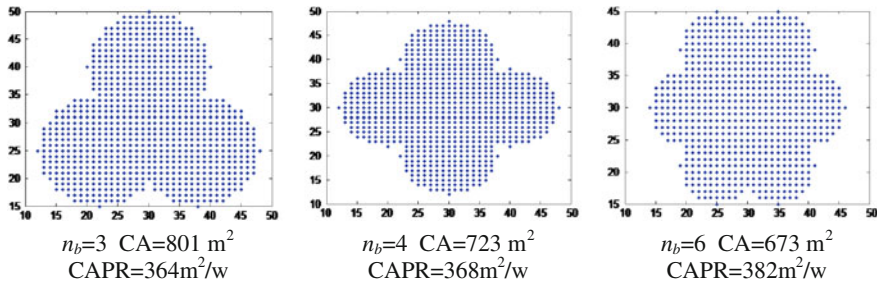


Fig. 7 Three various neighbor nodes coverage area and its CPPR

3.2 Coverage Power Efficiency and Simulation Results

Various neighbor node numbers construct various coverage area. Here, we define a discriminating value CAPR.

$$\begin{aligned} \text{CAPR (Coverage Area to Power consumption Ratio)} \\ = \text{coverage area/power consumption} \end{aligned}$$

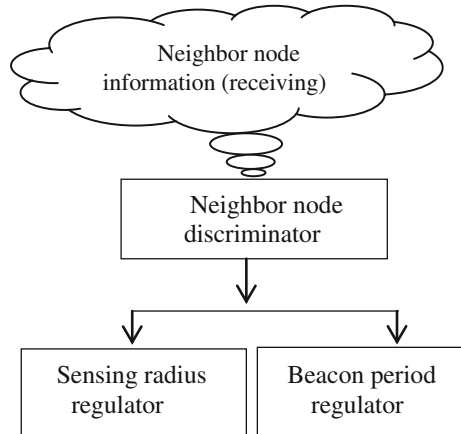
More higher CAPR, more higher efficient for wireless sensor network is. We select appropriate node position, then estimate the coverage area for each case. Results are shown in Fig. 7. For case 1, neighbor node number $n_b = 3$, CA = 801 m², CAPR = 364 m²/w. For case 2, neighbor node number $n_b = 4$, CA = 723 m², CAPR = 368 m²/w. For case 3, neighbor node number $n_b = 6$, CA = 673 m², CAPR = 382 m²/w. From data analysis, the network of $n_b = 6$ is better than other.

4 Self-regulated Mechanism

Different from node moving scheme for sensor deployment, here we propose a self-regulated sensing radius mechanism shown as Fig. 8.

This mechanism include a neighbor node discriminator, sensor can adjust its sensing radius according neighbor node in a period time. An example explaining: A sensor node is working with other 4 neighbor nodes with $r_c = 10$ m, $r_s = 8$ m initial. After a period time, this sensor discriminate its neighbor number changing to 6. Then this mechanism changes r_s from 8 to 6 m.

Fig. 8 A self-regulated mechanism for sensor node



5 Conclusions

Different from node moving scheme for sensor deployment, this text proposed a self-regulated scheme that sensor can adjust its radius of sensing range for high efficient power reaching in wireless sensor network. 6 neighbor nodes is an optimal selection for sensor working design.

References

1. Akyildiz F, Sankarasubramaniam SY, Cyirci E (2002) Wireless sensor networks: a survey. *Comput Netw* 38(4):393–422
2. Howard A, Mataric MJ, Sukhatme GS (2002) Mobile sensor network deployment using potential fields: a distributed, scalable solution to the area coverage problem. In: *Proceedings of 6th international conference on distributed autonomous robotic system*, Fukuoka, Japan, pp 299–308
3. Thangadurai N, Dhanasekaran R, Karthika RD (2013) Dynamic energy efficient topology for wireless ad hoc sensor networks. *WSEAS Trans Commun* 12(12):651–660
4. Costa DG, Guedes LA, Vasques F, Portugal P (2013) Redundancy-based semi-reliable packet transmission in wireless visual sensor networks exploiting the sensing relevancies of source nodes. *WSEAS Trans Commun* 12(9):468–478
5. Zou Y, Chakrabarty K (2003) Sensor deployment and target localization based on virtual forces. In: *Proceedings of IEEE INFOCOM*, pp 1293–1303
6. Wen J, Yang C, Huang Y (2014) Performance of hybrid virtual force algorithms on mobile deployment in wireless sensor networks. *WSEAS Trans Commun Art.* #61 13:558–566, 2014

Channel Equalization for MIMO LTE System in Multi-path Fading Channels

Hsuan-Fu Wang, Mu-Song Chen, Ching-Huang Lin
and Chi-Pan Hwang

Abstract The Long Term Evolution (3GPP-LTE) combines the Multi-input multi-output (MIMO) antenna and the Orthogonal Frequency Division Multiple Access (OFDMA) techniques to accomplish high speed data transmission. Comparisons of zero forcing (ZF), minimum mean square error (MMSE) and sphere decoding (SD) equalization methods with Turbo code are given under time-varying multi-path fading channel with Doppler frequency shift. The simulation shows that the MMSE and SD equalizer are more robust than the ZF equalizer as the Doppler frequency shift increases.

Keywords MIMO · LTE · ZF · MMSE · SD

1 Introduction

The Long-Term Evolution (LTE) air interface that introduced several technological evolutions as a means to support the advanced information services with universal access and high speed data exchange under restricted frequency band, transmission bandwidth

H.-F. Wang (✉)
Department of Electrical Engineering and Energy Technology,
Chung Chou University of Science and Technology, Changhua, Taiwan
e-mail: rex.wang.sige@gmail.com

M.-S. Chen
Department of Electrical Engineering, Da-Yeh University, Changhua, Taiwan
e-mail: chenms@mail.dyu.edu.tw

C.-H. Lin
Department of Electronic Engineering, National Yunlin University of Science
and Technology, Yunlin, Taiwan
e-mail: wilburlin@yuntech.edu.tw

C.-P. Hwang
Department of Electronic Engineering, National Changhua University of Education,
Changhua, Taiwan
e-mail: cphwang@cc.ncue.edu.tw

and power [1]. This includes block transmission technique using multi-carriers (MCs), multi-antenna systems (MIMO), base station (BS) cooperation, multi-hop relaying, as well as the all-over IP concept. The Orthogonal Frequency-Division Multiple Access (OFDMA) and Single-carrier Frequency-Division Multiple Access (SC-FDMA) methods are adopted for the down-link and up-link transmission for the air interface of the LTE [2].

An OFDM transmission scheme can represent a frequency-selective fading channel as a group of narrowband flat fading sub-channels. This in turn enables OFDM to provide an intuitive and simple way of estimating the channel according to reference signals or transmitting known data when the frequency spacing between subcarriers is sufficiently small. Accordingly, the OFDM receiver can recover the best estimate of the transmitted signal using a low-complexity frequency-domain equalizer (FDE) with a good estimate of the channel response at the receiver [3].

The first LTE Release introduces the MIMO operation that containing spatial multiplexing in addition to pre-coding and transmit diversity. The spatial multiplexing is sending signals from two or more different antennas with different data streams and by signal processing means in the receiver separating the data streams, therefore increasing the peak data rates by a factor of 2 or 4 based on 2-by-2 or 4-by-4 antenna configuration. The MIMO is one of the key technologies deployed in the LTE standards to support the advanced multi-media services with universal access and high speed information exchange under limited radio resources. Different types of MIMO method, including open-loop and closed-loop spatial multiplexing, are adopted in the LTE standard [4].

This paper is organized as follows. Section 2 introduces the baseband MIMO LTE system model. Section 3 presents the zero forcing (ZF), minimum-mean-squared error (MMSE) and sphere decoding (SD) equalizers. Section 4 presents simulation results. Section 5 presents the conclusions.

2 Simplified MIMO LTE System Model

Considering a simplified based-band MIMO LTE system that depicted in Fig. 1. The input binary data is first encoded by a Turbo encoder [5], then the encoded data is grouped and mapped based on the modulation scheme. Then, based on the layer mapping and precoding configuration [6], a single stream of modulated symbols is subdivided into multiple sub-streams destined for transmission via multiple antennas. Then the data sequence of length $X_i(k)$ for antenna i of the transmitter is transformed into time domain signal $x_i(n)$,

$$x_i(n) = \sum_{k=1}^{N-1} X_i(k) e^{j(2\pi kn/N)}, \quad (1)$$

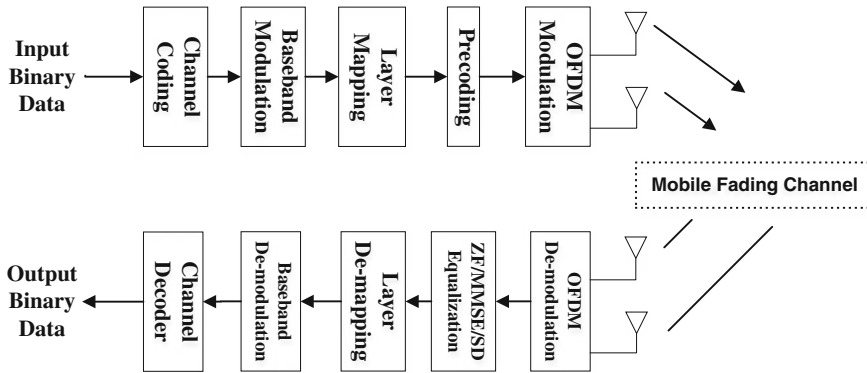


Fig. 1 The simplified baseband MIMO LTE system model

where N is the inverse discrete Fourier transform (IDFT) length. A guard time which is chosen to be larger than the expected delay spread and cyclic prefix (CP) which is the cyclically extended part of OFDM symbol are inserted into to eliminate inter-symbol interference (ISI) and inter-carrier interference (ICI). The resultant OFDM symbol is

$$x_{i,GCP}(n) = \begin{cases} x_i(N+n), & n = -N_{GCP}, -N_{GCP} + 1, \dots, -1 \\ x_i(n), & n = 0, 1, \dots, N - 1 \end{cases}, \quad (2)$$

where $N_{GCP} = N_G + N_{CP}$, N_G and N_{CP} are the length of the guard time and CP, respectively. The signal $x_{i,GCP}(n)$ is transmitted through the time varying frequency selective fading channel $h(n)$ with additive White Gaussian noise (AWGN) $w(n)$.

3 MIMO Receiver

In order to obtain the estimated OFDM symbols, the MIMO receiver performs the most the same operation as the transmitter with equalization signal processing. For the 2-by-2 MIMO configuration, the received signal \mathbf{r} that eliminating the AWGN,

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix}_{\mathbf{r}} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}_{\mathbf{H}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{\mathbf{x}}, \quad (3)$$

where \mathbf{H} is the channel matrix.

3.1 Zero Forcing Equalizer

According to (3), i.e., apply the channel matrix \mathbf{H}^{-1} to the both side of the equation, the estimated OFDM symbol is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{H}^{-1} \mathbf{r} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \quad (4)$$

As shown in (4), the zero forcing equalizer \mathbb{Z}_{ZF} is \mathbf{H}^{-1} [7].

3.2 Minimum Mean Square Error Equalizer

Considering the AWGN, the received signal \mathbf{y} is

$$\mathbf{y} = \mathbf{r} + \mathbf{w} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \quad (5)$$

Defining the error signal \mathbf{e} as

$$\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x} = \mathbb{Z}_{MMSE} \mathbf{y} - \mathbf{x}, \quad (6)$$

where, $\hat{\mathbf{x}}$ and \mathbf{x} are the estimated and ideal OFDM symbol, respectively. Applying the Eq. (6), i.e., minimizing the expected value of the error signal \mathbf{e} , the minimum mean square error equalizer \mathbb{Z}_{MMSE} [8] is

$$\mathbb{Z}_{MMSE} = \frac{\mathbf{H}^H}{\left(\|\mathbf{H}\|^2 + \sigma_w^2 \mathbf{I}_2 \right)}, \quad (7)$$

where $(\)^H$ denotes the Hermitian operation, σ_w^2 is channel noise variance and \mathbf{I}_2 is the 2-by-2 identity matrix.

3.3 Sphere Decoding Equalizer

The sphere decoding (SD) algorithm can be applied to find the maximum-likelihood (ML) solution for the Eq. (5), i.e., the SD evaluates all transmit signal vectors \mathbf{x} fulfilling the following criterion: $(\mathbf{y} - \mathbf{H}\mathbf{x})^H (\mathbf{y} - \mathbf{H}\mathbf{x}) < R^2$, where R is the search

radius of a hyper-sphere [1]. Hence, the ML estimate for the transmitted OFDM symbol is

$$\hat{\mathbf{x}} = \arg \min [(\mathbf{y} - \mathbf{H}\mathbf{x})^H(\mathbf{y} - \mathbf{H}\mathbf{x})] \quad (8)$$

4 Simulations

4.1 Simulation Configuration

The simulation parameters are indicated in Table 1 [9].

The channel models used for the simulations presented here are: Extended Pedestrian A model and Extended Vehicular A model. The multi-path fading is modeled as a tapped-delay line with a number of taps at fixed positions on a sampling grid. The gain associated with each tap is characterized by a distribution (Ricean with a K-factor > 0 , or Rayleigh with K-factor = 0) and the maximum Doppler frequency that is determined from the mobile speed. The definition of the 2 specific channels are shown in Tables 2 and 3.

4.2 Simulation Results

The bit error rate (BER) performance of the ZF, MMSE and SD MIMO receiver have simulated to obtain Figs. 2, 3 and 4. Figure 2 shows the comparison of the BER performance of the ZF equalizer in extended pedestrian A channel. It is clear that the BER performance as the Doppler frequency increased in Fig. 2. The BER

Table 1 Simulation parameters

Parameter	Specification
MIMO configuration	2×2 antennas
Layer mapping and precoding scheme	LTE downlink transmission mode 4 [1]
Bandwidth (MHz)	10
Date rate (Mbps)	10.3
OFDM symbol per slot	14
CP length (μs)	4.7
DFT length	1024
Modulation	QPSK
Channel coding	Turbo code [5], code rate = $\frac{1}{3}$, # of iteration = 5
Equalizer	ZF, MMSE, SD
Channel estimation	Ideal channel estimation
Doppler frequency (Hz)	0, 5, 70

Table 2 Extended pedestrian A model

Number of taps	Relative delay (ns)	Average power (dB)
1	0	0
2	30	-1.0
3	70	-2.0
4	90	-3.0
5	110	-8.0
6	190	-17.2
7	410	-20.8

Table 3 Extended vehicular A model

Number of taps	Relative delay (ns)	Average power (dB)
1	0	0
2	30	-1.5
3	150	-1.4
4	310	-3.6
5	370	-0.6
6	710	-9.1
7	1090	-7.0
8	1730	-12.0
9	2510	-16.9

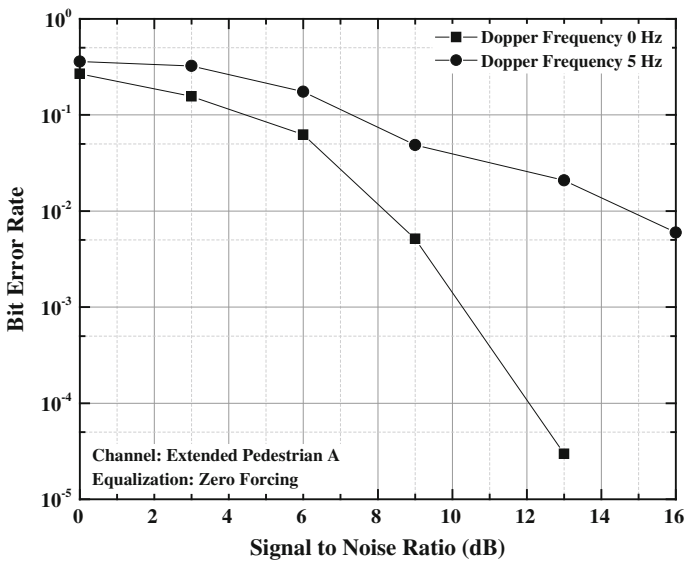


Fig. 2 BER performance of ZF equalizer in EPA channel

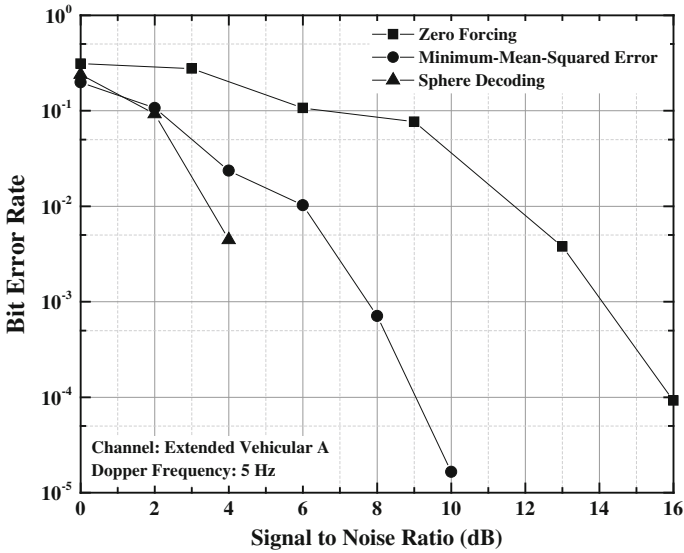


Fig. 3 Performance comparison: ZF/MMSE/SD equalization in EVA channel

performance comparison for ZF, MMSE and SD receivers in extended vehicular A is showed in Fig. 3. The MIMO receiver with the SD equalizer outperforms than that applying MMSE and ZF equalization algorithms. The same simulation results is also showed in Fig. 4 as the Doppler frequency increased in Fig. 4.

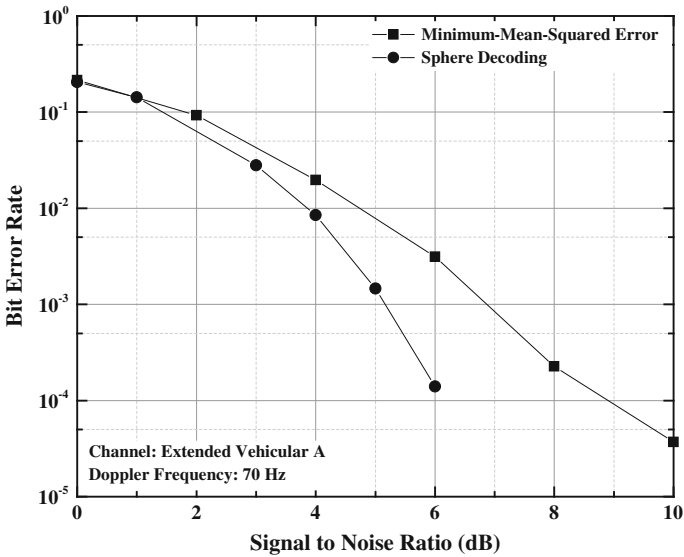


Fig. 4 Performance comparison: MMSE/SD equalization in EVA channel

5 Conclusion

The BER performance for different types of channel equalization techniques that applied to LTE in down-link time-varying mobile fading channel are evaluated in this paper. The SD and MMSE equalizers, due to their robustness against noise, they have better BER performance than the ZF equalizer. This is more evident as the Doppler frequency increases.

References

1. Ahmadi S (2013) LTE-advanced—a practical systems approach to understanding 3GPP LTE releases 10 and 11 radio access technologies. Elsevier, London
2. Marques da Silva M, Monteiro FA (2014) MIMO processing for 4G and beyond fundamentals and evolution. CRC Press, New York
3. Xiao P, Lin Z, Fagan A, Cowan C, Vucetic B, Wu Y (2011) Frequency-domain equalization for OFDMA-based multiuser MIMO systems with improper modulation schemes. *EURASIP J Adv Signal Process* 2011:73
4. Liu L, Chen R, Geirhofer S, Sayana K, Shi Z, Zhou Y (2012) Downlink MIMO in LTE-advanced: SU-MIMO vs. MU-MIMO. *IEEE Commun Mag* 50(2):140–147 Feb
5. Sun Y, Cavallaro JR (2011) Efficient hardware implementation of a highly-parallel 3GPP LTE/LTE-advance turbo decoder. *Integr VLSI J* 44(4):305–315
6. 3GPP (2011) Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation V10.0.0. TS 36.211, Jan 2011
7. Wang C, Au EKS, Murch RD, Mow WH, Cheng RS, Lau V (2007) On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error. *IEEE Trans Wireless Commun* 6(3):805–810 Mar
8. Kim N, Lee Y, Park H (2008) Performance analysis of MIMO system with linear MMSE receiver. *IEEE Trans Wireless Commun* 7(11):4474–4478 Nov
9. Sesia S, Toufik I, Baker M (2011) LTE—the UMTS long term evolution from theory to practice, 2nd ed. Wiley, New York

All-Digital High-Speed Wide-Range Binary Detecting Pulsewidth Lock Loops

Po-Hui Yang, Jing-Min Chen and Zi-Min Hong

Abstract This paper proposed a novel all-digital pulsewidth lock loops which adopted cyclic binary pulsewidth detector and cyclic delay line mechanism. This design has reduced the circuit area of delay line length increase under lower frequency operation, and it utilizes binary pulsewidth detection to lock output pulsewidth rapidly, whose locking time costs in only 25 duty cycles. Moreover, two delay lines is adopted in the pulsewidth generation mechanism circuit, and the cyclic delay line is employed under low frequency operation, but bypassed under high frequency operation for simplifying pulsewidth generating path. The output pulsewidth 25, 50, 75 % (by setting) could be generated by shift register, and the operating frequency range is 100 MHz to 3 GHz at CMOS 90 nm process simulation.

Keywords Duty cycle control · Pulsewidth lock loop · PWLL · Delay line

1 Introduction

In the high speed system on chip (SoC), clock signal will be affected by process, voltage, and temperature variations, then the clock distortion of previous stage makes the operation error with latter stages. Thus, the requirement of clock signal, the stability and adjustable of phase and width becomes higher, i.e. high-speed dynamic circuit, double sampling system, or positive and negative edge trigger circuit. Therefore, the adjustable and stable pulsewidth lock loops (PWLL) is an important sub-system for SoC. The PWLL research types divided into mixed-signal [1–4, 9] and all-digital [4–8]. The element parameters of mixed-signal structure are tuned difficultly at nanometer process or process variation. To adapt the current process, the rapid changed digital SoC chip design and nanometer process, therefore, this paper designed for an all-digital circuit. In all-digital architecture [5–8],

P.-H. Yang (✉) · J.-M. Chen · Z.-M. Hong

National Yunling University of Science and Technology, Douliu, Taiwan
e-mail: phyang@yuntech.edu.tw

the numbers of delay line are limited for the pulsewidth measurement and generation; consequently, it cannot operate at low frequency, except for increasing the delay line stages with larger circuit. This paper presents a new cyclic delay line structure which improves the operational frequency with fair circuit area.

2 Traditional All-Digital Pulsewidth Lock Loops and Duty Cycle Corrector

Pulsewidth lock provides 50 % and others pulsewidth selection, but duty cycle lock circuit only provides 50 %, between these two circuit, not only the different pulsewidth selection, but also the structure complexity. The former one is more complicated, and the other is simpler.

Figure 1 shows the traditional all-digital PWLL architecture [5], which composed of digital control delay line (DCDL), time-to-digital converter (TDC), pulsewidth comparator (PC), and up/down counter (UDC). The operation flow starts from the desire pulsewidth selection, then DCDL generates the initial outputs (CKoutput), subsequently, the output signal transfers to TDC for digital quantization measurement. In a result, the pulsewidth becomes digital code with ‘0’s and ‘1’s, and PC will compare the digital code with desired pulsewidth code, after that, the pulsewidth code is tuned by UDC and transmits to DCDL for pulsewidth control, consequently, repeats the operations until the output and desired pulsewidth are identical, finally, the pulsewidth will be locked.

The circuit architecture [6], as shown in Fig. 2, which composed of pulse generator (PG), half cycle time delay line (HCDL), and matching delay line (MDL). When the HCDL delay line completes the measurement, the negative edge time of S and R become half a clock cycle. The triggered SR latch generate a 50 % duty cycle. The operation frequency will be affected by delay line length of TDC and DCDL in Fig. 1, and so does HCDL in Fig. 2, in this paper, we propose a cyclic delay line to increase operation frequency range.

Fig. 1 Traditional all-digital pulsewidth lock loops block diagram

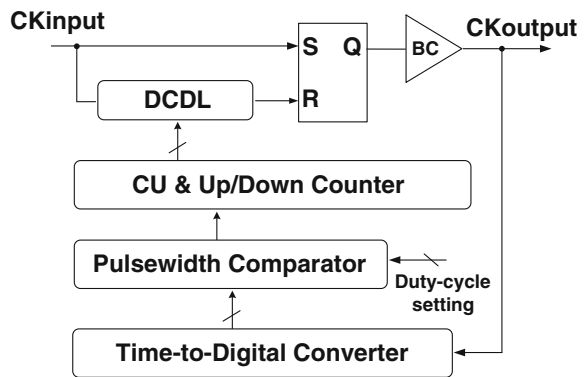
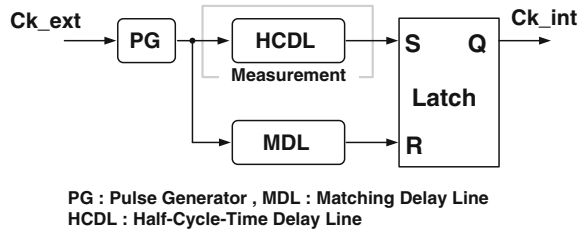


Fig. 2 All-digital duty cycle corrector block diagram



3 Operation Principle and Circuit Design

This paper proposed a major architecture, shown in Fig. 3, which composed of cyclic delay line binary pulsewidth detection and binary pulsewidth generator. The measuring mechanism of PWLL combined with cyclic binary pulsewidth detection (CBPD), 4-bit encoder, 7-bit counter, pulsewidth code shift register, control unit (CU), 10-bit comparator, and subtractor and adder. And the pulse generation mechanism composed of cyclic MUX cell delay line, cyclic matching delay,

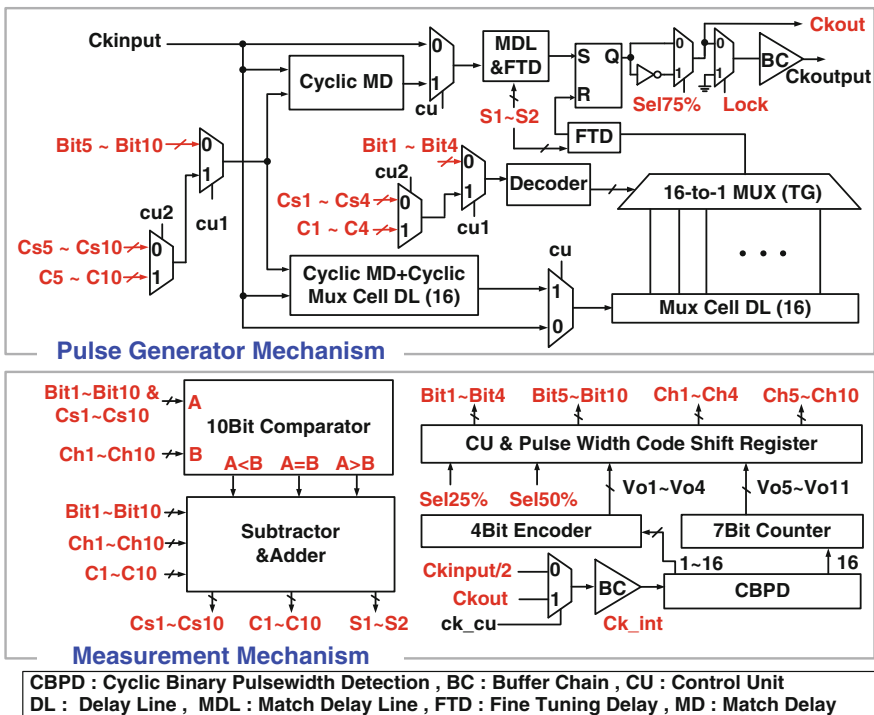


Fig. 3 All digital wide-range pulsewidth lock loops block diagram

matching delay line (MDL), 16-1 MUX, fine tuning delay (FTD), MUX cell delay line, buffer chain (BC), and SR latch.

The operation, at first, presets the external MUX ck_cu , $cu1$, and $cu2$ to '0', and divided the input signal ($CKinput$) into 2 ($CKinput/2$), the processed pulsewidth places pulsewidth of the original signal, and transfers to CBPD for pulsewidth detection. The CBPD combined with 16 delay cells, in the detection operation, the cyclic times of higher 7-bit cycle code (16, 32, 64, ..., 2048) are recorded by 7-bit counter, however, if the sum of quantization less than 15, the low 4-bit cycle code (1, 2, 4, 8) will be encoded by 4-bit encoder which are transmitted into shift register, and generate the 50 or 25 % pulsewidth code (Bit1–Bit10) through the divide-by-2 or divide-by-4 shift register. In the pulse generation mechanism, we adopt an identical delay line with CBPD, and each delay line composed of 16 MUXs with relative same delay time. The cyclic MUX delay line receives the higher 7-bit width code, and MUX delay line accepts lower 4-bit.

When this circuit operates at the high frequency range (>1 GHz), cu signal will become '0', simultaneously, the pulse generation mechanism only utilize one MUX delay line, thus, the detection also repeats only one times. If operation frequency exceeds 2 GHz, the pulsewidth detection error increase, because the delay cell of MUX delay line is also contribute the delay time. Therefore, we turn off the repeat detection mechanism while operation frequency higher than 2 GHz (Fig. 4).

Figure 5 shows the main architecture of CBPD, which composed of 16 MUXs and 17 DFFs, the 16th delay element will trigger D(16) and D(17) if pulsewidth larger than the whole delay time of 15 elements, simultaneously, D(17) becomes a narrow pulse generator, after that, 7-bit counter is triggered by the higher cycle code. Due to the detected pulsewidth only spends a cycle; the first quantization cycle value could be filtered by digital cycle filter. Figure 6 shows the cyclic

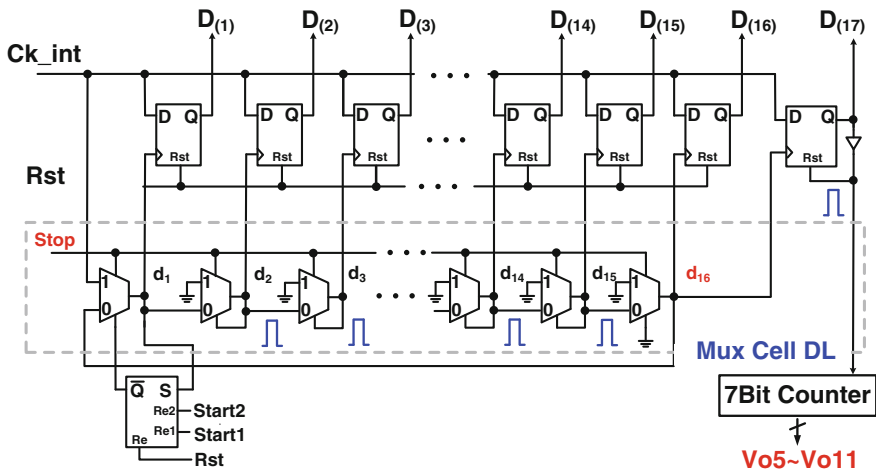


Fig. 4 Cyclic delay line binary pulsewidth detection circuit

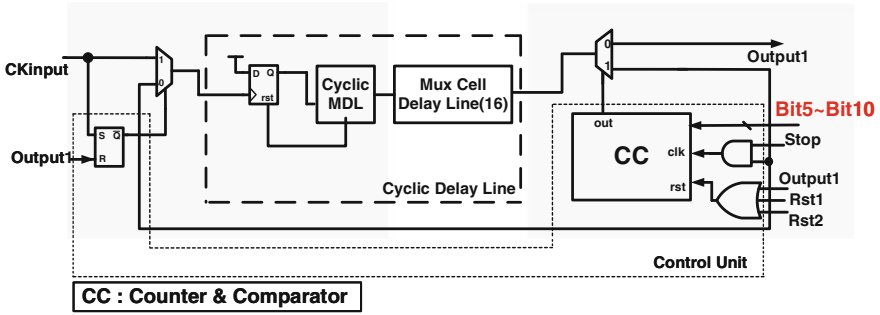


Fig. 5 Cyclic delay line circuit design

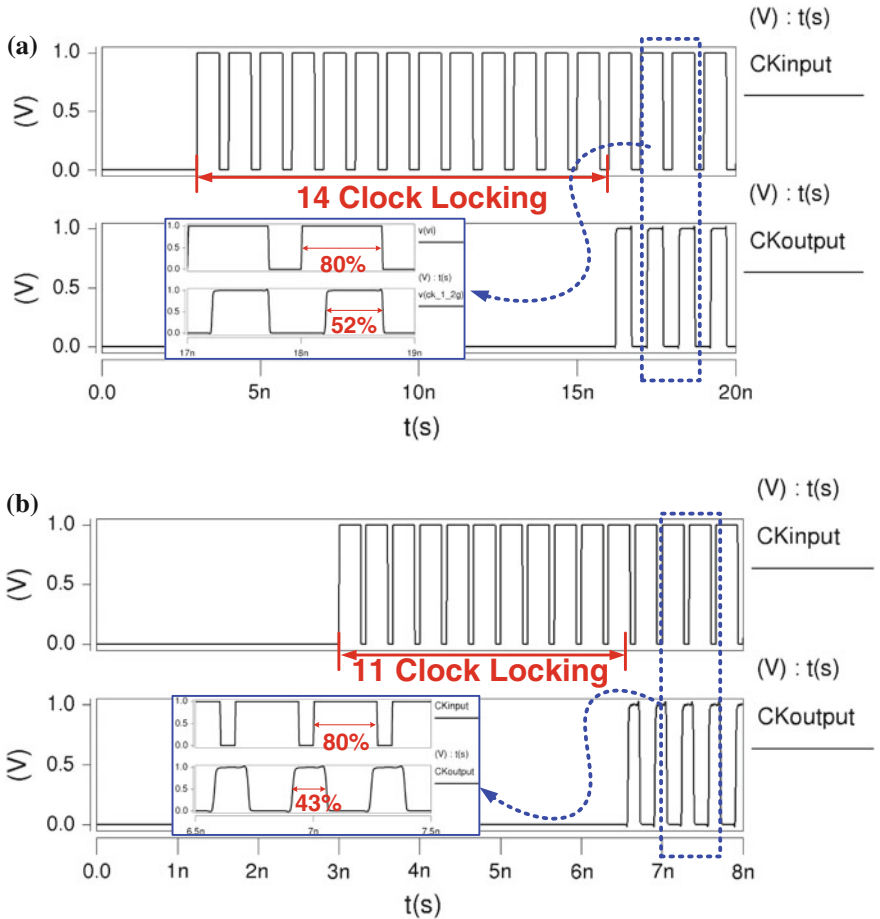


Fig. 6 Input 80 % duty cycle and output setting at 50 %, under operation frequency a 1 GHz, b 3 GHz

structure, all the control elements causes delay, therefore, the lump delay are not the same as prediction of 16 delay elements, for this reason, all the control elements are designed for matching with delay time.

Subsequently, the difference makes the delay time cannot be the same as the default binary values, and this situation occurs in the second detection. While the carry and borrow are happened at higher 7-bit and lower 4-bit, which cause the discontinuous binary value, thus the 3rd time detection is needed for the value correction. When operating frequency over 1 GHz, the lower 4-bit will turn off the cyclic delay line for saving power dissipation, then output pulse is adjusted by the fine tuning delay circuit, subsequently, output will be locked just at the 2nd time detection.

4 Simulation Results

The circuit is simulated at 90 nm CMOS process, operation frequency range is 100 MHz to 3 GHz, duty cycle could be locked at 25, 50, and 75 %. The Fig. 6a, b show operating at 1 and 3 GHz respectively, and the input duty cycle is 80 %, output duty cycle is set at 50 %. When operation frequency exceeds 1 GHz, it outputs through MUX delay line, subsequently, the duty cycle is locked in less than two times detection, among these operation frequency, 1 GHz only costs 14 duty cycles, the locked error is 2 %, and 3 GHz costs 11 duty cycles to lock, the locked error is 7 %. The Table 1 is the performance comparison with pervious studies.

Table 1 Performance comparison

Criteria	[7]Measured	[8]Simulation	[9]Simulation	This work
Technology	0.18 μm	90 nm (1.2 V)	0.13 μm	90 nm (1 V)
Architecture	Digital	Digital	Mixed analogy	Digital
Correction range	10–90 % @400 MHz 20–80 %@ 2 GHz	20–80 %	1–99 %	20–80 %
Correction duty cycle	50 %	50 %	10–90 %	25, 50, 75 %
Operation frequency	400 MHz to 2 GHz	500 MHz to 1 GHz	770 MHz to 1.05 GHz	100 MHz to 3 GHz
Locking time	Max@3.5 clock	46 clock	Max@40 ns	Max@25 clock
Resolution	70 ps	10 ps	N/A	30 ps
Duty cycle deviation	35 ps	50 \pm 0.5 %	N/A	Max@30 ps

5 Conclusion

The proposed binary cyclic detecting mechanism all-digital wide-range pulsewidth lock loops which achieved locking range from 100 MHz to 3 GHz operation frequency, and maximum locking cycle is 25. The new mechanism on low frequency operational pulsewidth measures and lock by using cyclic delay line loops. In the proposed design it can operate at 100 MHz. Furthermore, under high frequency operation the proposed new architecture has no complex loop control, and simplifies the accumulation delay works under high frequency operation. The pulsewidth generation mechanism adopted two delay lines architecture, one is combined with a complex cyclic delay line, the other delay line is selected delay time by MUX. When operating in high frequency, MUX directly outputs a preset pulsewidth. For conversion pulsewidth into digital code, we design delay line circuit in Hexadecimal, and use pulsewidth of continuous cyclic delay line detecting circuit, the wanted output pulsewidth code will get in three clock cycles. This paper has 7 % error under highest frequency operation. The wide-range performance of proposed PWLL agrees with the simulation results.

References

1. Cheng KH, Su CW, Wu CL, Lo YL (2004) A phase-locked pulsewidth control loop with programmable duty cycle. In: IEEE AP-ASIC, pp 84–87
2. Han SR, Liu SI (2004) A 500-MHz–1.25-GHz fast-locking pulse width control loop with presentable duty cycle. IEEE J Solid-State Circuits 39:463–468
3. Jovanović G, Mitü D, Stojčev M (2006) An adaptive pulsewidth control loop. IEEE Microelectron 626–629
4. Navidi MM, Abrishamifar A (2011) A fast lock time pulsewidth control loop using second order passive loop filters. In: IEEE ICEE, pp 1–5
5. Wang YM, Hu CF, Chen YJ, Wang JS (2005) An all-digital pulsewidth control loop, In: IEEE symposiums on VLSI circuits 2, pp 1258–1261
6. Wang YM, Wang JS (2004) An all-digital 50 % duty-cycle corrector. In: IEEE on ISCAS, pp 925–928
7. Gu JH, Wu JH, Gu DH, Zhang M, Shi LX (2012) All-digital wide range precharge logic 50 % duty cycle corrector. IEEE Trans VLSI Syst 20:760–764
8. Swathi R, Srinivas MB (2009) All-digital duty cycle correction circuit in 90 nm based on mutex. In: IEEE Computer Society annual symposium on VLSI (2009) 258–262
9. Weng RM, Lu YC, Liu CY (2009) A low jitter arbitrary-input pulsewidth control loop with wide duty cycle adjustment. In: IEEE ISCAS, pp 1301–1304

A BUS Topology Temperature Sensor Cell Design with System in Package Application

Po-Hui Yang, Jing-Min Chen and Ching-Ken Chen

Abstract This paper presents multi-chip temperature sensing technique with simple identification circuits for system in package (SiP) application, to monitor temperature of each functional chip on a system package substrate. The temperature sensor is activated by a simple decoding circuit, and then sends back the temperature dependent pulse. Moreover, by using the BUS topology the control unit connects every sensor chips with only two wires to transceive data, and the routing complexity will not be increased with increased the number of sensor chips. This circuit has implemented in 0.18 μm CMOS process, chip area is 0.02 mm^2 .

Keywords Temperature sensor · System in package · Multi-point temperature sensing

1 Introduction

System-in-package (SiP) are composed of function chips and passive components in a single package for improving performance, cost down, and reducing time of products onto the market. However, under the system design and integration, package structure is getting more and more complex, and integrates chips in a single package will significantly increase power. In addition, SiP towards miniaturization, which integrates more chips causes power density increases rapidly, therefore, heat problem becomes more difficult to solve, and heat spot is generated, which leads to package stress problem, consequently, the temperature issue becomes a much attention research topic. In the SiP architecture, besides the heat dissipation should be enhanced, the temperature monitoring is also very important.

The previous studies on embedded temperature sensors can be divided into digital, analog, and mixed signal. Analog type [1–6] has advantages of high resolution and low error, but it alone with large area, higher design difficulty, and not

P.-H. Yang (✉) · J.-M. Chen · C.-K. Chen
National Yunlin University of Science and Technology, Douliu, Taiwan
e-mail: phyang@yuntech.edu.tw

easily to be integrated into the digital system IC. Mixed signal [7, 8] also has large area, and poor temperature linearity. Digital structure has low power dissipation, high sampling rate, and low design complexity, and appropriate for multi-point sensing design. Especially, oscillator based temperature sensor [9, 10] will vary its frequency due to the carrier mobility and MOS transistor threshold voltage changed by the temperature. Basically, the MOS threshold voltage will drop when temperature raising, which makes the delay time increases and oscillation frequency decreases, conversely, the temperature goes down, and oscillation frequency increases, therefore, ring oscillator suits for adopting in multi-point sensing design with SiP.

2 The BUS Topology Multi-Point Temperature Sensing

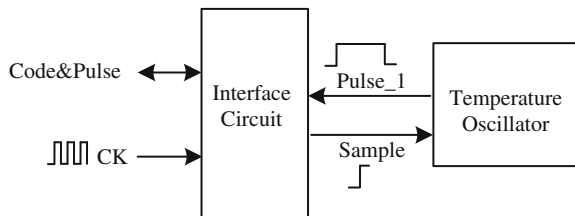
The BUS topology multi-point temperature sensor with SiP application chips schematic diagram is combined with a μ -controller and several temperature sensors. This proposed sensor's simple block diagram shows in Fig. 1, which composes of interface circuit and temperature sensor.

When temperature detection starts, as shown in Fig. 2, at first, the μ -controller will reset each temperature sensor, subsequently, the μ -controller sends a unique identification code of wanted test sensor, then CK1 and identification code will be synchronized by timing control circuit, the output of SIPO is also transmitted to register simultaneously, after four clocks, the identification code is loaded, then it will match with identification circuit.

When identification code is matched successfully, the sample signal will be sent to the sensor. After the sampling complete, temperature dependent pulsewidth will be sent back to timing control circuit, then it transfers to μ -controller by Code&Pulse for time-to-digital converting, and these are a complete temperature sensing operations. The circuit operating timing diagram shows in Fig. 3.

Interface circuit composes of timing control circuits, counter_4, serial-in-parallel-out (SIPO), and identification circuit. By timing control circuit, the initial value of circuit is

Fig. 1 BUS topology temperature sensor block diagram



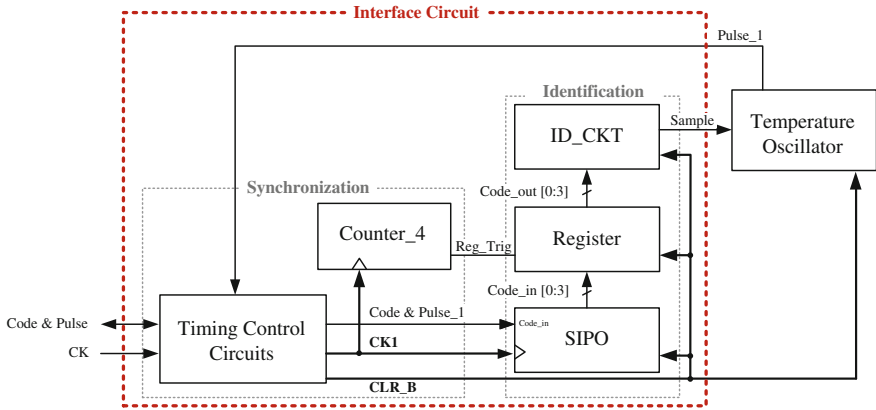


Fig. 2 The block diagram of proposed temperature sensor

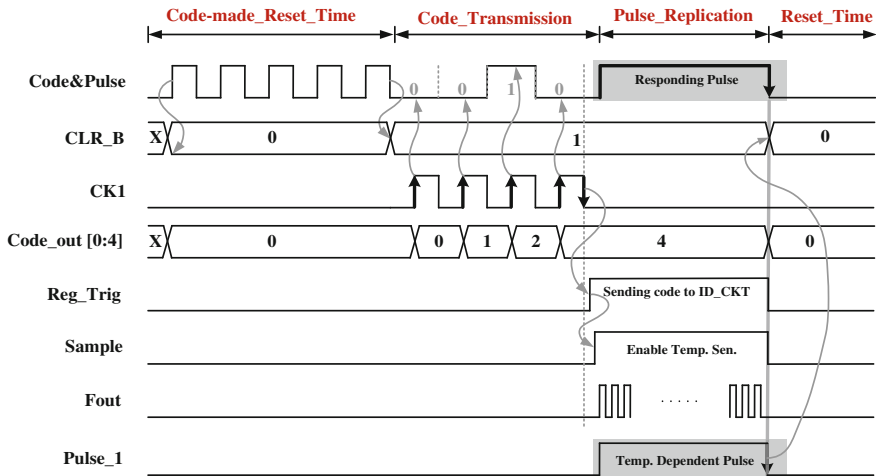


Fig. 3 The timing diagram of temperature sensing

cleared. The identification signal and CK are synchronized, then four clocks later the codes are identified by identification circuit, and sample signal is sent to activate sensor. Once codes matched, subsequently, temperature dependent pulse is sent back to timing control circuit. In this SiP case, the temperature dependent pulse will send to a μ -controller through the same wire Code&Pulse, these are a completely temperature sampling.

3 Implementation and Simulation Results

For a SiP application, the timing control circuit is shown in Fig. 4, before temperature monitoring, the μ -controller will send a reset signal to clear the initial value of whole chip, and timing control circuit will also synchronize the identification code and CK to make sure the unmatched sensor will not be activated.

Figure 5 is the identification circuit. When identification code and CK are synchronized, the code has to process with SIPO. At the same time, the processed

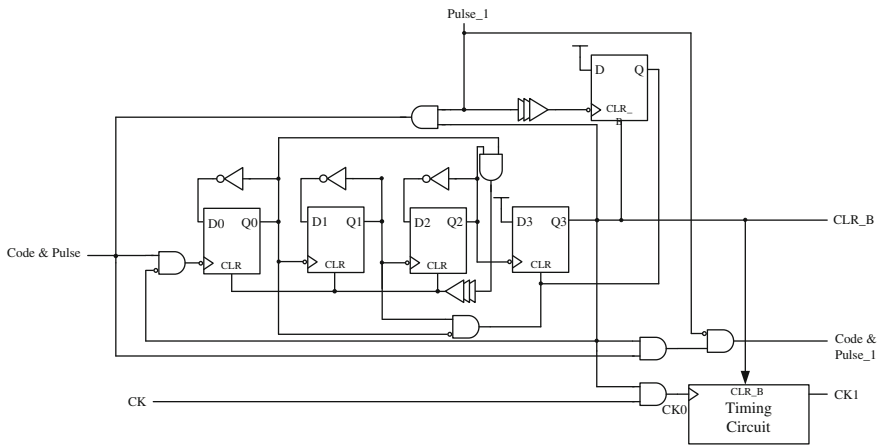


Fig. 4 Timing control circuit

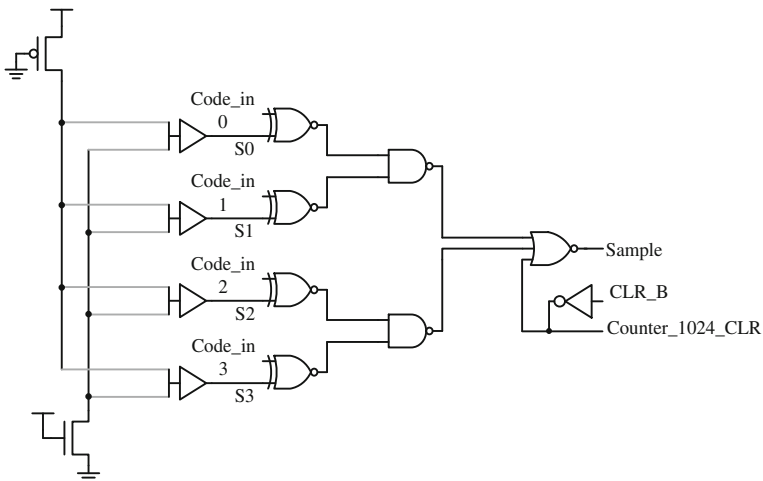


Fig. 5 The identification circuit

value will be saved into register. In this SiP case, each temperature sensor chip has to fabricate independently, therefore, the logic state S0, S1, S2 and S3 are defined with broken connection by laser cutter. The output digit will be serially sent out with SIPO, afterward, the sampling signal is sent to temperature sensors.

Temperature oscillator, shows in Fig. 6a, which composes with tri-state gate, positive edge DFF (PEDFF), oscillator, and counter_1024. The PEDFF is triggered when the Sample signal goes from '0' to '1', and Pulse_1 also turns into '1', then oscillator is activated. The temperature varying frequency (Fout) will be counted by a counter, and clear signal (CLR) is generated when counting achieves 1024 times, and the CLR signal leads to PEDFF's output '0', at the same time, Pulse_1 also changes from '1' to '0', then temperature dependent pulse will be sent back, and these are one time sensing operation, the timing diagram shows in Fig. 6b. The circuit senses temperature by the ring oscillator and their output is a temperature dependent frequency, then it soon be resolved into a digitalized code.

Table 1 is the performance comparison of traditional temperature sensors and the proposed BUS topology multi-point temperature sensor. According to the post-layout results, the proposed temperature sensor has a high sampling rate, and it reduces routing complexity, which needs only two wires to transmit signal, therefore, this paper is properly at monitoring multi-point temperature detection.

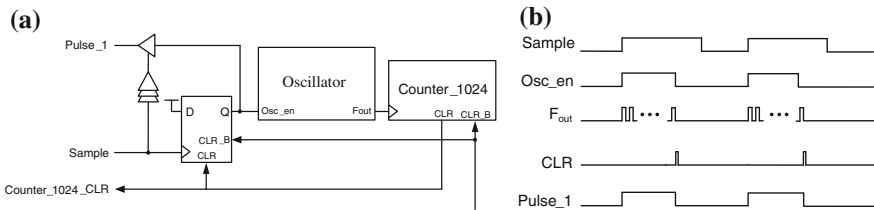


Fig. 6 Oscillator based temperature sensor a block diagram b timing diagram

Table 1 The performance comparison

Ref	[11]	[12]	[13]	[14]	[15]	This work
Error (°C)	-1.0 to +0.8	-1.8 to +2.3	-0.25 to +0.35	-5.1 to +3.3	±2	±3.99
Resolution (°C)	N/A	0.66	0.09	0.135	0.5	0.17
Power (µW)	25	N/A	36.7	15	60	220.7
Temp. range (°C)	50-125	0-100	0-90	0-60	20-80	0-100
Area (mm ²)	0.047	0.12	0.6	0.01	0.002	0.02
Sampling rate(Hz)	N/A	5 K	2	10 K	N/A	24.39 K
Process	90 nm	0.13 µm	0.35 µm	65 nm	65 nm	0.18 µm
Routing wire	N/A	N/A	N/A	N/A	N/A	2
Applications	SoC	SoC	SoC	SoC	SoC	SiP

4 Conclusion

This paper proposed a multi-point temperature sensor for SiP application, and utilized a μ -controller to connect all temperature sensors. A simple identification circuit which enabled by decoding code and sends back a temperature dependent signal for monitoring each block of package substrate. Owing to the proposed BUS topology, the routing complexity is reduced significantly, and it needs only two wires for control and temperature dependent signal transmission. Therefore, in this SiP case, a μ -controller with extra-small package is adopted to receive and convert the temperature signal. This paper presents an all-digital circuit structure; therefore it can easily integrate with other kinds of SiP chips, and properly implements the multi-point detection. Finally, it accurately monitors temperature variation at every one of blocks in SiP chips. This circuit is implemented in 0.18 μm CMOS process, chip area is 0.02 mm^2 , and ready for carrying out the SiP applications.

References

1. Bakker A, Huijsing JH (1996) Micropower CMOS temperature sensor with digital output. *IEEE J Solid-State Circuits* 31:933–937
2. Sanchez H, Phlip R, Alvarez J, Gerosa G (1997) A CMOS temperature sensors for PowerPC RISC microprocessors. In: *VLSI circuit, digest of technical symposium on IEEE CNF*, pp 13–14
3. Tuthill M (1998) A switched-current, switched-capacitor temperature sensor in 0.6 μm CMOS. *IEEE J Solid-State Circuits* 33:1117–1122
4. Pertijs MAP et al. (2001) A high-accuracy temperature sensor with second-order curvature correction and digital BUS interface. In: *Proceedings of IEEE ISCAS*, vol 1, pp 368–371
5. Pertijs M, Niederkon A, Xu M, Mckillop B, Bakker A, Huijsing J (2003) A CMOS temperature sensor with a 3σ inaccuracy of ± 0.5 $^{\circ}\text{C}$ from -50 $^{\circ}\text{C}$ to 120 $^{\circ}\text{C}$. In: *IEEE ISSCC Dig. Tech. Papers*, vol 1, pp 200–201
6. Pertijs M, Makinwa K, Huijsing J (2005) A CMOS temperature sensor with a 3σ inaccuracy of ± 0.1 $^{\circ}\text{C}$ from -55 $^{\circ}\text{C}$ to 125 $^{\circ}\text{C}$. *IEEE ISSCC Dig Tech Pap* 1:238–239
7. Chen P, Chen CC, Tsai CC, Lu WF (2005) A time-to-digital- converter-based CMOS smart temperature sensor. *IEEE J Solid-State Circuits* 40:1642–1648
8. Ituero P, Ayala JL, Marisa LV (2007) Leakage-based on-chip thermal sensor for CMOS technology. In: *Proceedings of IEEE ISCAS*, pp 3327–3330
9. Demassa TA, Ciccone Z (1996) *Digital integrated circuit*. Wiley, New York
10. Falanovsky IM, Allam A (2001) Mutual compensaetion of mobility and threshold voltage temperature effects with application in CMOS circuits. *IEEE Trans Circuits Syst I Fundam Theor Appl* 48:876–884
11. Sasaki M et al (2008) A temperature sensor with an inaccuracy of $-1/+0.8$ $^{\circ}\text{C}$ using 90-nm 1-V CMOS for online thermal monitoring of VLSI circuits. *IEEE Trans Semicond Manuf* 21:201–208
12. Woo K, Meninger S, Xanthopoulos T, Crain E, Ha D, Ham D (2009) Dual-DLL-Based CMOS all-digital temperature sensor for microprocessor thermal monitoring. In: *IEEE ISSCC*, pp 68–69
13. Poki C et al (2010) A time-domain SAR smart temperature sensor with curvature compensation and a 3σ inaccuracy of -0.4 $^{\circ}\text{C}$ to $+0.6$ $^{\circ}\text{C}$ over a 0 $^{\circ}\text{C}$ to 90 $^{\circ}\text{C}$ range. *IEEE J Solid-State Circuits* 45:600–609

14. Ha D, Woo K, Meninger S, Xanthopoulos T, Crain E, Ham D (2011) Time-domain CMOS temperature sensors with dual delay-locked loops for microprocessor thermal monitoring. *IEEE Trans Very Large Scale Integr Syst* 20(9):1590–1601
15. Xie S, Ng WT (2012) A 0.02 nJ self-calibrated 65 nm CMOS delay line temperature sensor. *IEEE Trans Circuits Syst* 3126–3129

The Off-Axis Parabolic Mirror Optical Axis Adjustment Method in a Wedge Optical Plate Lateral Shearing Interferometer

Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee and Ya-Hui Hsieh

Abstract The optical alignment of an off-axis parabolic (OAP) mirror has been successfully developed using the portable alignment device, the wedge optical plate lateral shearing interferometer and the CCD camera. In this method the optical axis of an OAP mirror is made parallel to the “five incident parallel laser beams” in the plane of incidence, by checking direction of these five reflected laser beams and changing the height and orientation of the OAP mirror. The lateral interferometer is referred to as the layout where opposite beams travel in difference directions, encountering exactly the same components until they emerge to form interference pattern. The lateral shearing interferometer uses to examine the parallel of five incident parallel laser beams. This fast aligning method for finding the optical axis of an OAP mirror can measure the Slant Focal Length deviation to an accuracy of 0.5 %.

Keywords Off-axis parabolic mirror · Optical axis adjustment

1 Introduction

The OAP mirror offers the advantage of an unobstructed aperture and minimizes system size, giving access to evaluate the performance of optical system testing and to ensure high resolution even in a short focal length compact spectrometer. The OAP mirrors shown in Fig. 1 are especially suitable for broadband and multiple wavelength applications due to their completely achromatic characteristics. When collimated light is incident parallel to the optical axis, a concave parabolic surface

F.-M. Yeh (✉) · S.-C. Lee

Department of BioIndustry Technology, Da Yeh University, Dacun, Taiwan
e-mail: optfmy@yahoo.com.tw

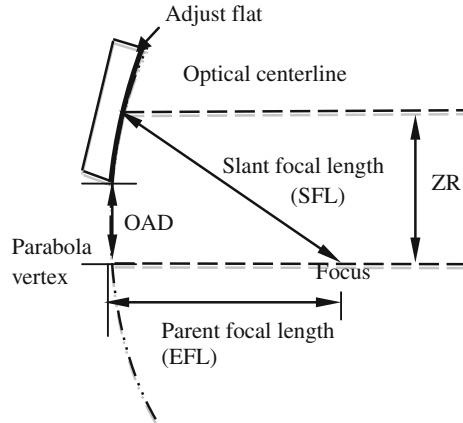
D.-C. Chen

Department of Electrical Engineering, Feng Chia University, Taichung, Taiwan

Y.-H. Hsieh

Department of Healthcare Information and Management, Ming Chuan University, Taipei, Taiwan

Fig. 1 Off-axis parabolic mirror



focuses the light into an excellent corrected point on the optical axis. The OAP mirror is a segment cut out of a large parent parabola. As the OAP mirror becomes more complex, precise optical axis alignment becomes more challenging. Mirror manufacture and alignment usually employ a sophisticated and expensive interferometer, such as laser unequal path interferometer, autocollimator, knife edge method, to measure the OAP mirror's characteristics, such as focal length, off-axis distance, and optical axis [1, 2]. Aligning optical axis of OAP mirror not only takes much of time, but also these methods cannot be used in UV and IR region. Because of the above problems, we establish a new optical technique to align the optical axis of an OAP mirror using the five parallel laser beams, laser triangulation range finder, the lateral interferometer and CCD camera. There are some advantages in this method: (1) It is a rapid and simple alignment method. (2) It simplifies the alignment optical system and lowers the cost. (3) The lateral shearing interferometer uses to examine the parallel of five incident parallel laser beams. (4) Because the rotation mechanism of the five parallel laser beams arrangement, it adjusts the OAP mirror with more freedoms of orientation. This is a fast method for aligning the optical axis of an OAP mirror and can measure the SFL deviation to accuracy of 0.5 %.

2 Basic Principle

The structure of the OAP mirror is shown in Fig. 1. The OAP mirror is specified as following. Parent focal length (PFL) is the focal length of the parent parabola. It defines the shape of the surface as $Z = ZR^2/(4 \cdot PFL)$, where ZR is radial distance from vertex and Z is sagittal depth of the surface. Slant Focal Length (SFL) is the distance between OAP mirror's mechanical center and parabola focus. Optical centerline is the line parallel to parent parabola optical axis and coming through the mechanical center of OAP. Before proceeding to measure the SFL and ZR of an

OAP mirror, it is necessary to align it correctly. An OAP mirror that is not aligned accurately enough due to the aberrations of the alignment reflected beam, the depth of focus and the size of the focal spot. In an ideal optical system, all rays of light from a point in the object plane would converge to the same point in the image plane, forming a clear image. The influences which cause different rays to converge to different points are called aberrations.

However, the parabola is not completely free of aberration; it has both coma and astigmatism. Since it has no spherical aberration, the position of the stop dose not change the amount of coma, which is given by Eq. 1.

$$Coma_s = \frac{u_p}{16(f/\#)^2} \quad (\text{Unit: radians}) \tag{1}$$

The amount of astigmatism is modified by the stop position. With the stop at the mirror the stigmatism is given by Eq. 2.

$$Astigmatism = \frac{u_p^2}{2(f/\#)} \tag{2}$$

Lateral shearing interferometry has been used extensively in versatile applications such as the testing of optical components and systems. It consists of duplicating wave front under study, displacing it laterally by a small amount, and obtaining the interference pattern between the original and the displaced wave fronts. In this paper, the parallel of five incident parallel laser beams was examined by the wedge optical plate lateral shearing interferometer. The collimation of laser light is done by adjusting whatever collimating optics is being used until the fringes produced by the test device are observed to be aligned. With perfectly plane or spherical wave fronts, straight fringes are obtained. Any departure from straightness or wiggles is indicative of aberrations in the system. The collimation test device consists of a piece of high quality BK-7 with very flat surfaces having a slight wedge angle between them. When a plane wave is incident at an angle of 45°, two reflected wave fronts result. These are separated laterally because of the plate thickness and angularity due to the wedge. The lateral separation is referred to as shear which is why the device is referred to as a shearing interferometer. With plane wave fronts incident, the area of overlap between the two reflected beams will show fringes when projected on a screen. The fringes will appear solely from the wedge angle and their spacing will be $d_f = \lambda/(2n\theta)$ where d_f is the fringe spacing, λ is the wavelength, n is the refractive index and θ is the wedge angle as shown at Fig. 2. If the laser beam is not perfectly collimated, the orientation of the fringes varies. When a non-collimated laser beam incident on a wedged optical plate, the path difference between the two reflected wavefronts is increased or decreased from the case of perfect collimation depending on the sign of the curvature. The pattern is then rotated and the beam's wavefront radius of curvature will be $R = (s * d_f)/[\lambda * \sin(\gamma)]$ where γ is the angular deviation of the fringe alignment from that of perfect collimation.

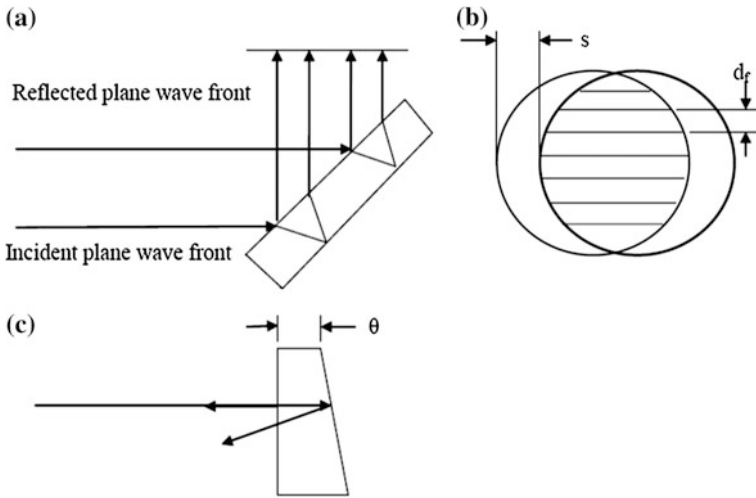
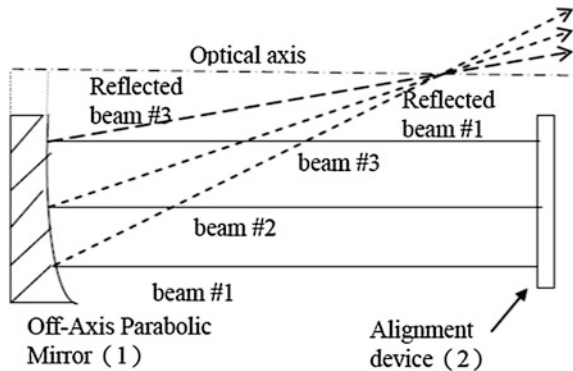


Fig. 2 Collimation test device **a** lateral shearing interferometer set up **b** fringe pattern **c** wedge optical plate

3 The Optical Alignment System

We develop a specific technique to align the optical axis of the OAP mirror. This alignment system can be applied to UV–IR regions by using the different wavelengths of alignment laser diode. The optical alignment system in Fig. 3 is composed of: (1) alignment device shown in Fig. 4, (2) the wedge optical plate lateral shearing interferometer shown in Fig. 5 is composed of wedge optical plate and OAP mirror. The Z is optical axis, (3) CCD Image camera and pinhole, and (4) optical table and mount (part (4) and (5) not shown). These are described as followings.

Fig. 3 The optical alignment system



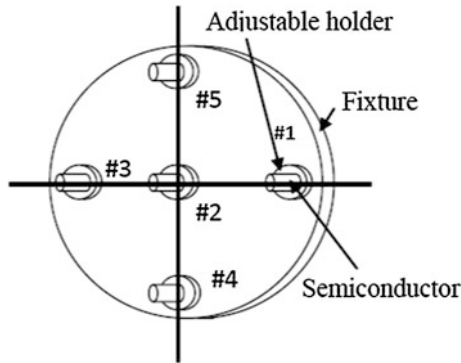


Fig. 4 The schematic of the alignment device

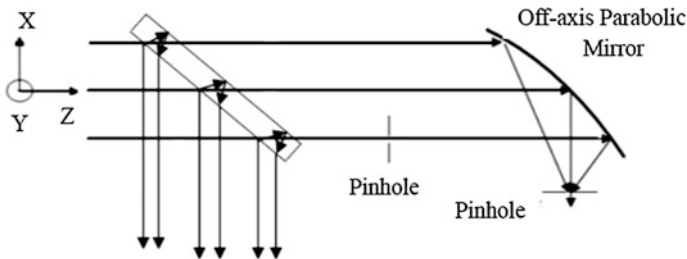
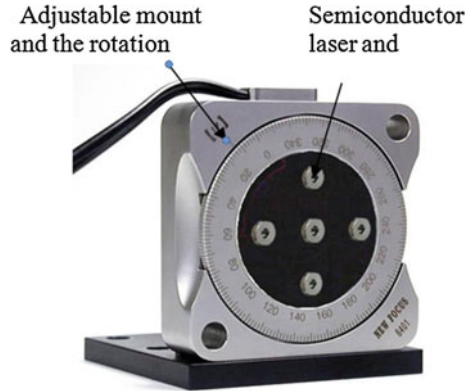


Fig. 5 The wedge optical plate lateral shearing interferometer for off-axis parabolic mirror optical axis adjustment

The portable alignment device includes five parallel semiconductor lasers, adjustable mount and the rotation mechanism, and precise adjustable holder shown in Fig. 6. The semiconductor laser beams on the fixture are made parallel each other with precise adjustable holder and perpendicular to the surface fixture. An adjustable holder for mounting and orienting the semiconductor laser is set by a removable retaining laser within a mount of ring.

The rotation mechanism adjusts the initially horizontal and vertical laser beams to any rotation angle α as shown in Fig. 6. For example, the five laser beams lines, i.e., #1, #2, #3, #4 and #5 from right to left and down to up in Fig. 6. When it clockwise rotates α degree, it shows #1, #2, #3, #4 and #5 from original direction to α direction in Fig. 6. In the rotation mechanism, it can rotate any α degree to adjust the OAP mirror orientation we want.

Fig. 6 Photograph of the portable alignment device

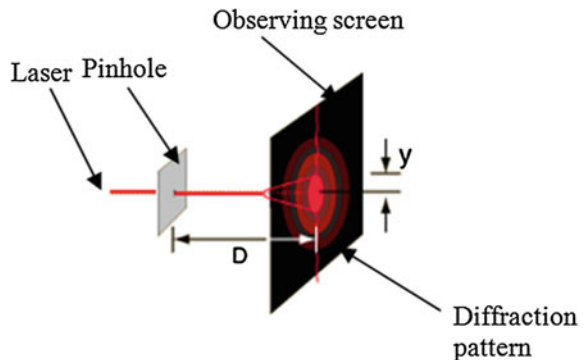


The pinhole has a size of 30 μm for alignment five laser beams and OAP mirror. When the diffraction pattern of concentric circular rings was appeared, the laser beam passes through the pinhole accurately. The diffraction patterns formed by a pinhole consist of a central bright spot surrounded by a series of bright and dark rings shown in Fig. 7. The diffraction experimental set-up includes semiconductor laser, pinhole and observing screen. We can describe the pattern in terms of the angle θ , representing the radius angle of each ring. If the aperture diameter is d (mm) and the wavelength is λ (mm), the radius angle θ of the first dark ring is given by

$$\sin\theta = 1.22(\lambda/d) = y/D \tag{3}$$

The distance between the pinhole and screen is D . The actual diffraction pattern of alignment pinhole is shown in Fig. 8. Figure 8a show the horizontal laser beam incident on the pinhole along optical axis of pinhole and Fig. 8b show the oblique laser beam incident on the pinhole at slope angle α . With this diffraction technique, the pinhole is used to precise align laser beam parallel to the OAP of the alignment system.

Fig. 7 Pinhole diffraction pattern



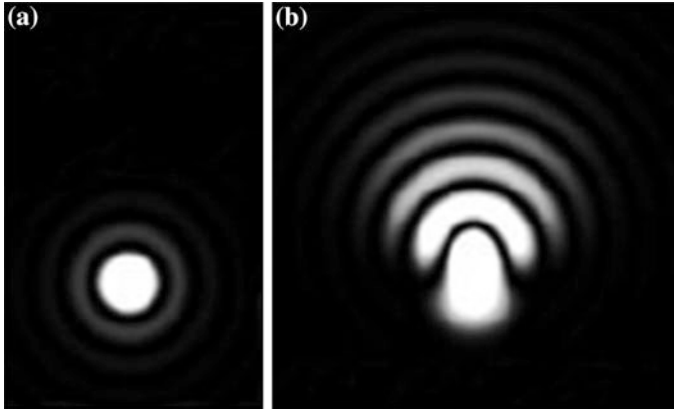


Fig. 8 The actual diffraction pattern for alignment pinhole **a** accurate alignment **b** alignment error

The OAP mirror is brought into focus onto the CCD camera by looking for the minimum spot diameter. The focal spot is then centered using the XYZ translation stage attached to the CCD camera. By varying the separation of the OAP mirror and detector, best focus is located when the displayed beam size is smallest.

4 Experiments and Results

The conditions and specifications used for the experiment are listed in Table 1. Experiment procedures are divided into two parts as follows:

4.1 *Pre-experiment of Five Semiconductor Laser Beams Fine Adjustment*

1. The fixture of alignment device is vertically (perpendicular) put on the optical table such that laser beam #2 is nearly parallel to the surface of the optical table
2. The precise adjustable holder is fine adjusted to make the laser beam #2 parallel to the surface of the optical table. To check the parallelism, we use the pinhole which moves on the optical table from position 1 to position 2 (about 50 cm distance, for SFL = 457 mm) with a fixed height and has observed the pinhole diffraction pattern that the laser beam passes through the pinhole accurately shown in Fig. 8a.
3. The plane mirror is put on the optical table at 3 m distance from the alignment device. To adjust the plane mirror, the reflected beam #2 is made parallel to the optical table. To check the plane mirror perpendicular to the optical table, we

Table 1 The experiment conditions

Off axis parabolic mirror	
Slant EFL	101.6 mm
Parent EFL	50.8 mm
Off-Axis angle	90°
Diameter	50.8 mm
Laser range finder	
Range of measurement	0.02 up to 60 m
Measuring accuracy	±1 mm
Time for a measurement	2.5–10 s
Light source	Red laser diode
Semiconductor laser of alignment device	
Wavelength (λ_p)	635 nm
Output power	3 mW
Beam divergence	<2 mrad
Operating current	25 mA
Beam diameter	3.3 mm
CCD camera	
Resolution	752 × 582
Spectral range	350–1100 nm
Wedge optical plate	
Optical substrate	BK7
Parallel	3 arc s
Diameter	55 mm

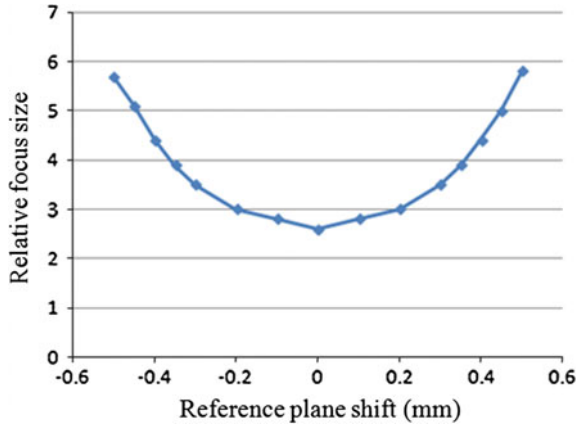
use a piece of paper to chop reflected beam #2, and we can easily see how well the reflected beam #2 overlap at the output hole of laser beam #2.

4. Step (2) is repeated until the beam #4, #5, #3 and beam #1 is parallel to the surface of the optical table.
5. We use a piece of paper to chop reflected beam #4, #5, #3 and #1, and we can easily see how well the reflected beam #4, #5, #3 and #1 overlap at the output hole of laser beam #4, #5, #3 and #1, respectively.
6. Finally, the parallel of five incident parallel laser beams can be examined by the wedge optical plate lateral shearing interferometer shown in Fig. 5.

4.2 Experiment with OAP Mirror

1. An OAP mirror is adjusted at first so that its optical axis is in the incident plane made by three horizontal parallel incident laser beams.
2. The OAP mirror is adjusted again so that the three horizontal parallel reflected beams come to a focus at the same point, i.e., the saggital optical plane of OPA is found.

Fig. 9 Focus size when reference plane shifted



3. Three vertical parallel incident beams are parallel to optical axis by adjusting the OAP mirror and the three vertical reflected beams also come to a focus at the same point obtained in (2), i.e., the tangential optical plane of OAP is found.
4. Capture the focus pattern at the reference plane (or near focal point) with the CCD camera. The reference plane is moved from left to right of the image plane to see if it fits small spot size.
5. The SFL of an OAP mirror and the off-axis distance are measured by laser ranger and trigonometry.

We measure best focus pattern curve that the reference focus size versus the position of the reference plane. Focus patterns sizes of reference plane are captured with CCD camera is shown in Fig. 9. The vertical coordinate is at different position of CCD camera. The relative focus size is normalized to reference focal point. The OAP mirror is brought into best focus onto the CCD camera by looking for the minimum spot diameter. The eleven focus patterns from left of the image plane to right of it are measured. The 30 % of the center of the maximum intensity is defined as the focus pattern of the OAP mirror.

The alignment of the OAP is finished when getting relative minimum focus size. This validates the experimental system. This alignment experiment repeatedly does twenty times. The test results show that the average SLF is 101.6 mm and the standard deviation is 0.755 mm or 0.5 %.

5 Conclusions

We have effectively presented the use of the five parallel laser beams alignment device, wedge optical plate lateral shearing interferometer and CCD camera to align the optical axis of OAP mirror. This system is designed for the optical engineers

needing lateral shearing interferometric and electro-optical practical techniques to achieve alignment of the off-axis optical systems. The advantages of the off-axis alignment techniques are: (1) it is simple to operate; (2) it simplifies the optical test system and lowers the cost [3], (3) it is less expensive to maintain the equipments, and (4) a simple phase shifting technique without moving element in lateral shearing interferometer.

References

1. Lee YH (1992) Alignment of off-axis parabolic mirrors with two parallel He-Ne laser beams. *Opt Eng* 31:2287–2292
2. Barkhouser R, Ohl RG (1999) Interferometer alignment and figure testing of large (0.5 m) off-axis parabolic mirrors in a challenging clean room environment. *Proc SPIE* 3782:601–614
3. Chrzanowski K (2007) Evaluation of infrared collimators for testing thermal imaging systems. *Opto-Electron Rev* 15:82–87

Two-Mirror Telescope Optical Axis Alignment by Additive Color Mixing Method

Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee and Ya-Hui Hsieh

Abstract In this paper, the two wavelength laser color alignment device, the additive color mixing method and the CCD camera is used to rapidly and accurately align the optical axis of a two-mirror telescope (TMT). In this method the optical axis of a TMT is made parallel to the “five incident parallel laser beams” in the plane of incidence, by checking direction of these five reflected laser beams and changing the height and orientation of the Two-mirror telescope. The two wavelength laser color alignment device emit five laser beams at two different visible wavelengths, including blue lasers with 405 nm at horizontal axis and the other red lasers with 645 nm at vertical axis. The combined dot will become magenta once parallel laser beams focus on focus point of two-mirror telescope and their beams overlap. The blue laser beam and red laser beam are added to the magenta light by the additive color mixing. The additive color mixing method uses to find the best focus and get minimum spot diameter of the TMT. This fast aligning method for finding the optical axis of a two-mirror telescope can measure the effective focal length deviation to an accuracy of 0.9 %.

Keywords Two-mirror telescope · Additive mixing color

F.-M. Yeh (✉) · S.-C. Lee
Department of BioIndustry Technology, Da Yeh University,
Changhua, Taiwan
e-mail: optfmy@yahoo.com.tw

D.-C. Chen
Department of Electrical Engineering, Feng Chia University,
Taichung, Taiwan

Y.-H. Hsieh
Department of Healthcare Information and Management,
Ming Chuan University, Taipei, Taiwan

1 Introduction

The two-mirror telescope offers the advantage of a large aperture and minimizes system size, giving access to play an important role in astronomy and other applications. The TMT shown in Fig. 1 are especially suitable for broadband and multiple wavelength applications due to their completely achromatic characteristics. When collimated light is incident parallel to the optical axis, a concave parabolic surface focuses the light into an excellent corrected point on the optical axis. As optical alignment of the TMT becomes more complex, precise optical axis alignment becomes more challenging. Mirror manufacture and alignment usually employ a sophisticated and expensive interferometer, such as laser unequal path interferometer, autocollimator, knife edge method, to measure the TMT's characteristics, such as effective focal length, back focal length, and optical axis [1, 2]. Aligning optical axis of TMT not only takes much of time, but also these methods cannot be used in UV and IR region. Because of the above problems, we establish a new optical technique to align the optical axis of a two-mirror telescope using the five parallel laser beams, laser triangulation range finder, the additive color mixing method and CCD camera. There are some advantages in this method: (1) It is a rapid and simple alignment method. (2) It simplifies the alignment optical system and lowers the cost. (3) The additive color mixing method uses to find the best focus and get minimum spot diameter of the two-mirror telescope. (4) Because the rotation mechanism of the five parallel laser beams arrangement, it adjusts the TMT with more freedoms of orientation. This is a fast method for aligning the optical axis of a two-mirror telescope and can measure the EFL deviation to accuracy of 0.9 %.

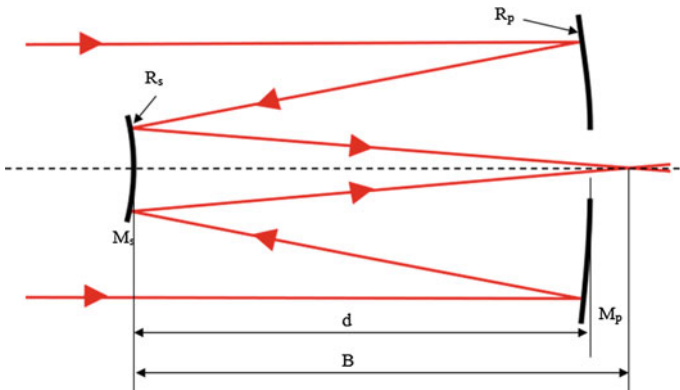


Fig. 1 Two-mirror telescope configuration

2 Basic Principle

The structure of the TMT is shown in Fig. 1. These optical system parameters, effective focal length (f), back focal length between from the secondary mirror vertex to system focus (B) and separation between the mirror vertices (d), the radius of curvature R_p of the primary mirror (M_p) and the radius of curvature R_s of the secondary mirror (M_s), respectively, in a TMT configuration are [1]

$$R_p = \frac{2df}{B-f}, \quad (1)$$

$$R_s = \frac{2dB}{B+d-f}. \quad (2)$$

Figure 1 shows the TMT configuration considered in our study and the parameters used. The conic constant of the primary mirror (k_p) can be adjusted according to the conic constant chosen for the secondary (k_s) for a two-mirror system corrected for third-order spherical aberration:

$$k_p = \left(\frac{B}{f}\right) \cdot \frac{(B+d-f)}{(B-f)^3} \cdot \left[(B+d-f)^2 k_s + (f+d-B)^2\right] - 1. \quad (3)$$

The classical Cassegrain or Gregorian configuration is obtained if both mirrors are independently corrected for third-order spherical aberration, which leads to the following conic constants:

$$k_p = R_p^3 \frac{(f-B)^3}{8d^3 f^3}, \quad (4)$$

$$k_s = R_s^3 \frac{(f-d-B)(f+d-B)^2}{8d^3 B^3}. \quad (5)$$

The Ritchey-Chrétien telescope (RC), which is also corrected for third-order coma, has a specific secondary conic constant given by

$$k_s = R_s^3 \frac{[2f(B-f^2) + (f-d-B)(f+d-B)(d-f-B)]}{8d^3 B^3}. \quad (6)$$

The difficulties of TMT's optical axis alignment are: (a) To solve "the error of design and manufacture" with "the precision of alignment"? (b) To keep the system stability when separating mirrors from the alignment mechanisms and fixed on the frames? (c) To finish the alignment with high performance in a shorter period? (d) To find the best focus and minimum spot diameter of the system. In this paper, the additive color mixing method uses to find the best focus and get minimum spot diameter of the TMT. The laser diodes are monochromatic light with the energy

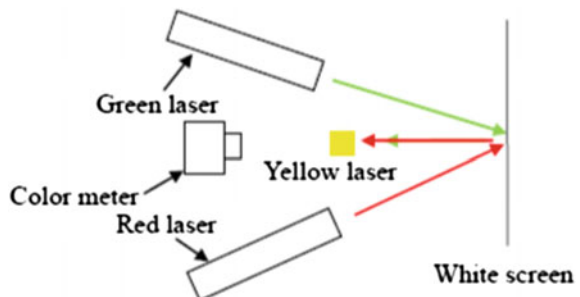
centered on a single wavelength. The red and the green laser are arranged so their outputs fall side by side on a white sheet of paper that the distinct red and green dots will be seen. Now move the lasers so that their beams overlap on the white screen; the combined dot will become yellow. A schematic diagram of the experimental configuration using red and green filters is shown below. If the beams are carefully adjusted, the region of overlap should be yellow. Simultaneous presence of reflected red and green appears yellow. Additive color mixing is the method of creating color by mixing various proportions of two or three distinct stimulus primary colors of light. These colors are commonly red, green, and blue; however they may be any wavelengths to stimulate distinct receptors on the retina of the eye and that the stimuli come from separate monochromatic sources. Grassmann's law is an empirical result about human color perception. The chromatic sensation can be described in terms of an effective stimulus consisting of linear combinations of different light colors. If a test color is the combination of two other colors, then in a matching experiment based on mixing primary light colors, an observer's matching value of each primary will be the sum of the matching values for each of the other test colors when viewed separately (Fig. 2).

If light beam 1 and 2 are the initial colors, and the observer chooses (R_1, G_1, B_1) as the light power of the primaries that match light beam 1 and (R_2, G_2, B_2) as the light power of the primaries that match light beam 2, then if the two lights beam were linear combined, the matching values will be the sums of the components. Precisely, they will be (R, G, B) , where:

$$\begin{aligned} R &= R_1 + R_2 \\ G &= G_1 + G_2 \\ B &= B_1 + B_2 \end{aligned} \quad (7)$$

A mixture of any two colors (light sources C_1 and C_2) can be matched by linearly adding together the mixtures of any three other colors that individually

Fig. 2 Additive mixing color



match the two source colors. This is Grassman’s second law of color mixture. It can be extended to any number of source colors.

$$C_3(C_3) = C_1(C_1) + C_2(C_2) = [R_1 + R_2](R) + [G_1 + G_2](G) + [B_1 + B_2](B) \quad (8)$$

where R is light power units of red, G is light power units of green and B is light power units of blue.

3 The Optical Alignment System

We develop a specific technique to align the optical axis of the two-mirror telescope. This alignment system can be applied to UV–IR regions by using the different wavelengths of alignment laser diode. The optical alignment system in Fig. 3 is composed of: (1) the two wavelength laser color alignment device shown in Fig. 4. #1, #2 and #3 are blue laser and #4 and #5 are red laser, (2) the additive color mixing shown in Fig. 5 are composed of two wavelengths laser color alignment device, two-mirror telescope and color meter (CL-200A). The Z is optical axis, (3) CCD Image camera and pinhole, and (4) optical table and mount (part (4) and (5) not shown). These are described as followings.

The two wavelengths laser color alignment device includes five parallel semiconductor lasers, adjustable mount and the rotation mechanism, and precise adjustable holder shown in Fig. 6. The semiconductor laser beams on the fixture are made parallel each other with precise adjustable holder and perpendicular to the surface fixture. An adjustable holder for mounting and orienting the semiconductor laser is set by a removable retaining laser within a mount of ring.

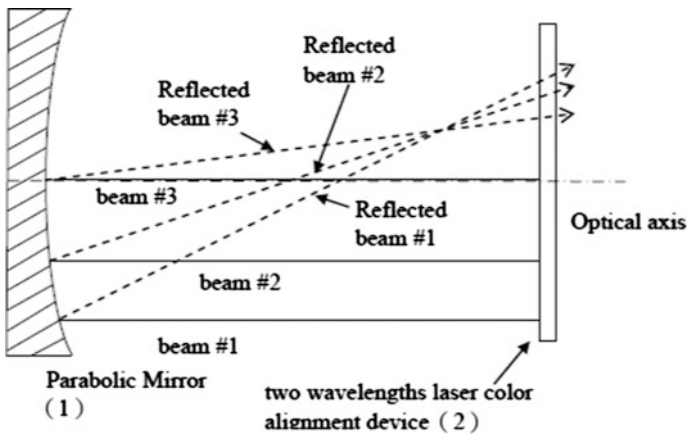


Fig. 3 The optical alignment system

Fig. 4 The schematic of two wavelength laser color alignment device

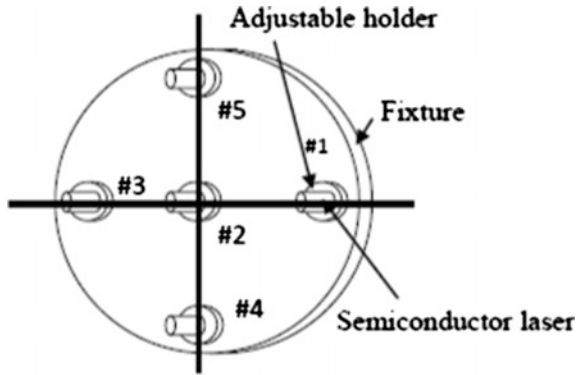


Fig. 5 TMT optical axis adjustment by additive color mixing method

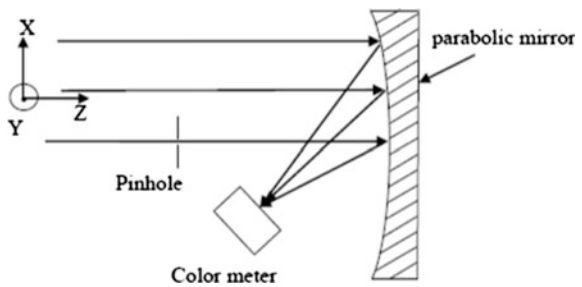
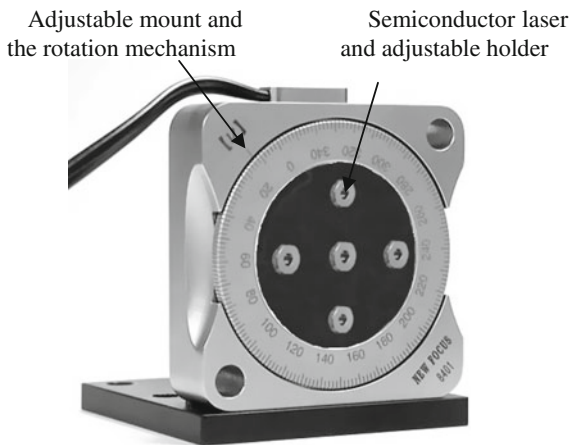


Fig. 6 Photograph of the portable alignment device



The rotation mechanism adjusts the initially horizontal and vertical laser beams to any rotation angle α as shown in Fig. 6. For example, the five laser beams lines, i.e., #1, #2, #3, #4 and #5 from right to left and down to up in Fig. 6. When it

clockwise rotates α degree, it shows #1, #2, #3, #4 and #5 from original direction to α direction in Fig. 6. In the rotation mechanism, it can rotate any α degree to adjust the Two-mirror telescope orientation we want.

The pinhole has a size of $30\ \mu\text{m}$ for alignment five laser beams and Two-mirror telescope. When the diffraction pattern of concentric circular rings was appeared, the laser beam passes through the pinhole accurately. The diffraction patterns formed by a pinhole consist of a central bright spot surrounded by a series of bright and dark rings shown in Fig. 7. The diffraction experimental set-up includes semiconductor laser, pinhole and observing screen. We can describe the pattern in terms of the angle θ , representing the radius angle of each ring. If the aperture diameter is d (mm) and the wavelength is λ (mm), the radius angle θ of the first dark ring is given by

$$\sin \theta = 1.22(\lambda/d) = y/D \tag{10}$$

The distance between the pinhole and screen is D . The actual diffraction pattern of alignment pinhole is shown in Fig. 8. Figure 8a show the horizontal laser beam incident on the pinhole along optical axis of pinhole and Fig. 8b show the oblique laser beam incident on the pinhole at slope angle α . With this diffraction technique, the pinhole is used to precise align laser beam parallel to the TMT of the alignment system.

Fig. 7 Pinhole diffraction pattern

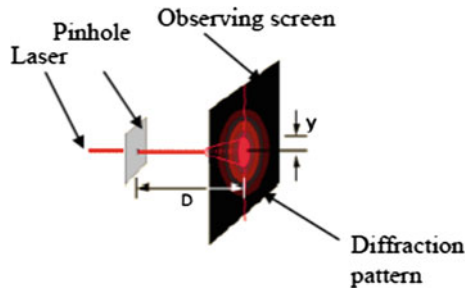
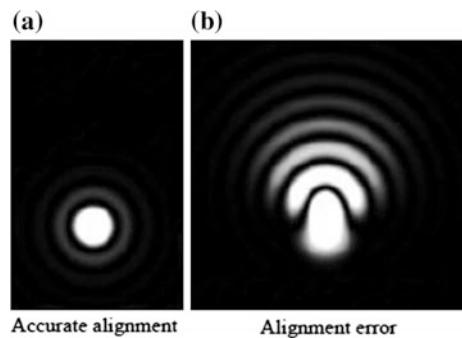


Fig. 8 The actual diffraction pattern for alignment pinhole.
a Accurate alignment.
b Alignment error



The Two-mirror telescope is brought into focus onto the CCD camera by looking for the minimum spot diameter. The focal spot is then centered using the XYZ translation stage attached to the CCD camera. By varying the separation of the Two-mirror telescope and detector, best focus is located when the displayed beam size is smallest.

4 Experiments and Results

The conditions and specifications used for the experiment are listed in Table 1. The whole optical alignment consists of two phases: initial alignment and the optical axis of TMT alignment. Second phase includes the alignment results and EFL measurement.

Table 1 The experiment conditions

<i>Two-mirror telescope</i>	
EFL	3750 mm
BFL	1200 mm
Diameter	800 mm
R_p, R_s	2352.94, 1097.14 mm
k_p, k_s	-1.0, -3.66
<i>Laser range finder</i>	
Range of measurement	0.02 up to 60 m
Measuring accuracy	± 1 mm
Time for a measurement	2.5–10 s
Light source	Red laser diode
<i>Red alignment laser</i>	
Wavelength (λ_p)	635 nm
Output power	3 mW
Beam divergence	< 2 mrad
Operating current	25 mA
Beam diameter	3.3 mm
<i>Green alignment laser</i>	
Wavelength (λ_p)	532 nm
Output power	5 mW
Beam divergence	< 1.2 mrad
Beam diameter	12 mm
<i>CCD camera</i>	
Resolution	752×582
Spectral range	350–1100 nm
<i>Beam splitter</i>	
Reflectance	50 % (at $\lambda = 635$ nm)
Transmittance	50 % (at $\lambda = 635$ nm)

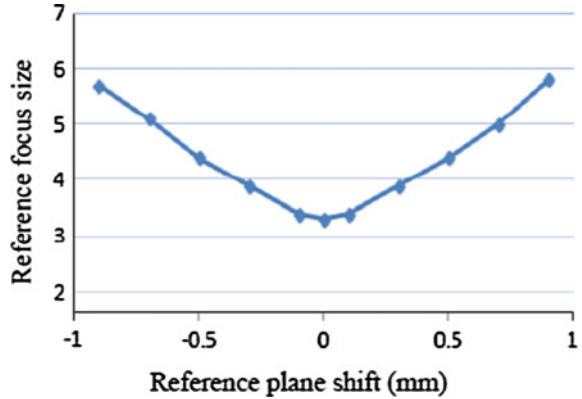
4.1 Initial Alignment

- (1) The fixture of alignment device is vertically (perpendicular) put on the optical table such that laser beam #2 is nearly parallel to the surface of the optical table.
- (2) The precise adjustable holder is fine adjusted to make the laser beam #2 parallel to the surface of the optical table. To check the parallelism, we use the pinhole which moves on the optical table from position 1 to position 2 (about 50 cm distance, for SFL = 457 mm) with a fixed height and has observed the pinhole diffraction pattern that the laser beam passes through the pinhole accurately shown in Fig. 8a.
- (3) The plane mirror is put on the optical table at 3 m distance from the alignment device. To adjust the plane mirror, the reflected beam #2 is made parallel to the optical table. To check the plane mirror perpendicular to the optical table, we use a piece of paper to chop reflected beam #2, and we can easily see how well the reflected beam #2 overlap at the output hole of laser beam #2.
- (4) Step (2) is repeated until the beam #4, #5, #3 and beam #1 is parallel to the surface of the optical table.
- (5) We use a piece of paper to chop reflected beam #4, #5, #3 and #1, and we can easily see how well the reflected beam #4, #5, #3 and #1 overlap at the output hole of laser beam #4, #5, #3 and #1, respectively.
- (6) Finally, the parallel of five incident parallel laser beams can be examined by the color meter shown in Fig. 5.

4.2 The Optical Axis of TMT Alignment

- (1) A primary mirror of TMT is adjusted at first so that its optical axis is in the incident plane made by three horizontal parallel incident laser beams.
- (2) The primary mirror is adjusted again so that the three horizontal parallel reflected beams come to a focus at the same point, i.e., the sagittal optical plane of primary mirror is found.
- (3) Three vertical parallel incident beams are parallel to optical axis by adjusting the primary mirror and the three vertical reflected beams also come to a focus at the same point obtained in (2), i.e., the tangential optical plane of primary mirror is found.
- (4) Capture the focus pattern at the reference plane (or near focal point) with the CCD camera. The reference plane is moved from left to right of the image plane to see if it fits small spot size.
- (5) The optical axis of secondary mirror is the same alignment procedure as the optical axis of primary mirror.
- (6) The EFL and BFL of a TMT are measured by laser ranger and trigonometry.

Fig. 9 Focus size when reference plane shifted



We measure best focus pattern curve that the reference focus size versus the position of the reference plane. Focus patterns sizes of reference plane are captured with CCD camera is shown in Fig. 9. The vertical coordinate is at different position of CCD camera. The relative focus size is normalized to reference focal point. The two-mirror telescope is brought into best focus onto the CCD camera by looking for the minimum spot diameter. The eleven focus patterns from left of the image plane to right of it are measured. The 30 % of the center of the maximum intensity is defined as the focus pattern of the Two-mirror telescope.

The alignment of the TMT is finished when getting relative minimum focus size. This validates the experimental system. This alignment experiment repeatedly does twenty times. The test results show that the average EFL is 3750 mm and the standard deviation is 0.755 mm or 0.9 %.

5 Conclusions

We have effectively presented the use of the five parallel laser beams alignment device, additive color mixing method and CCD camera to align the optical axis of two-mirror telescope. This system is designed for the optical engineers needing the additive color mixing and electro-optical practical techniques to achieve alignment of the telescope optical systems. The advantages of the TMT alignment techniques are: (1) it is simple to operate; (2) it simplifies the optical test system and lowers the cost [3, 4], (3) it is less expensive to maintain the equipments, and (4) a simple additive color mixing technique.

References

1. Smith WJ (2008) Modern optical engineering, 4th edn. McGraw Hill, New York, Chap. 5, p 74, Chap. 18, pp 508–514
2. Schmid T, Thompson KP, Rolland JP (2010) Misalignment-induced nodal aberration fields in two-mirror astronomical telescopes. *Appl Opt* 49:D131–D144
3. Orlenko EA, Cherezova TYu (2005) Off-axis parabolic mirrors: a method of adjusting them and of measuring and correcting their aberrations. *J Opt Technol* 72:306–312
4. Chrzanowski K (2007) Evaluation of infrared collimators for testing thermal imaging systems. *Opto-Elect Rev* 15:82–87

Design of Relay Lens Based on Zero Seidel Aberrations

Kuang-Lung Huang, Yu-Wei Chan, Jin-Jia Chen and Te-Shu Liu

Abstract Seidel aberration has been used successfully in finding starting points for a lens system, especially in designing a relay lens, which may have the symmetric condition with aperture stop, and the benefit of zero Seidel aberrations of com, distortion, and transverse chromatic aberration. In this paper, a relay lens has been designed based on the principle of zero Seidel aberrations. The MTF of the design is near diffraction limited, which has an excellent image performance of high resolution.

Keywords Seidel aberrations · Lens design · Relay lens

1 Introduction

Relay lens are used to relay an image from one place to another, as shown in Fig. 1. They are 1:1 relay lens, unit power copy lens, rifle sight, and eyepiece relay [1]. The use of Seidel aberration for initial condition of a lens design has been discussed in some references [2–4]. That is because the role of Seidel aberration is so important in finding a good solution of a potential lens. In this paper, a 1:1 relay lens has been designed based on zero Seidel aberrations, which shows the value of symmetrical lens design principle. A thin lens layout of controlling the Petzval sum and longitudinal chromatic aberration via glass selection has also been discussed.

K.-L. Huang

Department of Materials and Energy Engineering, Mingdao University,
369, Wen Hwa Rd., Peetow, Changhua 52345, Taiwan

Y.-W. Chan (✉)

Department of Hospitality Management, Chung Chou University of Science
and Technology, Changhua, Taiwan
e-mail: ywchan@dragon.ccut.edu.tw

J.-J. Chen · T.-S. Liu

Department of Electrical Engineering, National Changhua University
of Education, Changhua 50074, Taiwan

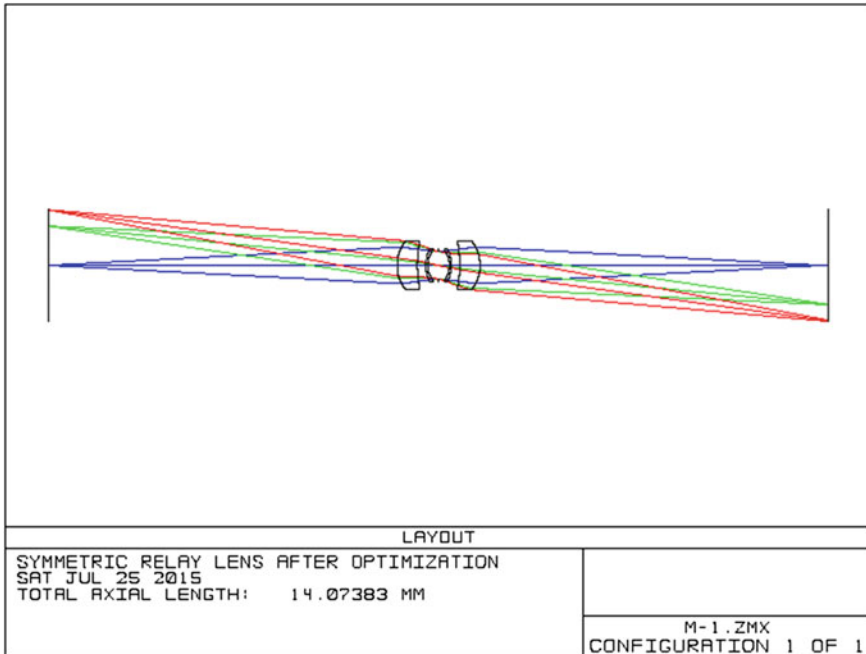


Fig. 1 1:1 relay lens

2 Thin Lens Layout for 1:1 Relay Lens

Seidel aberrations consist of 5 monochromatic aberrations, which are spherical aberration (S_I), coma (S_{II}), astigmatism (S_{III}), Petzval sum (or field curvature, S_{IV}), and distortion (S_V), and 2 chromatic aberrations, which are longitudinal chromatic aberration (C_L) and transverse chromatic aberration (C_T). The thin lens layout for 1:1 relay lens is processed by 3 steps, which are (1) controlling power, Petzval sum and longitudinal chromatic aberration via glass selection, (2) handling coma, distortion, and transverse chromatic aberration by mirror image part of the lens to the aperture stop (A.S.), and (3) minimizing spherical aberration and astigmatism by bending the lens shapes. The procedure is described as follows.

2.1 Controlling Power, Petzval Sum and Longitudinal Chromatic Aberration

Power, Petzval sum and longitudinal chromatic aberration controlling is important in thin lens layout. That is because the power (or magnification) has to be fixed, the

Petzval sum and the longitudinal chromatic aberration are related with glass materials (e.g. index n and Abbe value V). The thin lens equations of two lens system for total power (K), zero longitudinal chromatic aberration ($C_L = 0$) and Petzval sum ($S_{IV} = 0$) are listed as follows:

$$\sum_{n=1}^2 hK = h_1K_1 + h_2K_2 = h_1K \quad (1)$$

$$\sum_{n=1}^2 \frac{hK}{V} = \left(\frac{h_1^2}{V_1}\right)K_1 + \left(\frac{h_2^2}{V_2}\right)K_2 = 0 \quad (2)$$

$$\sum_{n=1}^2 \frac{K}{n} = \left(\frac{1}{n_1}\right)K_1 + \left(\frac{1}{n_2}\right)K_2 = 0 \quad (3)$$

where,

- h marginal ray height at each lens (e.g. lens 1, 2 ...)
- K Total power of lens
- K_1 Power of lens 1
- K_2 Power of lens 2
- n_1 Index of lens 1
- n_2 Index of lens 2

The lens power K_1 and K_2 can be obtained by solving Eqs. (1), (2) and (3) simultaneously, which satisfied the zero Petzval sum and zero longitudinal chromatic aberration conditions.

2.2 Handling Coma, Distortion, and Transverse Chromatic Aberration

The condition for zero coma, distortion, and transverse chromatic aberrations is obtained by mirror image the half part of lens structure to the aperture stop [5], as shown in Fig. 1. This result comes from the fact that the aberration of the components before the aperture cancelled exactly to that of the symmetric components after the aperture stop.

2.3 Minimizing Spherical Aberration and Astigmatism

At this stage, the remained Seidel aberrations are spherical aberration and astigmatism, which are shape dependent aberrations [6]. A bending technique has been

introduced to reduce these two aberrations to zero, which is changing the lens shape while keeping the focal length constant.

3 Design Examples

A 1:1 relay lens has been designed, which is composed of four symmetric lenses. The specification of the lens is listed as follows:

- Magnification: -1
- F-number: $f/10$
- Elements: 4
- Image height: 1 mm
- MTF at all fields: $\geq 0.5 @ 55$ cycles/mm.

3.1 Setting Half Part of the Relay Lens

Having selected two glasses, e.g. SF2 ($n_d = 1.647689$, $V_d = 33.8483$) and BK10 ($n_d = 1.497821$, $V_d = 66.9545$), and setting effected focal length of the right-hand side of the stop of the lens to 4.28 mm, the total power is controlled by Eq. (1); the longitudinal chromatic aberration is equal to zero by using Eq. (2) and the Petzval sum is also zero via Eq. (3). The lens data is listed in Table 1, and the Seidel aberration is listed in Table 2, which shows that both values of S_{IV} and C_L are all equal to zero.

3.2 Symmetric Relay Lens with Aperture Stop

In order to have zero coma, distortion, and transverse chromatic aberrations, the right-hand of the relay lens is mirror imaged with aperture stop to form the

Table 1 Lens data for right half-part of the relay lens

Surface	Curvature	Thickness (mm)	Glass material
Object	0.0000000	Infinity	
1.	-1.0579900	0.000	SF2
2.	0.0000000	0.693	
3.	0.0000000	0.000	BK10
4.	-1.2512910	6.314	
Image	0.0000000		

Table 2 Seidel aberration for right half-part of the relay lens

Surface	S_{IV}	S_I	S_{II}	S_{III}	S_V	C_L	C_T
1.	0.006	0.593	0.003	0.000	0.000	0.563	0.003
2.	0.000	0.426	0.086	0.017	0.003	0.365	0.073
3.	0.000	-0.552	-0.111	-0.022	-0.004	-0.230	-0.046
4.	-0.006	-7.212	-0.673	-0.063	-0.006	-0.698	-0.065
Total	0.000	-6.745	-0.696	-0.068	-0.007	0.000	-0.036

Unit microns

Table 3 Lens data for symmetric relay lens

Surface	Curvature	Thickness (mm)	Glass material
Object	0.0000000		
1.	1.2512910	6.314	BK10
2.	0.0000000	0.693	
3.	0.0000000	0.000	SF2
4.	1.0579900	0.500	
A.S.	0.0000000	0.500	
6.	-1.0579900	0.000	SF2
7.	0.0000000	0.693	
8.	0.0000000	0.000	BK10
9.	-1.2512910	6.314	
Image	0.0000000		

complete structure of the lens. The lens data is listed in Table 3, and the Seidel aberration is listed in Table 4, which has additional zero coma, distortion, and transverse chromatic aberrations.

Table 4 Seidel aberrations for symmetric relay lens

Surface	S_{IV}	S_I	S_{II}	S_{III}	S_V	C_L	C_T
1.	-0.006	-7.212	-0.379	-0.020	-0.001	-0.698	-0.037
2.	0.000	-0.552	-0.088	-0.014	-0.002	-0.230	-0.037
3.	0.000	0.426	0.068	0.011	0.002	0.365	0.058
4.	0.006	0.593	-0.022	0.001	0.000	0.563	-0.020
A.S.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6.	0.006	0.593	0.022	0.001	0.000	0.563	0.020
7.	0.000	0.426	-0.068	0.011	-0.002	0.365	-0.058
8.	0.000	-0.552	0.088	-0.014	0.002	-0.230	0.037
9.	-0.006	-7.212	0.379	-0.020	0.001	-0.698	0.037
Total	0.000	-6.745	0.000	-0.045	0.000	0.000	0.000

Unit microns

3.3 *Bending Lens Shape for Reducing Spherical and Astigmatism Aberrations*

The bending technique has been used on lens 2 and 3 for reducing spherical aberration and then lens 1 and 4 for zero astigmatism purpose. In order to keep zero longitudinal chromatic aberration, the lens thicknesses were added on lens 1 and 4 as parameters for bending. The seven Seidel aberrations are near zero, as shown in Table 5. The only remain Seidel aberration is longitudinal chromatic aberration (C_L), which has 0.007 micron, coming from the thickness change. The transverse ray aberration plot before optimization shows that the dominated aberration is only spherical aberration, as shown in Fig. 2. Figure 3 shows the MTF plot before optimization. This can be a good starting point for further optimization.

3.4 *Further Optimization*

The lens has so far is mainly following the thin lens theory, which normally has zero lens thickness and need further optimization to check the image performance via real ray tracing. The transverse ray fan plot after optimization, as shown in Fig. 4, shows that the aberration has been reduced to 10 %, compared to that of in Fig. 2. Figure 5 shows the MTF plot after optimization and those values are also improved compared to those of in Fig. 3.

Table 5 Zero Seidel aberration for symmetric relay lens

Surface	S_{IV}	S_I	S_{II}	S_{III}	S_V	C_L	C_T
1.	-0.579	-18.282	-1.779	-0.173	-0.073	-1.263	-0.123
2.	0.206	-0.003	0.022	-0.164	-0.317	0.027	-0.202
3.	-0.843	-0.722	0.272	-0.096	0.331	-1.210	0.426
4.	1.216	19.056	-2.872	0.433	-0.249	2.449	-0.369
A. Stop	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6.	1.216	19.057	2.872	0.433	0.249	2.449	0.369
7.	-0.843	-0.722	-0.272	-0.096	-0.331	-1.210	-0.426
8.	0.206	-0.003	-0.022	-0.164	0.317	0.027	0.202
9.	-0.579	-18.281	1.779	-0.173	0.073	-1.263	0.123
Total	0.000	0.000	0.000	0.000	0.000	0.007	0.000

Unit microns

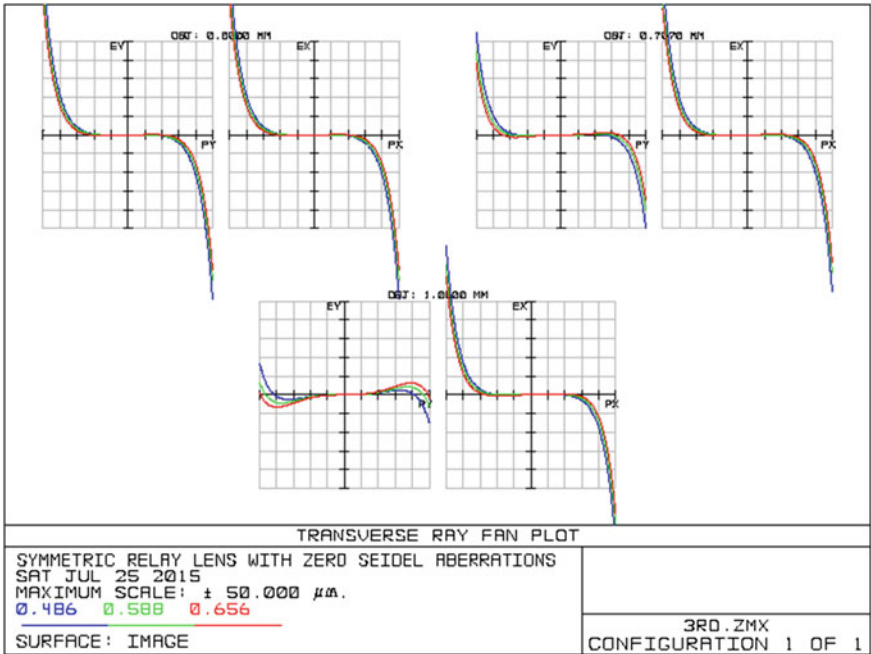


Fig. 2 Ray fan plot before optimization

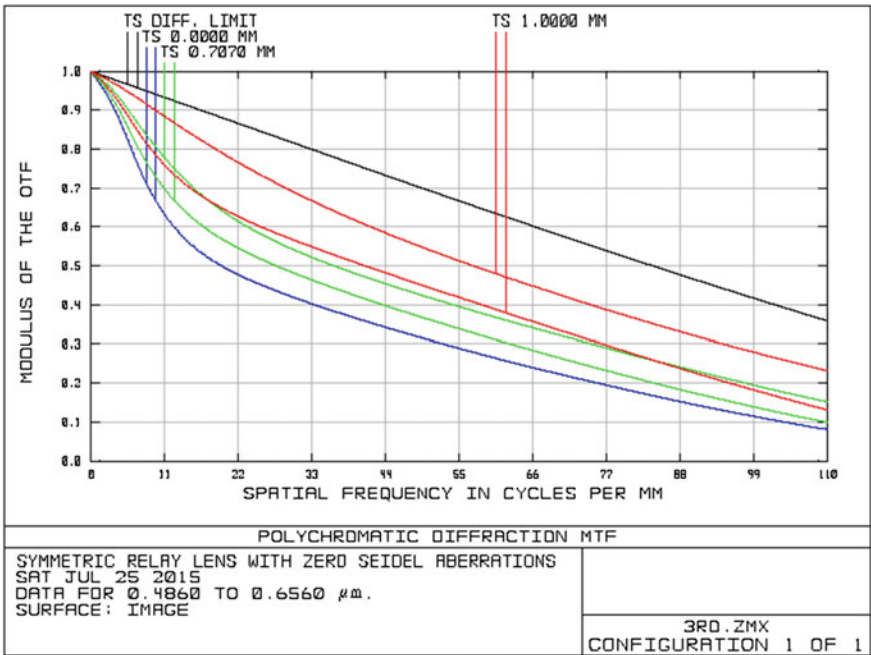


Fig. 3 MTF plot before optimization

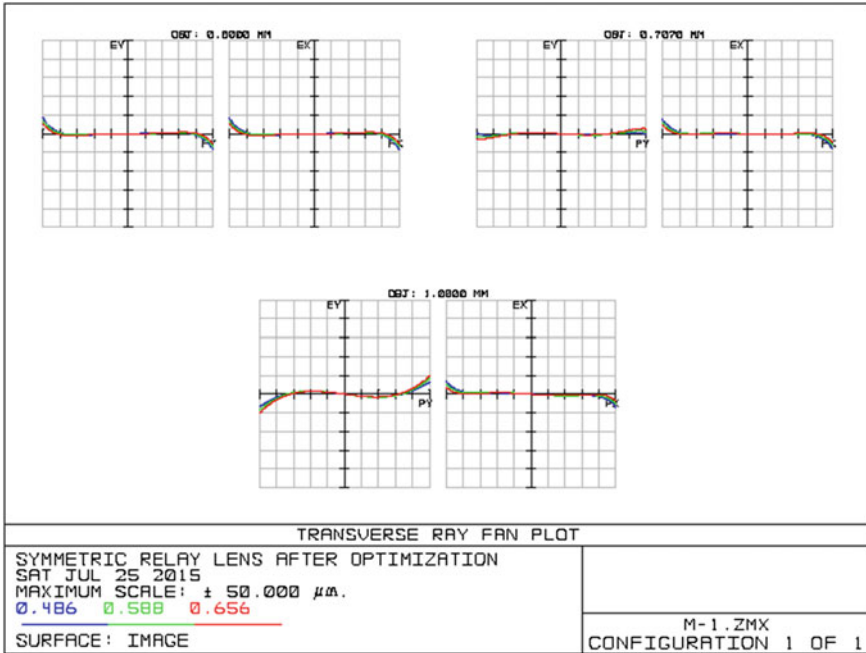


Fig. 4 Ray fan plot after optimization

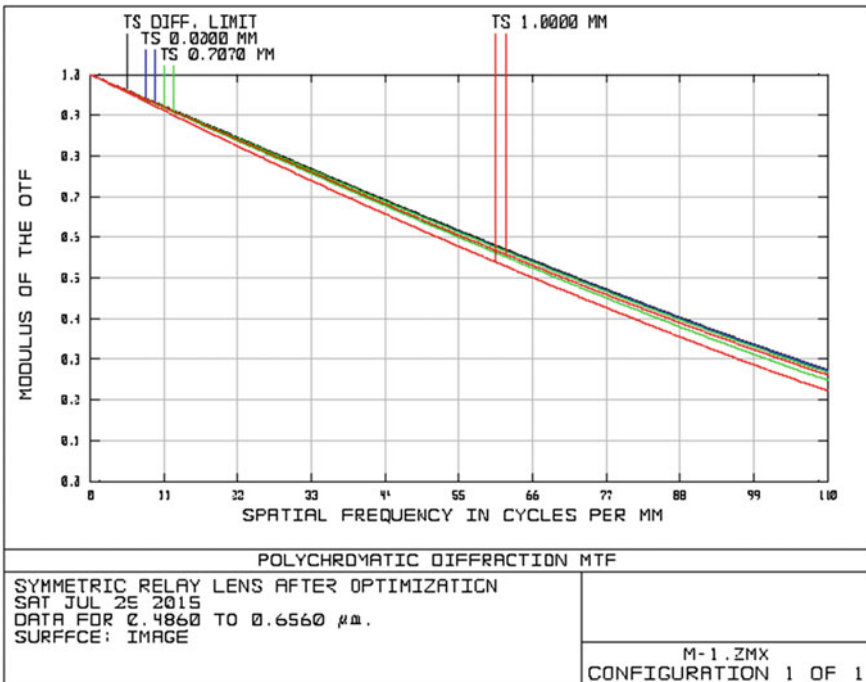


Fig. 5 MTF plot before optimization

4 Conclusion

This research provides the methodology of 1:1 symmetric relay lens design based on zero Seidel aberrations. The results obtained from this method can provide a good starting point for later optimization. A 1:1 symmetric relay lens design has been designed, which has an excellent image performance. The MTF of the lens is near diffraction limited.

References

1. Laikin M (2001) Lens design, Chap. 14. Marcel Dekker, New York
2. Smith WJ (1990) Modern optical engineering, Chap. 12. McGraw-Hill, New York
3. Welford WT (1986) Aberrations of optical system, Chap. 12. Adam Hilger, Bristol
4. Fischer RE et al (2008) Optical system design, Chap. 1. SPIE Press, Bellingham
5. Smith WJ (1990) Modern optical engineering, Chap. 3. McGraw-Hill, New York, pp 39–40
6. Kidger MJ (1992) Principles of lens design, critical reviews of optical science and technology. SPIE, CR41, pp 30–53

The Optical Spectra Analysis of 4 LED White-Light Sources Passing Through Different Fogs

Chien-Sheng Huang, Ching-Huang Lin, Guan-Syuan Hong and Hsuan-Fu Wang

Abstract As the prosperous development of LED fabrication technology, the solid lighting has been a very popular issue in illumination. Not only does the LEDs light source provide a choice of energy-saving and longer lifetime, but also it could change its color-temperature depending on weather or traffic conditions. However, due to the droplets in fog, the propagation of light will be affected differently according to the wavelength. In this paper, the spectra of 4 LED white-light sources, which were RGBY-mixed or phosphors-activated LEDs with high or low color temperature, were explored before and after passing through 5 different fog conditions. The results showed that the color temperature alteration of each white-light source was so different from clear to heaviest fogs. The spectra analysis suggested that the short wavelength (blue) suffered more scattering attenuation than the longer one (red) while passing through fog.

Keywords Fog · Color temperature · Solid lighting

1 Introduction

Nowadays, the progress of white LEDs has made the lighting technology into a whole new era of energy saving and longer lifetime. However, the heat management is important for LED due to its negative impact on the lifetime and optical performance, e.g., the light decay [1]. For illumination safety, an acceptable chromaticity shift should be established by scientific validation. As the substrate temperature of LED increases, the chromaticity of white LEDs will shift due to the phosphor

C.-S. Huang · C.-H. Lin (✉) · G.-S. Hong
Department of Electronic Engineering, National Yunlin University of Science and Technology, Douliu, Taiwan
e-mail: wilburlin@yuntech.edu.tw

H.-F. Wang
Department of Electrical Engineering and Energy Technology,
Chung Chou University of Science and Technology, Changhua, Taiwan

stability [2]. However, they are all conducted under clear air environments. On a fog environment, the refractive index and scattering coefficient are both dependent on the light wavelength [3]. Therefore, the chromaticity shift may also occur when white light passes through fog. In this research, 2 popular types of white light LEDs were chosen, which were phosphors-activated and RGBY-mixed, respectively. Furthermore, each white LED type would be categorized into 2 correlated color temperatures (CCT), low and high, and there were 4 LED white-light sources used in the experiments. Each one was tested in a pipe under 5 fog conditions, respectively, and the chromaticity and spectra were measured at 5 m away.

2 Experiment

A flexible tube and smoke-producer were used, and data were collected at 5 m away. 5 fog levels were determined by a green laser. Its light intensity recorded at 5 m away was used as 5 fog levels index.

The well known attenuation formula [4] is used for the illuminance of the laser and the white LEDs,

$$I = I_0 e^{-(\alpha x)}, \quad (1)$$

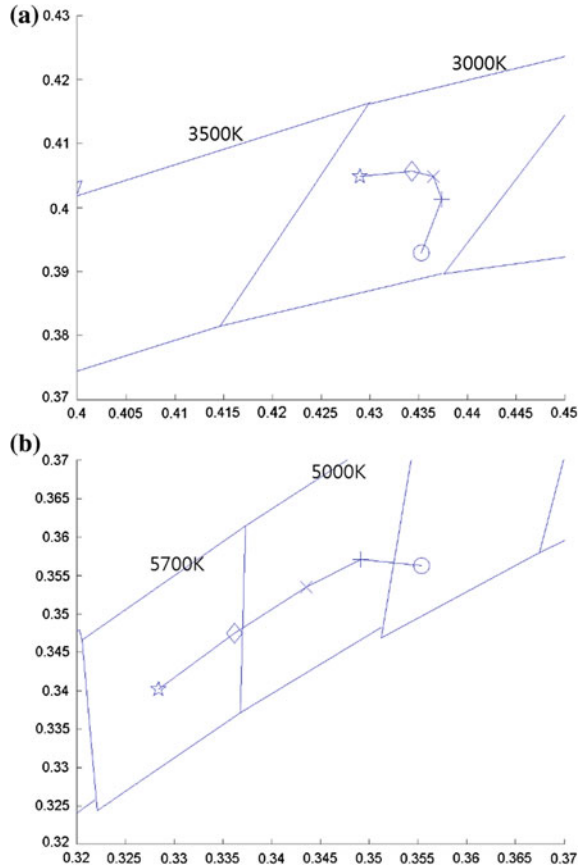
In this formula, I_0 is the illuminance of sources at the entrance of the pipe, and x is the distance between this plane and experimental planes. The constant of attenuation α could be deduced by data collected at 5 m. Hereafter, 5 levels of fog density were characterized to clear, light, mediate, heavy, and heavier.

As the pipe was stable in one fog level, one LED white-light source would shine at the entrance of the pipe, and its chromaticity and spectra were recorded at 5 m away. The low CCT of phosphors-activated white LED was 3000 K, and that of RGBY-mixed white LED was 3500 K, respectively. The high CCT of phosphors-activated white LED was 5700 K, and that of RGBY-mixed white LED was 6500 K, respectively.

3 Results and Discussion

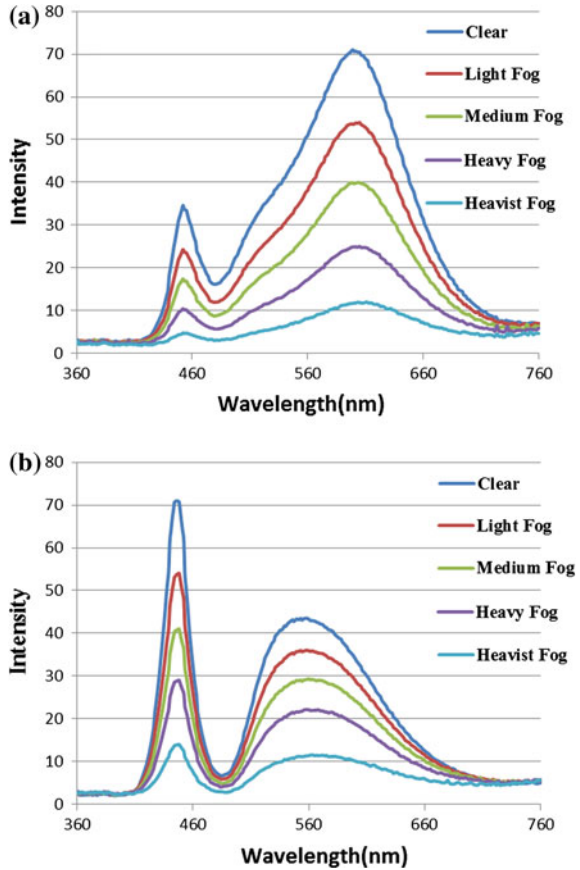
This research aims to make a framework for learning system on cloud computing environment. Figure 1 showed the chromaticity shifts of phosphors-activated white LEDs under 5 fog levels at 5 m away, and Fig. 2 showed the corresponding spectra. For low CCT phosphors-activated white LEDs, its CCT would also remain in the area of 3000 K under all the 5 fog levels. However, for high CCT phosphors-activated white LEDs, its CCT would shift from 5700 to 4500 K as the fog levels increased from clear to heavier. According to the spectra shown in Fig. 2, the intensity of each wavelength decreased as the fog levels increased. This would be explained by

Fig. 1 The chromaticity shifts of phosphors-activated LEDs white-light sources with **a** 3000 K, and **b** 5700 K color temperature passing through 5 m fogs of different levels: clear (*star*), light (*diamond*), mediate (*cross*), heavy (*plus*), and heavier (*circle*)



attenuation formula (1). However, the attenuation was dependent on the wavelength in such a way that the short wavelength would suffer more attenuation than the long wavelength did. The phosphors-activated white LEDs would have the same blue light to activate the broad yellow band phosphors. And it was obviously shown in Fig. 2 that the attenuation of blue wavelength part for low or high CCT white LEDs was similar. Nevertheless, for low CCT one, the broad yellow band peak was closer to long wavelength side than that of high CCT one. Thus the attenuation on the broad yellow band would be different for both of them. There were two main noticed phenomena. First, the peak of the broad yellow band shifted a little to the long wavelength as the fog levels increased. Second, the shape of the broad yellow band would change slightly, too. CCT of white LEDs was determined by the relative intensity ratio of wavelength, and a red shift of CCT would occur when the long wavelength part relatively dominated. The high CCT white LEDs suffered more attenuation in the broad yellow band while passing through fog, and their peak shifting were also more obvious than the low CCT ones did. As a result, while fog levels increased, the high CCT white LEDs changed its CCT to 5000 K under mediate and heavy fog levels, and even to 4500 K

Fig. 2 The spectra of phosphors-activated LEDs white-light sources with **a** 3000 K, and **b** 5700 K color temperature passing through 5 m fogs of different levels



under heavier fog levels, respectively. However, for low CCT white LEDs, the red shifting was towards the violet margin and its CCT stills remained in the 3000 K area.

Figure 3 showed the chromaticity shifts of RGBY-mixed white LEDs under 5 fog levels at 5 m away, and Fig. 4 showed the corresponding spectra. For low CCT RGBY-mixed white LEDs, its CCT would also remain in the area of 3500 K under all the 5 fog levels. However, for high CCT RGBY-mixed white LEDs, its CCT would shift from 6500 to 4500 K as the fog levels increased from clear to heavier. According to the spectra shown in Fig. 4, the intensity of each wavelength decreased as the fog levels increased. Following the discussion in last paragraph, and due to the narrow band characteristics of red, green, blue, and yellow component LEDs, the peaks of each component did not shift. However, the relative intensity ratio would still change, and would result in the change of CCT while passing through fog. For the low CCT one, it shifted zigzag towards violet margin, too. For high CCT one, it changed its CCT to 5700 K under light and mediate fog levels, and 5000 K under heavy fog level, and even to 4500 K under heavier fog levels, respectively.

Fig. 3 The chromaticity shifts of RGBY-mixed LEDs white-light sources with **a** 3500 K, and **b** 6500 K color temperature passing through 5 m fogs of different levels: clear (*star*), light (*diamond*), mediate (*cross*), heavy (*plus*), and heavier (*circle*)

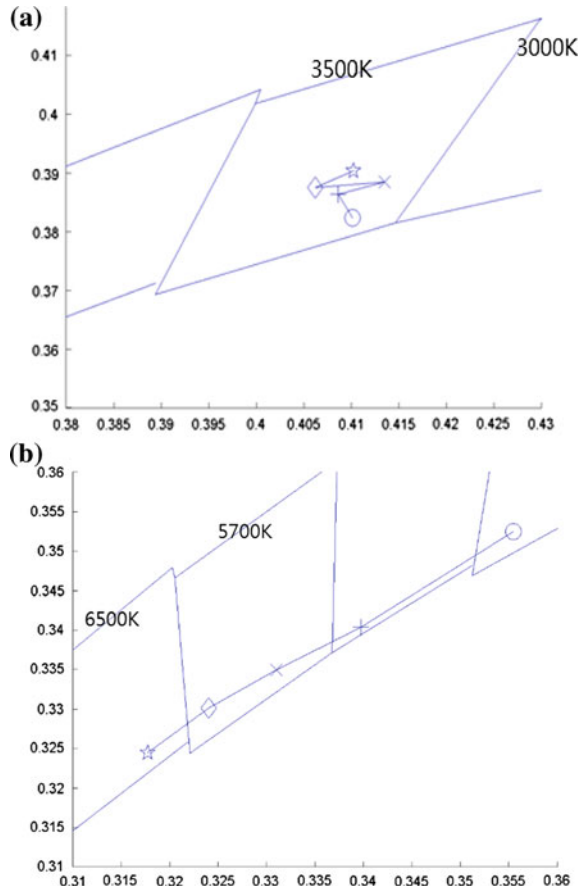


Table 1 showed the measured attenuation constants of 4 white-light LEDs, i.e., RGBY-mixed type white LEDs with low and high CCT, and phosphors-activated ones with low and high CCT, under 5 fog levels, respectively. The light intensity of the source at the entrance and 5 m away were both measured, and the attenuation constant was calculated by using formula (1). The phosphors-activated white LEDs showed a little higher attenuation than RGBY-mixed one under all fog levels, regardless of low or high CCT. It should be noticed that the calculated attenuation constant was attributed to all the wavelengths in the visible range, and was a result of effective weighting on each wavelength component. Since the phosphors-activated white LEDs consisted of broader wavelengths than RGBY-mixed one, it was reasonable that the former exhibited a little higher calculated attenuation constant than the latter did.

Compared with the low CCT white sources, the high CCT white sources, no matter made of RGBY-mixed or phosphors-activated LEDs, exhibited higher attenuation constants while passing through clear, light, and mediate fog levels.

Fig. 4 The spectra of RGBY-mixed LEDs white-light sources with **a** 3500 K, and **b** 6500 K color temperature passing through 5 m fogs of different levels

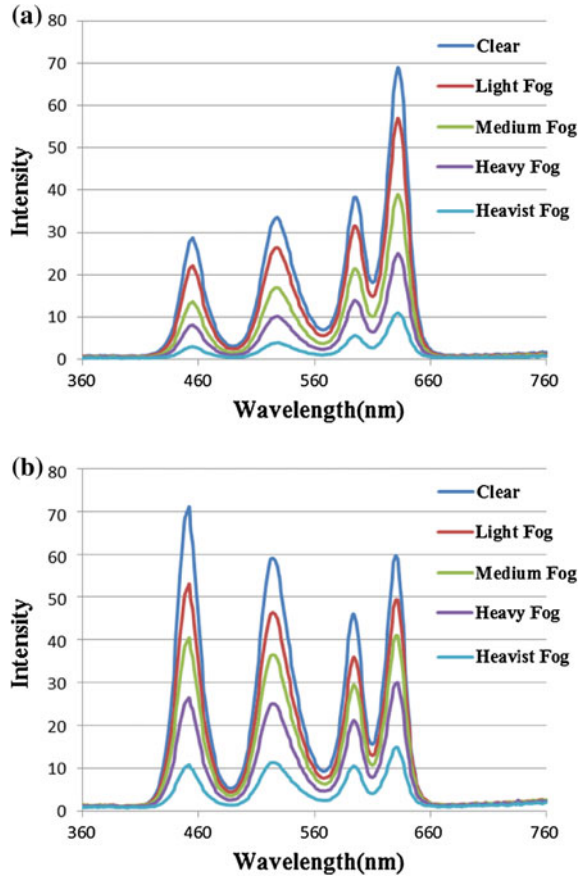


Table 1 The attenuation constants of 4 white-light LEDs

Source	RGBY-mixed		Phosphors-activated	
	3500 K	6500 K	3000 K	5700 K
Fog				
Clear	0.304	0.310	0.308	0.321
Light	0.343	0.369	0.363	0.375
Mediate	0.418	0.420	0.423	0.431
Heavy	0.507	0.482	0.517	0.500
Heavier	0.663	0.621	0.672	0.645

However, when the fog levels increased to heavy and heavier, the situation were reversed, and the low CCT sources showed higher attenuation constant. From Figs. 2 and 4, it was obvious that the depression in the wavelength range of 500–530 nm could be the major reason.

4 Results and Discussion

The 4 LED white-light sources showed different characteristics while they shined through 5 fog levels. The low CCT sources, no matter made of phosphors-activated or RGBY-mixed LEDs, as they shined through fog, would exhibit stable CCT than the high CCT ones did, which otherwise exhibited a red shift of CCT towards low. The calculated attenuation of light intensity was related to the type of the white-light sources. The phosphors-activated LEDs exhibited a little higher attenuation than RGBY-mixed one.

Acknowledgements We would like to express our thanks to the graduate Shao-Chiang Gan for his help in the experimental setup. And also we thank the Ministry of Science and Technology of Taiwan to support this paper (NSC-103-2220-E-224-001).

References

1. Narendran N, Gu Y (2005) Life of LED-based white light sources. *IEEE/OSA J Disp Technol* 1:167–171
2. Dal Lago M, Meneghini M, Trivellin N, Mura G, Vanzi M, Meneghesso G, Zanoni E (2012) Phosphors for LED-based light sources: thermal properties and reliabilities issues. *Microelectron Reliab* 52:2164–2167
3. Strutt J (1871) On the scattering of light by small particles. *Phil Mag* 41:447–454
4. Beer A (1852) Determination of the absorption of red light in colored liquids. *Annalen der Physik und Chemie* 86:78–88

High Resolution Camera Lens Design for Tablet PC

Ching-Huang Lin, Hsien-Chang Lin, Ta-Hsiung Cho,
Hsuan-Fu Wang and Cheng-Chieh Tseng

Abstract Tablet pc with camera has become the standard function in recent years. Especially high definition camera is the primary specification of high level tablet pc. In this paper, we design a 13 mega-pixels lens system by Zemax. The system includes five plastic aspheric lenses, an IR cut off filter and a sensor cover glass. The F number and FOV of the camera are 2.4 and 64° respectively. The sensor have 13 mega-pixels which is made by OmniVision, the maximum resolution is 4224×3120 and the pixel size is $1.3 \mu\text{m}$. The design result shows that RMS spot radius at different fields are small, so it is closed to the diffraction limit. The MTF value is more than 0.45 in most fields of view at 1/2 Nyquist sampling frequency.

Keywords Optical design · Tablet pc · High resolution

1 Introduction

The tablet pc is small in size and capable of carrying a functional mobile computer. The main input is the touchscreen display, which allow the user through body parts or stylus to operate it. The traditional computer mouse or keyboard commonly use

C.-H. Lin (✉) · C.-C. Tseng

Department of Electronic Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan
e-mail: wilburlin@yuntech.edu.tw

H.-C. Lin

Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology, Yunlin, Taiwan

T.-H. Cho

Department of Optometry, Shu Zen Junior College of Medicine and Management, Kaohsiung, Taiwan

H.-F. Wang

Department of Electrical Engineering and Energy Technology, Chung Chou University of Science and Technology, Changhua, Taiwan

to carry out operations, but the tablet pc device via the built-in virtual keyboard, handwriting recognition and speech recognition to complete. The functions of tablet pc include the photographs, documentation, video player, applications (app) and so on. In particular vertical markets, such as financial services systems, medical information systems and logistics management systems, which can via the environment needs to make the special functions of tablet pc.

In 1968, computer scientist Kay [1] proposed personal, portable information manipulator at Palo Alto Research Center, and described in his paper as the Dynabook in 1972 [2], the idea of a portable personal computer followed. Since 1980, several new input techniques of personal computer were developed, including the touchpad (Gavilan SC, 1983), handwriting recognition (Linus Write-Top, 1987). In 1989, the Palm company published the first commercial tablet pc, GridPad, nowadays some people think this is the world's first tablet pc.

The end of January 2010, Apple introduced a mobile computer called the iPad [3], which the position of the product is between mobile phones and notebook computers, product size is 7.75×5.82 inches, thickness of 13.4 mm. The user interface is the touch screen, but does not carry the camera module. At the end of 2010, Microsoft Windows XP Tablet PC Edition product tried to call Tablet PC, making the Tablet PC name has become more well-known. In 2011, Steve Jobs unveiled a new product iPad2. The iPad2 added to the camera module, including VGA of front camera and 0.7 mega-pixels of back camera. Thickness of product reduced to 9 mm, making the tablet pc forward the thin and high resolution.

With the camera modules continuous developed, tablet pc in resolution from VGA to nowadays common 2 and 3 mega-pixels, or even 5 and 8 mega-pixels, it has become the primary specification of high level tablet pc. But with the increase in pixels, the volume has not increased too much, attributed to the rapid development of complementary metal oxide semiconductor (CMOS) sensors, the pixel size from $5 \mu\text{m}$ to reduce the common 1.75 and $1.4 \mu\text{m}$, or even $1 \mu\text{m}$, and therefore under the same conditions as pixels, they can get smaller volume [4].

Compare Apple in October 16, 2014 announced the iPad Air 2, including 1.2 mega-pixels of front camera and 8 mega-pixels of back camera that the F number are 2.2 and 2.4, respectively. The primary back camera used five-piece optical lens and pixel size $1.12 \mu\text{m}$, adopting BSI technology of CMOS sensor. In this study, we designed a high-pixel lens for tablet pc with Zemax, using 1/2.6 inch CMOS sensor with maximum resolution 4224×3120 , the pixel size $1.3 \mu\text{m}$ and F number 2.4, which made the total track length of the lens less than 7 mm.

2 Initial Structure Selection

An excellent design usually begins with the choice of the initial configuration. There are general two ways to select initial structure for optical designers. It is difficult to create an initial configuration with Gaussian optics principle by designers, which requires many years experience and wealthy aberration theoretical

Table 1 Design specifications

Item	Specifications
Wavelength	486–656 nm
Half of image height	3.46 mm
F/#	2.4
FOV	64°
Effective focal length	5.54 mm
Relative lamination	>50 %
Optical distortion	<2 %
Total track length	<10 mm
MTF at 385 lp/mm	>10 %
Lens structure	5P2G

knowledge. The best way is to select one suitable initial configuration directly from the patent and then initialize it. We choose the lens of patent libraries as initial configuration, and do further optimize to meet the design’s requirement.

The initial configuration use 5P1G configuration, effective focal length is 5.8 mm, F number is 2.45, half full of view is 33.5°. We set the default wavelength by F, d, C (486.1, 587.6, 656.3 nm) as visible light. The image height on the sensor is equal to one half of diagonal length of the CMOS, and it is 3.46 mm according to data of the effective area of CMOS. The target F/# is set to be 2.4 and the half view angle to be 32°. The effective focal length can be calculated 5.53 mm, according to Eq. (1).

$$\tan \omega = \frac{\text{image height}}{\text{focal length}} \tag{1}$$

The above calculations are the theoretical value without considering the distortion. If considering the lens distortion less than 2 %, the focal length should be enlarged to 5.54 mm. The total track length must be less than 10 mm, the MTF of all fields should be greater than 10 %. According to the above conditions, we initialize the system to meet the requirements. The design specifications are as shown in Table 1.

3 Sensor Selection

There are two types of the image sensor: the charge coupled device (CCD) and the complementary metal oxide semiconductor. Generally CCD superiors to CMOS in Sensitivity and noise, so far most of the high end digital cameras and high end consumer products both use CCD as the sensing element. However, CMOS-based cameras, which consume significantly less power, while offering higher levels of integration and a lower overall system Bill of Materials, compare better with CCD-based systems [5].

The CMOS production process enables the integration of all camera functions on a single chip, which significantly reduce component count, cost and board space. The CMOS image sensors allows for quick and easy application designs, which enable less time to market for systems manufacturers. Therefore, all manufacturers have been actively involved in CMOS development such as a backside illuminated (BSI) sensor [6] so that the pixel size can be reduced, sensitivity increased, and reduction of noise. It does not produce too dark situation. BSI represents a revolution in the mass production of CMOS image sensors (CIS), which adopt a fundamentally different approach to traditional pixel architectures.

BSI offers CIS architectures for generations to come by enabling continued improvements in sensitivity, color reproduction and image quality while the design shrink down to 0.9 μm pixels by using backside illumination technology. Currently CMOS sensors have been the mainstream camera lens for the tablet pc, the major international manufacturers of CMOS have Aptina, OmniVision, Sony, Samsung, STmicro, Toshiba, etc. [7].

We chose the image sensor OV13860 which is a 13 mega-pixels 1/2.6 inch CMOS chip provided by Omnivision Co., adopting the backside illumination technology. The sensor's image area is 5.5541 mm \times 4.1134 mm, and its diagonal can be calculated as 6.911 mm. In order to avoid the dark corner of the sensor's edge, the half image height must be slightly greater than the diagonal of the sensor's image area, so we set the half image height as 3.46 mm.

According to Nyquist sampling theorem, we can calculate the maximum range of spatial frequency. The single pixel size of the sensor is 1.3 μm \times 1.3 μm . Pixel size determines the spatial frequency of the lens, and they satisfy the spatial frequency = $1/(2 \times \text{pixel size}) = 385 \text{ lp/mm}$.

The MTF (Modulation Transfer Function) value is often used to determine the image quality of the optical system is good or bad. The higher MTF curve indicates that the optical system can deliver more information to make better the image quality.

In order to realize excellent image quality, the MTF of 0.7 field should be achieved to 45 % at 192.5 lp/mm and to 10 % at 385 lp/mm of spatial frequency. The basic specifications are shown in Table 2.

Table 2 Basic specifications of CMOS

Sensor	OV13860 (1/2.6")
Resolution	4224 \times 3120
Pixel size	1.3 μm \times 1.3 μm
Image area	5.5541 mm \times 4.1134 mm
Diagonal length of unit sensor	6.911 mm
Nyquist frequency	385 lp/mm
Lens chief ray angle	<33.4°

4 Material Selection

There are a variety of optical materials, such as glass, plastics, optical crystal, semiconductor, ceramic, film, etc., according to their needs for different applications. Most manufacturers based on cost considerations, the lenses often used optical plastic materials. The optical plastic materials have a low cost, light weight, high luminous transmittance, easy processing and so on. The design of the lens use aspherical plastic surface to simplify the system structure, increase the relative aperture lens and the imaging quality.

Polymethylmethacrylate (PMMA) and Polycarbonate (PC) are two commonly used optical plastic materials. PC has a high refractive index and cheap, but compared to other injection-molded materials has large birefringence. PMMA has cheap, but high water absorption less for the camera lens. For the above reasons, some more excellent optical plastic appear. Such as Apel, Topas of Cyclic Olefin Copolymer (COC) and Zeonex, Arton of Cyclo Olefin Polymer (COP) and Osaka Gas Chemicals of Polyester [8].

COC has better heat resistance and moisture resistance capability. The highly transparent resin also possesses excellent properties in terms of low birefringence and low water absorption that are important in optical components. The birefringence less than 20 and water absorption less than 0.01, and it can reduce injection molding effect on image quality of optical lenses. Its has also high transparency, the light transmittance reach to 91 %.

COP has low moisture and high temperature environmental stability. Its has the higher molding temperature capability to improve birefringence and the water absorption significantly less than PC and PMMA. It has a refractive index of 1.53 and abbe number of 56.

OKP features have high refractive index, low birefringence, low abbe number and high fluidity. OKP is a special polyester for optical use arising from coal chemistry, it has a high refractive index of 1.6 or more, extremely low birefringence, and high fluidity. Therefore, it is easy to obtain high performance injection-molded objects and films. They are be used as a standard plastic material for camera.

In this paper, the first, third and fifth pieces of lenses adopt E48R material, and second and forth pieces of lenses adopts OKP4HT, and sixth and seventh pieces of

Table 3 Optical plastic materials (1)

Property	PMMA	TEJIN AD5503	OSAKA OKP4
Polymer	Acrylate	PC	Polyester
Refractive index (n_d)	1.49	1.58	1.61
Abbe number (v_d)	58	30	27
Transmittance (%)	92.5	89	90
Birefringence (nm)	13	>80	<20
Heat distortion temperature (I)	101	124	135
Water absorption (%)	0.3	0.2	0.14

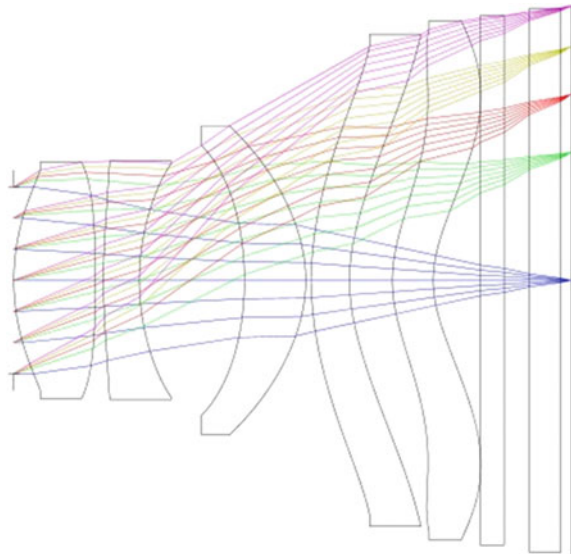
Table 4 Optical plastic materials (2)

Property	APEL 5014DP	TOPAS 5013L	ARTON FX4727	ZEONEX E48R
Polymer	COC	COC	COP	COP
Refractive index (n_d)	1.54	1.53	1.52	1.53
Abbe number (v_d)	56	56	52	56
Transmittance (%)	90	91	93	92
Birefringence (nm)	2	<20	42.6	32.7
Heat distortion temperature (I)	125	123	110	122
Water absorption (%)	<0.01	<0.01	0.05	<0.01

lenses adopt BK7, as IR cut off filter, filtering 700–1100 nm near-infrared wavelength and sensor cover glass, respectively. The common optical plastic materials are as shown in Tables 3 and 4.

5 Simulation Results and Discussions

The optimized lens configuration is shown in Fig. 1, the total track length of the imaging lens system is 5.6 mm, with an effective focal length of 7.001 mm, and of a back focal length 0.837 mm. The lens has a FOV of 64° , the image height is 6.92 mm

Fig. 1 Layout of the system

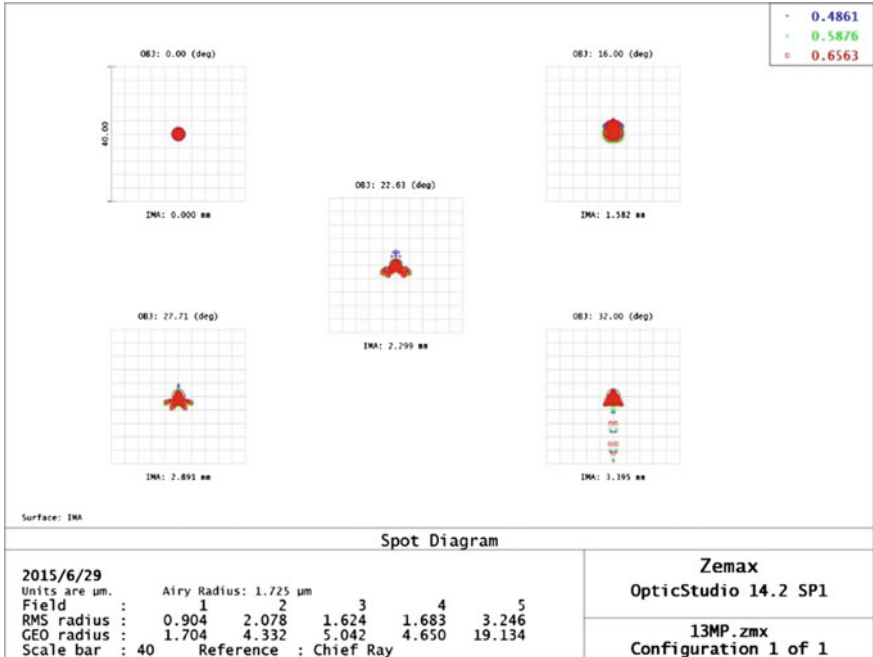


Fig. 2 Spot diagram

which is a little larger than the CMOS diagonal 6.911 mm, which ensure the CMOS sensor of image without vignetting. The chief ray angle is less than 33.4°.

5.1 Spot Diagram

Figure 2 shows the spot diagram. The airy disk of the spot diagram is 1.725 mm, and the fifth field has the largest RMS radius which is about 3.246 mm, the RMS radius is close to the airy disk radius, the image is close to the diffraction limit.

5.2 MTF

The MTF curve for different FOVs is shown in Fig. 3. MTF is a commonly standard to evaluate the imaging quality on optical system. In this design, the MTF value of 0.7 field at 192.5 lp/mm is more than 45 % and more than 10 % at 385 lp/mm.

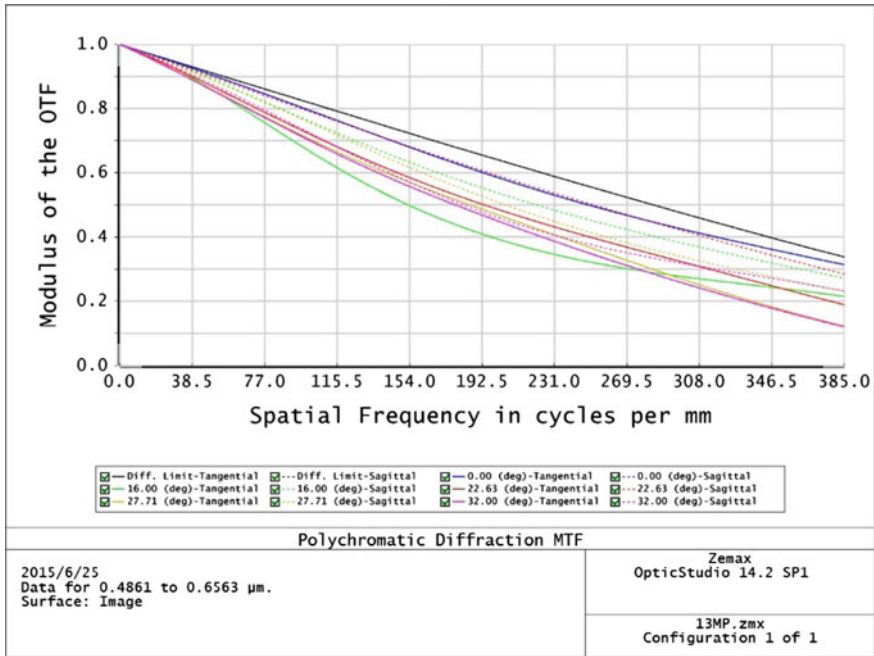


Fig. 3 MTF

5.3 Field Curvature and Distortion

The curvature and distortion of the lens is shown in Fig. 4. The field curvature is less than 0.05 mm. The larger the angle of view is greater distortion. The optical distortion is less than 2 %, the human eye will not perceive it.

5.4 Longitudinal Aberration

Figure 5 shows the longitudinal aberration. The low spherical aberration is controlled between -0.05 and 0.05 mm.

5.5 Later Color

Figure 6 shows the Later color. The lateral color of the maximum field is within the $1.3 \mu\text{m}$, which is less than a pixel size to avoid color shift on the sensor.

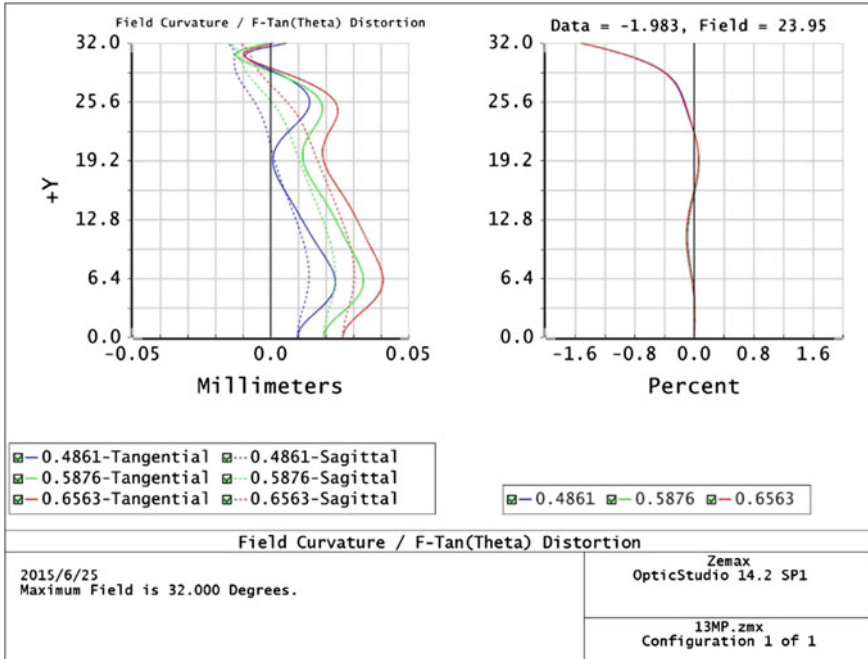


Fig. 4 Field curvature and distortion

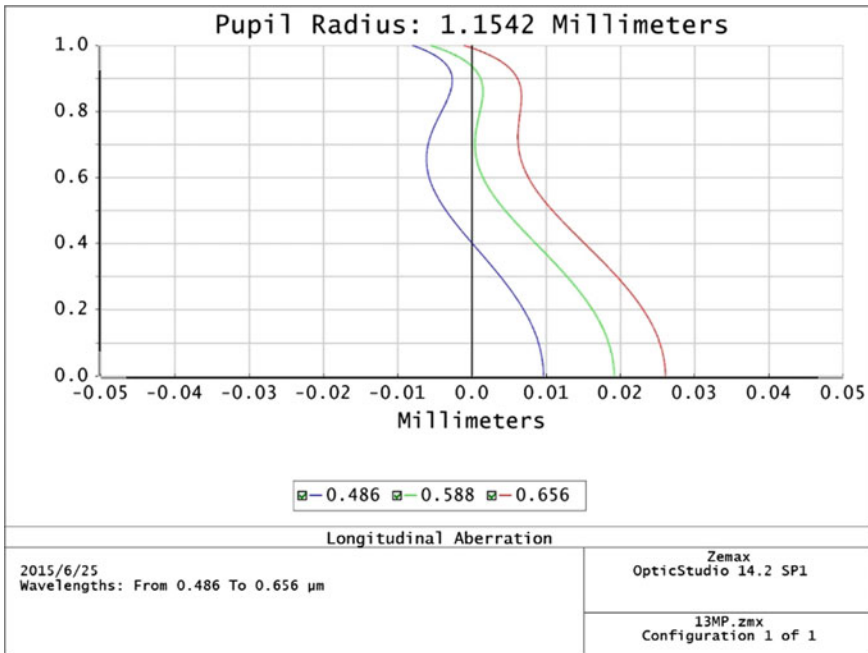


Fig. 5 Longitudinal aberration

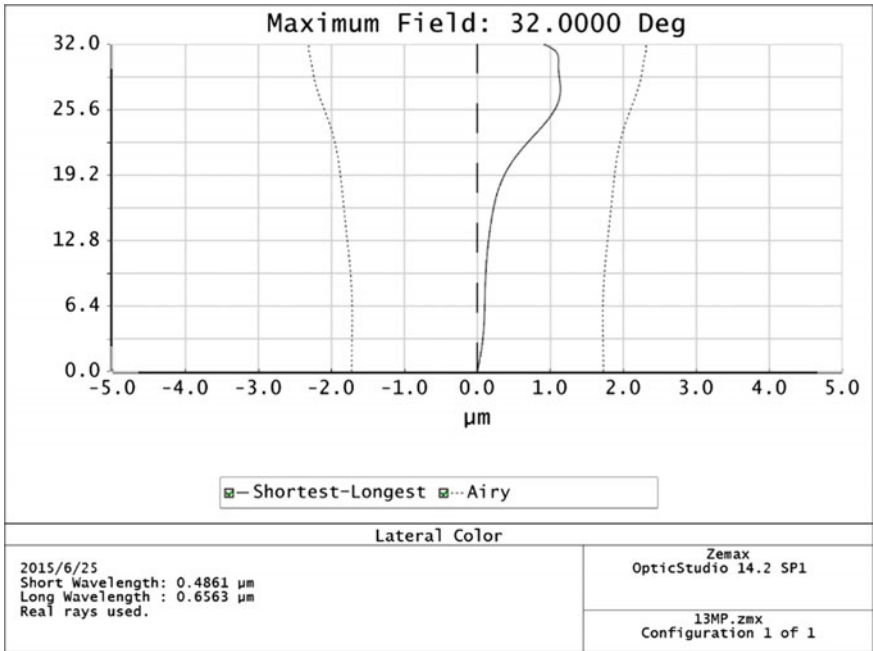


Fig. 6 Later color

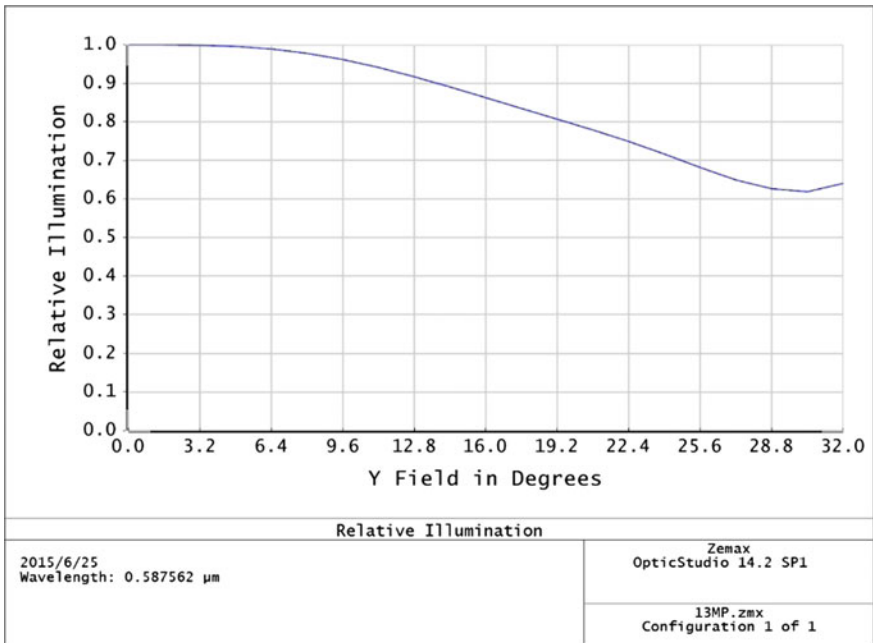


Fig. 7 Relative illumination

5.6 Relative Illumination

Figure 7 shows the relative illumination, which is the ratio of intensity at edge to the one at center. The relative illumination must be great than 50 % for all field view, which avoid edges of the field produce vignetting.

$$RI = \frac{E_{edge}}{E_{center}} \quad (2)$$

6 Results and Discussion

In this paper, we propose a 13 mega-pixels camera lens for tablet pc based on Zemax. The lens consists of 5 plastic aspheric lenses, an infrared glass filter and a sensor cover glass. The image sensor is OV13860 which is a 13 mega-pixels 1/2.6 inch CMOS chip from Omnivision, adopting the backside illumination technology. The lens has an effective focal length of 5.54 mm, a F number of 2.4, a field of view of 64°, and a total length of 7.001 mm. The MTF value of designed laptop lens are greater than 0.45 at 192.5 lp/mm. According to the above results, it can meet the requirements of tablet pc camera lens with high resolution.

References

1. Kay AC (1972) A personal computer for children of all ages. Xerox Palo Alto Research Center
2. Maxwell JW (2006) Tracing the dynabook: a study of technocultural Transformations. University of British Columbia
3. Apple Inc., <https://www.apple.com>
4. Bareau J, Clark P (2006) The optics of miniature digital camera modules. In: International optical design conference. <http://profiles.spiedigitallibrary.org/summary.aspx?DOI=10.1117%2f12.944333&Name=Matthew+P.+Rimmer>
5. Jin X, Liu Z, Chen J (2010) CMOS vision sensor with fully digital image process integrated into low power 1/8-inch chip. Chin Opt Lett 8(3):282–285. <http://profiles.spiedigitallibrary.org/summary.aspx?DOI=10.1117%2f12.944333&Name=Matthew+P.+Rimmer>
6. OminVision Co., OminBSI. <http://www.ovt.com/technologies/technology.php?TID=2>
7. Huang CY, Hung MC, Jhu TL, Tzeng GH (2010) Selecting a CMOS sensor by using a fuzzy MCDM framework. In: International conference on system science and engineering, pp 119–123
8. Bäumer S (2005) Handbook of plastic optics. Wiley-VCH, Weinheim

Two-Wavelength Optical Microscope Optical Axis Adjustment by Five Incident Parallel Laser Beams

Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee and Wei-Hsin Chen

Abstract The optical alignment of two-wavelength optical microscope (TWOM) has been successfully developed using the portable alignment device, right-angle prism and the CCD camera. In this method the optical axis of two-wavelength optical microscope is made parallel to the “five incident parallel laser beams” in the plane of incidence, by checking direction of these five reflected laser beams and changing the height and orientation of the MWOM. Right angle prisms are generally used to achieve a 90° light path bend. This produces a left-handed image and depending on the orientation of the prism, the image may be inverted or reverted. The prism uses to examine the vertical angle of between input laser beam and output laser beam of a plurality of light source. This novel method can rapidly and accurately align the optical axis of MWOM.

Keywords Right angle prism · Laser · Alignment

1 Introduction

The high-accuracy two-wavelength optical microscope provides a high contrast of an image and much information about a sample. It is suitably applicable for obtaining not only an enlarged image of a sample, but also an enough amount of information about the chemical composition. It permits, even with a very slight intensity of illuminating light, achievement of a long life of an excited state relative to multiple resonance absorption, a best contrast of image, and abundant information about a biological sample. Optical microscopes of various structures have so far been developed and used. As a result of recent progress made in peripheral technologies including laser and digital image technologies, optical microscope

F.-M. Yeh (✉) · S.-C. Lee

Department of BioIndustry Technology, Da Yeh University, Changhua, Taiwan
e-mail: optfmy@yahoo.com.tw

D.-C. Chen · W.-H. Chen

Department of Electrical Engineering, Feng Chia University, Taichung, Taiwan

systems of a further higher accuracy have been developed. The conventional microscopes have an insufficient picture contrast quality including, an insufficient amount of available information, and instability of observation. The TWOM includes a plurality of light sources, a wavelength selector independently varying the wavelength of the individual light sources, polarization plane rotators on the optical path for each light source and optical microscope.

Aligning optical axis of TWOM not only takes much of time, but also these methods cannot be used in UV and IR region. Because of the above problems, we establish a new optical technique to align the optical axis of a TWOM using the five parallel laser beams, laser triangulation range finder, prism device and CCD camera. There are some advantages in this method: (1) It is a rapid and simple alignment method. (2) It simplifies the alignment optical system and lowers the cost. (3) The prism device uses to examine the vertical angle of input and output laser beams. (4) Because the rotation mechanism of the five parallel laser beams arrangement, it adjusts the TWOM with more freedoms of orientation. This is a fast method for aligning the optical axis of a TWOM.

2 Basic Principle

Exciting electrons on the valence electron orbit of molecules composing the sample. When short wavelength light is necessary at this point, an SHG second harmonics oscillator 1 is installed as required after the dye laser 1 to reduce wavelength. The other the dye laser 2 is adjusted to a resonance wavelength λ_2 capable of causing transition from the primary excited state to a secondary one by exciting electrons on a saturated orbit. An SHG 2 is similarly provided after the dye laser 2 to reduce wavelength. These steps permit achievement of light beams of two wavelengths including the resonance wavelength λ_1 and wavelength λ_2 from the single light source. The propagation directions of the light of wavelength λ_1 and wavelength λ_2 are changed by mirrors 1 and 2, respectively, brought together by a beam splitter 2 into a single optical path. The light of each wavelength is enlarged by a telescope and irradiated through a condenser lens onto a sample. The molecules composing the sample are excited by these irradiated laser beams, and an absorption image is obtained through the then absorption process of electrons. A light-emitting image is available under the effect of fluorescence or phosphorescence emitted upon when the excited state decays to the ground state. The images form images on the retina of the observer by using an eye lens through a beam splitter 3 after enlargement by an objective. With a view to permitting observation of only the image of the beam of wavelength λ_2 by cutting the beam of wavelength λ_1 , a filter 1 may from time to time be inserted as required after the eye lens. The optical path of the beam reflected by the beam splitter 3 provided between the objective and the eye lens is changed, and a transmission image can simultaneously be observed by means of a TV camera installed separately from the eye lens on the thus changed optical path. To permit observation of only the image formed by the beam of wavelength λ_2 by cutting the

beam of wavelength λ_1 , a filter 2 a similarly be inserted between the beam splitter 3 and the TV camera. A polarization plane rotator 1 is provided between the mirror 1 and the beam splitter 2 on the optical path for the light of wavelength λ_1 , and a polarization plane rotator 2 is provided between the mirror 2 and the beam splitter 2 on the optical path for the light of wavelength λ_2 . This system is thus possible to obtain an absorption image from an absorption process of the molecules composing a sample from the ground state to the primary excited state. The absorption cross section during excitation from the ground state to the primary excited state is assumed to be σ_1 , the life of the primary excitation, τ , and the absorption cross section during excitation from the primary to the secondary excited states, a σ_2 . If the photon flux of the resonance wavelength λ_1 is I_0 , the irradiation time, T , and the density of molecules to be observed through resonance absorption, N_0 , then, the density N of molecules in the ground state at time T would be expressed by the following equilibrium equation:

$$\frac{dN}{dt} = -I_0\sigma_1N + \frac{N_0 - N}{\tau} \tag{1}$$

The density n of molecules in the primary excited state after the lapse of time T under initial conditions including $t = 0$ and $N = N_0$ is expressed by the following equation (Fig. 1):

$$n = \left[1 - \exp\left\{-\left(I_0\sigma_1 - \frac{1}{\tau}\right)T\right\}\right] \frac{N_0I_0\sigma_1\tau}{1 + I_0\sigma_1\tau} \tag{2}$$

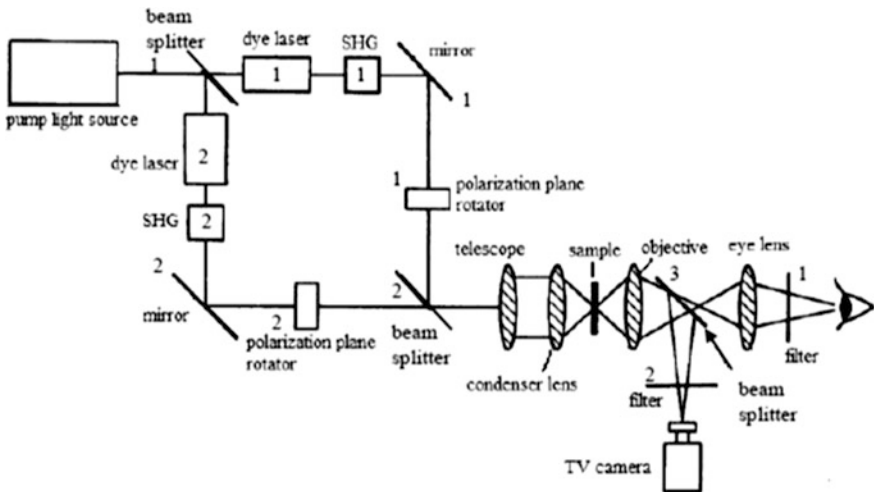


Fig. 1 A configuration diagram of the two-wavelength optical microscope

3 The Optical Alignment System

We develop a specific technique to align the optical axis of the TWOM. This alignment system can be applied to optical axis adjustment of multi-configuration optical system by using the different wavelengths of alignment laser diode, reflector, prism and other optical component. The optical alignment system in Fig. 2 is composed of: (1) alignment device shown in Fig. 3 is used to optical axis alignment of TWOM, shown in Fig. 4 (two-wavelength laser light generator optical axis adjustment) and Fig. 5 (Optical microscope optical axis adjustment), respectively, (2) two right angle prisms shown in Fig. 4 is used to optical axis alignment of the two-wavelength laser light generator. The Z is optical axis, (3) CCD Image

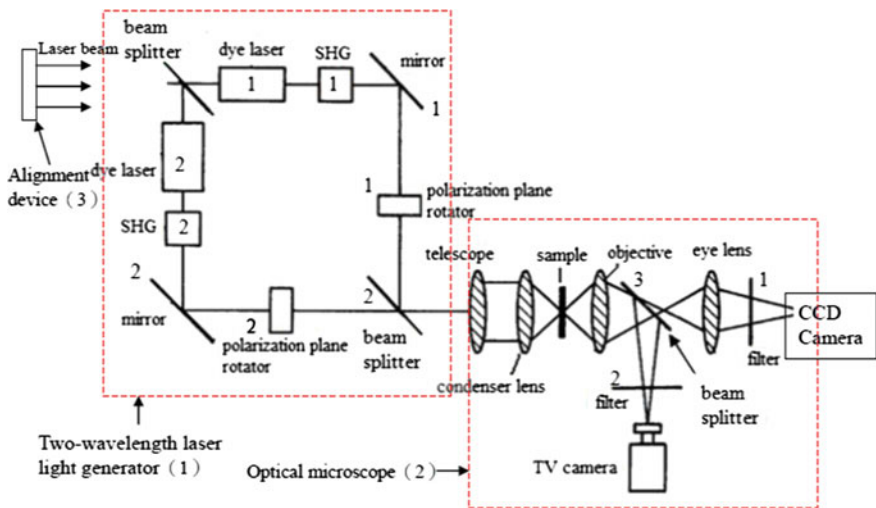
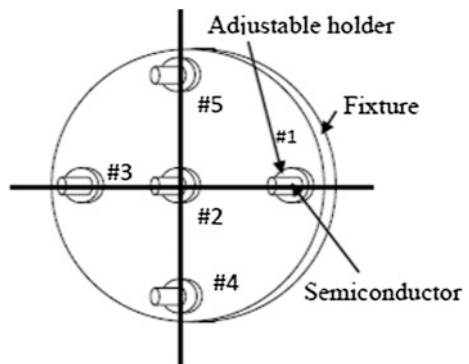


Fig. 2 The optical alignment system

Fig. 3 The schematic of the alignment device



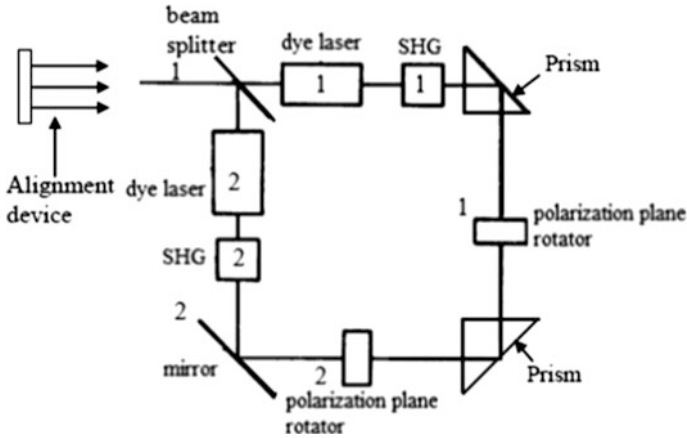


Fig. 4 Two-wavelength laser light generator optical axis adjustment

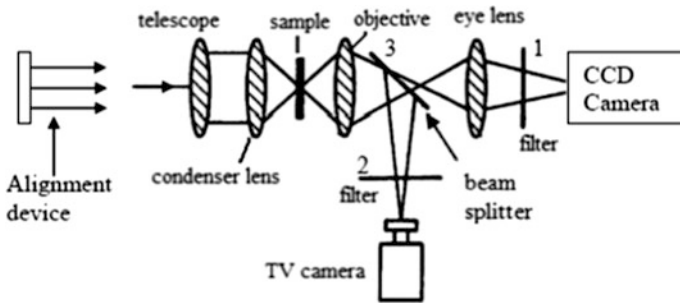


Fig. 5 Optical microscope optical axis adjustment

camera and pinhole, and (4) optical table and mount (not shown). These are described as followings.

The portable alignment device includes five parallel semiconductor lasers, adjustable mount and the rotation mechanism, and precise adjustable holder shown in Fig. 6. The semiconductor laser beams on the fixture are made parallel each other with precise adjustable holder and perpendicular to the surface fixture. An adjustable holder for mounting and orienting the semiconductor laser is set by a removable retaining laser within a mount of ring.

The rotation mechanism adjusts the initially horizontal and vertical laser beams to any rotation angle α as shown in Fig. 6. For example, the five laser beams lines, i.e., #1, #2, #3, #4 and #5 from right to left and down to up in Fig. 6. When it clockwise rotates α degree, it shows #1, #2, #3, #4 and #5 from original direction to α direction in Fig. 6. In the rotation mechanism, it can rotate any α degree to adjust the OAP mirror orientation we want.

Fig. 6 Photograph of the portable alignment device

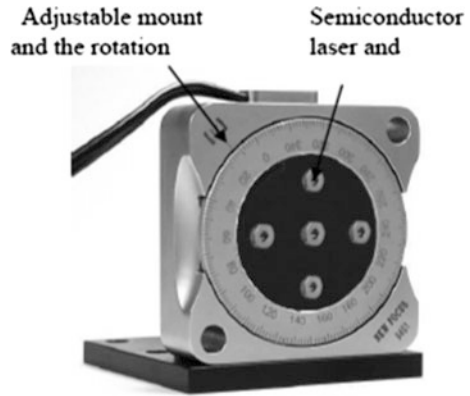
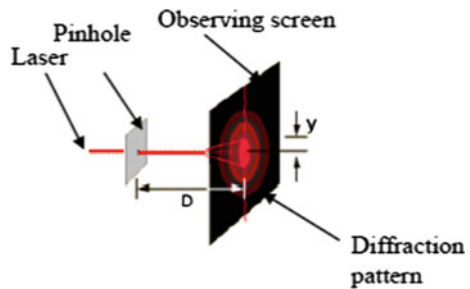


Fig. 7 Pinhole diffraction pattern



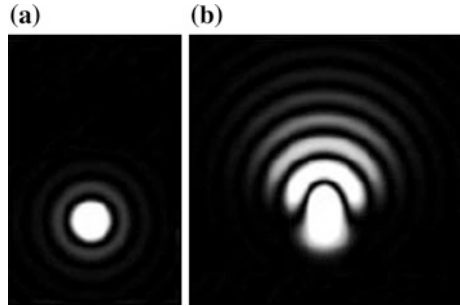
The pinhole has a size of 30 μm for alignment five laser beams and OAP mirror. When the diffraction pattern of concentric circular rings was appeared, the laser beam passes through the pinhole accurately. The diffraction patterns formed by a pinhole consist of a central bright spot surrounded by a series of bright and dark rings shown in Fig. 7. The diffraction experimental set-up includes semiconductor laser, pinhole and observing screen. We can describe the pattern in terms of the angle θ , representing the radius angle of each ring. If the aperture diameter is d (mm) and the wavelength is λ (mm), the radius angle θ of the first dark ring is given by

$$\sin \theta = 1.22(\lambda/d) = y/D \tag{3}$$

The distance between the pinhole and screen is D . The actual diffraction pattern of alignment pinhole is shown in Fig. 8. Figure 8a show the horizontal laser beam incident on the pinhole along optical axis of pinhole and Fig. 8b show the oblique laser beam incident on the pinhole at slope angle α . With this diffraction technique, the pinhole is used to precise align laser beam parallel to the TWOM of the alignment system.

The TWOM is brought into focus onto the CCD camera by looking for the minimum spot diameter. The focal spot is then centered using the XYZ translation stage attached to the CCD camera. By varying the separation of the TWOM and detector, best focus is located when the displayed beam size is smallest.

Fig. 8 The actual diffraction pattern for alignment pinhole.
a Accurate alignment.
b Alignment error



4 Experiments and Results

The conditions and specifications used for the experiment are listed in Table 1. Experiment procedures are divided into two parts as followings.

Table 1 The experiment conditions

<i>Optical microscope</i>	
Eye lens	10X, view field diameter 2.12 mm
Objective	10X, 0.25 N.A.
<i>Laser range finder</i>	
Range of measurement	0.02 up to 60 m
Measuring accuracy	± 1 mm
Time for a measurement	2.5–10 s
Light source	Red laser diode
<i>Semiconductor laser of alignment device</i>	
Wavelength (λ_p)	635 nm
Output power	3 mW
Beam divergence	< 2 mrad
Operating current	25 mA
Beam diameter	3.3 mm
<i>CCD camera</i>	
Resolution	752×582
Spectral range	350–1100 nm
<i>Right angle prism</i>	
Optical substrate	BK7
Entire face length and width (mm)	$56.6 * 40$
Beveled edges (mm)	40
Surface quality (scratch-dig)	20–10
Surface flatness ($\lambda @ 633$ nm)	$\lambda/8$
Angle tolerance	± 3 min

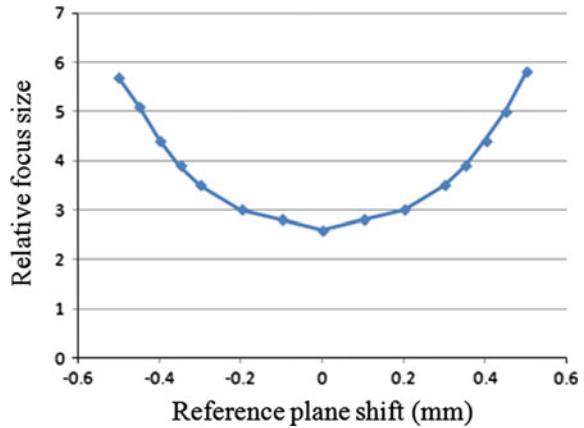
4.1 *Pre-experiment of Five Semiconductor Laser Beams Fine Adjustment*

- (1) The fixture of alignment device is vertically (perpendicular) put on the optical table such that laser beam #2 is nearly parallel to the surface of the optical table.
- (2) The precise adjustable holder is fine adjusted to make the laser beam #2 parallel to the surface of the optical table. To check the parallelism, we use the pinhole which moves on the optical table from position 1 to position 2 (about 50 cm distance, for system size = 1.5 m × 1 m) with a fixed height and has observed the pinhole diffraction pattern that the laser beam passes through the pinhole accurately shown in Fig. 8a.
- (3) The plane mirror is put on the optical table at 3 m distance from the alignment device. To adjust the plane mirror, the reflected beam #2 is made parallel to the optical table. To check the plane mirror perpendicular to the optical table, we use a piece of paper to chop reflected beam #2, and we can easily see how well the reflected beam #2 overlap at the output hole of laser beam #2.
- (4) Step (2) is repeated until the beam #4, #5, #3 and beam #1 is parallel to the surface of the optical table.
- (5) We use a piece of paper to chop reflected beam #4, #5, #3 and #1, and we can easily see how well the reflected beam #4, #5, #3 and #1 overlap at the output hole of laser beam #4, #5, #3 and #1, respectively.

4.2 *Experiment with TWOM*

- (1) Two-wavelength laser light generator including two prisms, mirror, and beam splitter is adjusted at first so that its optical axis is in the incident plane made by three horizontal parallel incident laser beams as shown in Fig. 4.
- (2) Two-wavelength laser light generator including two prisms, mirror, and beam splitter is adjusted again so that the three horizontal parallel reflected beams come back the same point of alignment device.
- (3) Three vertical parallel incident beams are parallel to optical axis by adjusting the two prisms, mirror, and beam splitter of two-wavelength laser light generator and the three vertical reflected beams also come back the same point of alignment device.
- (4) The two prisms are replaced by mirror and beam splitter as shown in Fig. 2 and optical axis alignment of two-wavelength laser light generator was repeated step (1) to step (3).
- (5) The optical axis alignment of optical microscope is the same two-wavelength laser light generator.

Fig. 9 Focus size when reference plane shifted



- (6) Capture the focus pattern at the reference plane (or near focal point) with the CCD camera. The reference plane is moved from left to right of the image plane to see if it fits small spot size.
- (7) The EFL and BFL of objective and eye lens are measured by laser ranger and trigonometry.

We measure best focus pattern curve that the reference focus size versus the position of the reference plane. Focus patterns sizes of reference plane are captured with CCD camera is shown in Fig. 9. The vertical coordinate is at different position of CCD camera. The relative focus size is normalized to reference focal point. The optical microscope is brought into best focus onto the CCD camera by looking for the minimum spot diameter. The eleven focus patterns from left of the image plane to right of it are measured. The 30 % of the center of the maximum intensity is defined as the focus pattern of the optical microscope.

The alignment of the optical microscope is finished when getting relative minimum focus size. This validates the experimental system. This alignment experiment repeatedly does twenty times. The test results show that the average EFL is 15.545 mm and the standard deviation is 0.755 mm or 0.5 %.

5 Conclusions

We have effectively presented the use of the five parallel laser beams alignment device, right angle prism and CCD camera to align the optical axis of TWOM. This alignment system can be applied to optical axis adjustment of multi-configuration optical system by using the different wavelengths of alignment laser diode, reflector, prism and other optical component. The advantages of the two-wavelength alignment

techniques are: (1) it is simple to operate; (2) it simplifies the optical test system and lowers the cost [1], (3) it is less expensive to maintain the equipments, and (4) a simple optical axis alignment of multi-configuration optical system technique.

Reference

1. Chrzanowski K (2007) Evaluation of infrared collimators for testing thermal imaging systems. *Opto-Elect Rev* 15(2):82–87

The Correlation Analysis Between the Non-contact Intraocular Pressure and Diopter

Feng-Ming Yeh, Der-Chin Chen, Shih-Chieh Lee
and Ching-Chung Chen

Abstract This paper discusses the relevance of the non-contact intraocular pressure (non-contact tonometer, NCT) and the diopter value measurements of refractive errors. Methods for the measurement of both eyes IOP of 192 patients, and binocular eye refraction as well, then the measured data were statistically analyzed. Research on age, gender, which eye, refractive power and refractive status (i.e. myopia, face, and hyperopia) affects the values of non-contact tonometer measurements. The results of refractive measurements were as follows: myopia, emmetropia and hyperopia accounted as 47.92, 15.62 and 36.46 % respectively. IOP measurement of all ages: age of 10 years old group (17.74 ± 3.34) mmHg, 10–20 years old group (17.85 ± 3.34) mmHg. According to gender, the male gender (17.94 ± 3.12) mmHg, and female (17.92 ± 3.23) mmHg. Group of the different eye, right eyes IOP was (18.05 ± 3.12) mmHg, left eyes IOP was (17.88 ± 3.17) mmHg. According to refractive errors: myopia group IOP was (17.83 ± 2.95) mmHg, hyperopia group IOP was (18.12 ± 2.94) mmHg, and emmetropia group IOP was (17.81 ± 3.16) mmHg. The results are: As we age, non-contact tonometry values tended to increase; gender has no effect on IOP measurement; detecting IOP found the value of the right eyes higher than the left eye; the refractive errors of (hyperopia, myopia and emmetropia); The values of IOP of myopia and hyperopia are higher than the emmetropia eyes.

Keywords IOP · Refractive status · Diopter

F.-M. Yeh (✉) · S.-C. Lee
Department of BioIndustry Technology, Da Yeh University, Changhua, Taiwan
e-mail: optfmy@yahoo.com.tw

D.-C. Chen
Department of Electrical Engineering, Feng Chia University, Taichung, Taiwan

C.-C. Chen
School of Optometry, Kang-Ning University, Tainan, Taiwan

1 Introduction

IOP is the pressure of the contents of the eye acting on the wall of the eye. Normal IOP will maintain normal morphology of eyes, provide nutrition and metabolism to an avascular structure such as intraocular lens and maintain the intraocular fluid circulation. Glaucoma is the most relevant IOP eye disease. IOP related to aqueous humor production, aqueous discharge and the pressure of sclera venous. Generation of aqueous humor formation is a major factor of IOP.

Under normal circumstances, aqueous humor production rate, aqueous discharge rate and the volume of the content of the eye are in a dynamic equilibrium, if the imbalance could lead to pathological intraocular pressure. To know the IOP of normal and pathological values is important for clinicians. Normal IOP range is 10–21 mmHg, if IOP > 21 mmHg as is abnormal. There are few normal whose IOP higher than 21 mmHg, but it does not cause any damage to the optic nerve and visual function. However, there are some glaucoma patients with normal range of intraocular pressure has caused glaucomatous optic nerve and visual impairment.

The clinical significance of intraocular pressure measurement is very important. The non-contact tonometer (non-contact tonometer, NCT) is very easy to operate, generally used as part procedure of visual function examination. We analyze the factors of non-contact tonometry values of refractive errors patients. The results reported below, to guide the correct clinical assessment of intraocular pressure.

2 Method

Study object is 192 patients chosen from the vision screenings during 2014–2015, aged from 7 to 20 years old. Among them, 105 cases of men (accounting for 54.68 %), female 87 cases (accounting for 45.32 %). All the selected objects excluded eye disease other than refractive errors. The process of visual function examination is as follow which shown in Fig. 1.

Subject category: reception guides, leads the subject to fill in personal information, case history, asks and understands subject visual needs, if the patient had old glasses, measurement of the power of spectacles.

Preliminary functional tests, the tests as of the following items:

Color vision test is to screen the acquired or genetic color vision defects.

Stereopsis measures the angle of stereopsis and if has any suppression.

Worth's four-dots is to assess the fusion ability of distant and near objects, to check any suppression or eye deviation, and whether any unilateral suppression spots.

Accommodative facility to measure the ability of accommodative accuracy and flexibility.

Near point accommodation test is measure the amplitude of accommodation by converting the near point accommodative distance and the focusing ability of the lens.

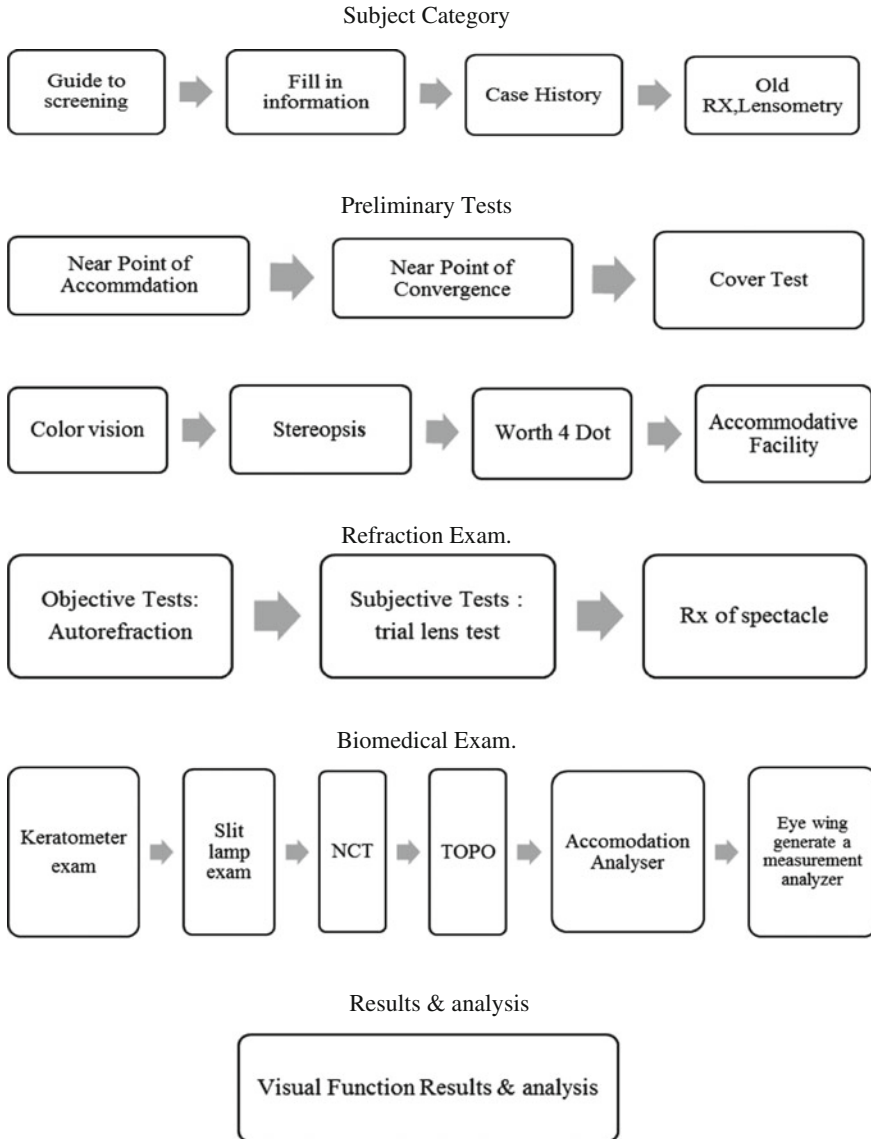


Fig. 1 Visual function screening flow chats

Near point convergence is measured the eye inward ability while maintaining fusion.

Cover test is mainly used to differentiate patients with tropia or phoria, alternating or permanent.

Objective auto refractor measurements, using NIDEK Company's AR-310A (auto kerato-refractometer), the annular measurement plus SLD (Super Luminescent Diode), ultra-high sensitivity CCD. All information collected from pupil area within the range of 4, 2 mm minimum measurable pupil diameter, and increased accuracy significantly to small pupil persons. Three measurements per eye, if 3 times result difference of more than 0.5 D then re-do the measurement. The refractive errors data is the value which is automatically averaged three times by auto refractor. Then through the subjective refractive measurements to decide final prescription of glasses and trial frame that.

Biotechnology detection instruments contains:

Keratometer can be used objectively to measure in the degrees and corneal curvature.

Slit lamp examination can be used to assess physiological state of the eye, including the assessment of the cornea, conjunctiva, anterior chamber, iris, lens, etc. Also could check entropion, inverted eyelashes, conjunctivitis, keratitis, cataract, dry eye and other eye diseases.

Non-contact tonometer detected by airflow flattening the cornea, so the head of the tonometer is not in contact with the eye when eye pressure is measured. When measuring IOP, the discharge air pressure rapidly increased, the pressure increase with time is linear. In order to detect a flatten area of the cornea, the instrument also issued a directional beam to the cornea, the reflected beam is photocell received. When the central cornea to flatten area of 3.06 mm diameter, the emitted light to reach the maximum amount photocell. In other words, the air pressure is emitted when the maximum reflected light is measured IOP. IOP within the normal scope, the relevance is very good, and the use of non-contact tonometer also avoid cross-infection, simple non-contact tonometer operation, generally without anesthesia, before the test, patients should be told not afraid of the air spray. Non-contact tonometer measures the moment of the intraocular pressure, so should be repeated few times and taking the mean value in order to reduce errors. Non-contact tonometer has a calibration system can be calibrated periodically.

In this study, the instrument which measure the intraocular pressure produced by the Japanese company Topcon CT-80-type non-contact tonometer, which employs spectral region OCT technology combined with non-mydratic retinal camera, infrared light-based lighting sources, excluding glare simple bright light and increase the comfort of the patients, with an array of built-in LCD fixation target, complemented by focusing divisive 3D OCT-1000 operation. The signals can be adjusted to best fit the situation, ensure the accuracy of the measurement. There are a double sense of measurements: one for the light sensor, and the other is the use of quantitative air pressure sensor ejected onto the cornea, there will be pressure to produce, by reflect back pressure power to measure IOP values.

Measurement process start from asking the patients to sit upright, adjust the lift height, the patient comfortable seating and avoid excessive bend, rise, asking the patient place the low jaw on the lower jaw holder, forehead against the forehead band, and then adjust the height of the jaw holder, adapted to measure the eye level

height. Explain to the patient prior to the measurement of intraocular pressure gas discharge situation, to avoid tension, measured three times per eye, 3 times result difference of more than 3 mmHg were re-measured to ensure the accuracy of measurement. Take 3 times of the average value measurements.

Corneal Topography instrument can measure the entire cornea curved arc, previous over the contour shape graph analysis, accurate measurement of corneal thickness, not only it is central thickness, but also the shape of the entire thickness of the cornea whether it is normal or not. It is also analyzed corneal front and back contours to diagnosis whether there is keratoconus degenerative disease or not. So early detection and prevention can to reduce any problems which can cause blindness.

Eye accommodation adjust measuring analyzer, measurement and analysis of accommodation of the eye.

Eye wing generate a measurement analyzer, check conjunctiva whether in the long-term exposed to ultraviolet or infrared radiation, will the conjunctiva cells degenerate, becomes chronic inflammation, tissue becomes thicker, grown pterygium or not.

After analysis of all the visual function test results, the explanation and advice will be given to patients.

3 Result

Research on age, gender, which eye, refractive power and refractive status (i.e. myopia, face, and hyperopia) affects the values of non-contact tonometer measurements. The results of refractive measurements were as follows: myopia, emmetropia and hyperopia accounted as 47.92, 15.62 and 36.46 % respectively.

IOP measurement of all ages: In the age of 10 years group: (17.74 ± 3.34) mmHg, 10–20 years old group (17.85 ± 3.34) mmHg, as age grow, the IOP tend to increase. As shown in Table 1.

According to gender, the male gender (17.94 ± 3.12) mmHg, and female (17.92 ± 3.23) mmHg. There are not significant different related IOP and gender. As shown in Table 2.

Table 1 Age versus IOP

Age	Number	IOP
Below 10 years old	75	17.74 ± 3.34
10–20 years old	117	17.85 ± 3.34

Table 2 Gender versus IOP

Sex	Number	IOP
Male	105	17.94 ± 3.12
Female	87	17.92 ± 3.23

Table 3 Eye different versus IOP

Eye different	Number	IOP
Right eye	192	18.054 ± 3.12
Left eye	192	17.88 ± 3.17

Table 4 Refractive errors versus IOP

Refractive errors	Number	IOP
Hyperopia	70	18.12 ± 2.94
Emmetropia	30	17.81 ± 3.16
Myopia	92	17.83 ± 2.95

According to right eye or left eye: Group of the right eyes IOP was (18.05 ± 3.12) mmHg, left eyes IOP was (17.88 ± 3.17) mmHg. There is tendency that right eye IOP higher than IOP of left eye. As shown in Table 3.

According to refractive errors: myopia group IOP was (17.83 ± 2.95) mmHg, hyperopia group IOP was (18.12 ± 2.94) mmHg, and emmetropia group IOP was (17.81 ± 3.16) mmHg. Non-contact IOP of myopia group is higher than that of Emmetropia. Non-contact IOP of hyperopia group is also higher than that of emmetropia. As shown in Table 4.

4 Discussion

192 cases of patients received vision screenings, non-contact tonometer are measured: the right eye IOP values ranging from 11–27 mmHg, left IOP values ranging from 11–30 mmHg. Most of IOP in patients examined were within the normal range, very few intraocular pressure exceeds 21 mmHg, the highest is up to 30 mmHg.

IOP study found, procedure explanation can help to eliminate the tension of patients, adjusting posture, when measured pull the upper lid from the orbital rim (do not oppress the eye), to avoid eyelashes, eyelids resulting measurement error.

Comparison myopia, hyperopia group with emmetropia group respectively, myopia IOP (17.83 ± 2.95) mmHg, hyperopia group (18.12 ± 2.94) mmHg, emmetropia group (17.81 ± 3.16) mmHg. This result is similar to the results from other researchers [1], myopia IOP higher than Emmetropia's. Studies have found that the higher myopia is, the higher values corresponding intraocular pressure is [2]. But studies have shown [3], when IOP measured, the degree of shape of the eye wall deformation is related to eye ball wall material characteristics and the form of the eye wall. Material Features of eye ball wall is mainly refers to the degree of elasticity of the eye wall. Myopia, especially high myopia, as the axial extension of the eye content expansion, leading the eye wall weakened, reduce wall stiffness ball and makes intraocular pressure measurements lower than the actual IOP. Therefore, for patients with high myopia, if the IOP is high value, should be careful examined, to make sure there is no glaucoma.

The study found that the IOP of age of 10 years old was (17.74 ± 3.34) mmHg, 10–20 years old group (17.85 ± 3.34) mmHg, the effect of age exists for IOP and IOP increased as age increased. Most studies have shown that IOP was positively correlated with age [2, 4]. IOP may vary with age, blood pressure, pulse rate, obesity, leading to increased IOP. Clinical also observed teenagers IOP in the developmental stages at a high state. But the other studies suggest that there is no correlation between the two [4], and even research [1] found that with age, IOP decreased.

This study find eye different affect IOP. Right eye IOP was (18.05 ± 3.12) mmHg, left eye IOP was (17.88 ± 3.17) mmHg. But from the clinical analysis, the difference was not statistically significant.

Gender analysis in this study shown that influence of gender on intraocular pressure, male (17.94 ± 3.12) mmHg, female (17.92 ± 3.23) mmHg, IOP of male slightly higher than IOP of female, but from the clinical analysis, the difference was not statistically significant.

References

1. Stamper RL, Lieberman MF, Drake MV (2001) Becher-shaffer's diagnosis and therapy of the glaucomas, 7th edn. CV Mosby Co, St Louis, pp 75–78
2. Liu L, Zhou Y-H (2006) Multiple factors analysis of the factors affecting intraocular pressure of myopic patients. *Recent Adv Ophthalmol* 26(2):133–136
3. Foster PJ, Wong JS, Wong E et al (2000) Accuracy of clinical estimates of intraocular pressure in Chinese eyes. *Ophthalmology* 107(10):1816–1821
4. Shimmyo M, Ross AJ, Moy A et al (2003) Intraocular pressure, Goldmann applanation tension, corneal thickness, and corneal curvature in Caucasians, Asians, Hispanics, and African Americans. *Am J Ophthalmol* 136(4):603–613

Automated Tool Trajectory Planning for Spray Painting Robot of Free-Form Surfaces

Wei Chen and Yang Tang

Abstract Automated spray painting is an important process in the manufacturing of many products. In order to ensure computational efficiency, a new tool trajectory optimization scheme based on T-Bézier curve is developed. And a T-Bézier basis is presented in trajectory optimization problem. The tool trajectory is formed through offsetting the distance between spray gun and the free-form surface along the normal vectors. Automotive body parts, which are free-form surfaces, are used to test the scheme. The results of experiments have shown that the trajectory planning algorithm achieves satisfactory performance. This algorithm can also be extended to other applications.

Keywords Spray painting robot · Trajectory planning · Surface modeling · T-Bézier curve · Experiment

1 Introduction

Automated spray painting is an important process in the manufacturing of many products, such as automobiles, furniture, airplanes; etc. Surface modeling is the first step of trajectory optimization for spray painting robot. The shape of workpiece and the tool parameters can strongly influence the quality of painting [1]. In order to achieve the new spraying operation standards, new surfaces modeling of workpiece algorithms for spray painting robot are active research for many years. Surface modeling based on parametric surface in spray painting is presented by Antonio et al. [2]. For a product with a parametric surface; two basic approaches can be

W. Chen (✉)

School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, China

e-mail: cwchenwei@aliyun.com

Y. Tang

School of Science, Jiangsu University, Zhenjiang, China

e-mail: ty800117@ujs.edu.cn

applied. The first one is called the section approach. Spray painting robot trajectories are generated by intersecting the target surface with a series of parallel equidistant section planes. The second is called the offset curve approach that generates a start curve on the target surface, and then constructs the subsequent paths by offsetting the start curve along a family of curves orthogonal to the start curve [3, 4]. However, both the two basic approaches are too complicated to model large free-form workpiece surfaces for spray painting [5].

Now, achieving uniform paint thickness for free-form surfaces is still a challenging research topic due to the complex geometry of free-form surfaces [6]. And automated trajectory planning has been widely studied. Chen and Zhao [7] proposed a new optimization algorithm of the path planning for spray painting robot of workpiece surfaces. But the algorithms are too complicated to optimize spray painting trajectory on large free-form surfaces in automobile manufacturing. Chen et al. [8, 9] developed a automatic tool path planning for a free-form surface, and the experimental results illustrate the feasibility and availability of the method. But their algorithms couldn't resolve tool trajectory optimization problem. Yu et al. [10], Gasparetto [11] developed an automatic tool path planning for a free-form surface. Li et al. [12] proposed a new trajectory optimization scheme and a model of paint deposition rate for a free-form surface was established. Li et al. [13] proposed trajectory optimization of spray painting robot for a free-form surface based on adapted genetic algorithm. However, due to the process is complex and very time-consuming, their algorithms couldn't resolve robot trajectory optimization problem in automotive spray painting. The paint thickness function for free-form surfaces is not considered, and the optimal time is not satisfying.

In this paper, a new trajectory optimization scheme for large free-form surfaces in automobile manufacturing is developed. And a free-form surface model is approximated by a set of flat patches. Each patch is treated individually to generate robot trajectories. And a new trajectory optimization scheme by T-Bézier curve is developed. Automotive body parts, which are large free-form surfaces, are used to test the scheme. The results of experiments have shown that the trajectories optimization algorithm achieves satisfactory performance.

2 The Model of a Large Free-Form Surface

The paint deposition rate function on a plane according to the experiment data is considered. And assuming that the shape of spray painting from the gun is a cone and the distribution model of spray is shown in [14].

To obtain time-efficient spray painting robot trajectories and sufficiently utilize the workspace of the robot, a grid approximation of a free-form surface is adopted in CAD modeling. The CAD model of a free-form surface can be formulated as:

$$M = \{T_j : j = 1 \dots M\} \tag{1}$$

where T_j is the j th grid on the free-form surface; M is the number of grids.

During spray painting, a free-form surface is only covered by a spray cone at each time instant. The patch forming method is based on minimizing the maximum deviation angle of spray cones. To optimize the paint thickness on a free-form surface, the maximum deviation angle of every spray cone has to be minimized. Assume that there are N grids which are covered by a spray cone and the spray cone is projected to a plane. The normal of the plane must be a vector which minimizes the inaxiniuin deviation angle of tile spray cone.

A flat patch is a collection of connected grids, which correspond to a certain area of continuous part surface and satisfy the constraint: the angle between the normal of any grid in the patch and the average normal of the patch is within certain threshold. The average normal of a patch is defined as follows:

$$\vec{n}_a = \frac{\sum_{j=1}^p s_j \vec{n}_j}{\sum_{j=1}^p s_j} / \left\| \frac{\sum_{j=1}^p s_j \vec{n}_j}{\sum_{j=1}^p s_j} \right\| \tag{2}$$

Here s_j is the area of grid T_j , and \vec{n}_j is the normal of T_j .

An area-weighted average is employed to calculate the average normal, which reflects the fact that bigger grids have more contribution to the average normal than smaller ones. The average normal is used to indicate the direction of a patch.

The maximum-area-direction of a flat patch is the direction that when the patch is orthographically projected along this direction, the image of the patch has the maximum area. The image area of a patch in orthographic projection with scaling factor $m = 1$ is:

$$S = \sum_{j=1}^p s_j |\vec{n}_j \cdot \vec{v}_a| \tag{3}$$

Here \vec{v}_a is the direction of orthographic projection. To find \vec{v}_a such that S is maximum, we have:

$$\frac{dS}{d\vec{v}_a} = 0 \tag{4}$$

The results means the maximum-area-direction is the same direction as the area-weighted ‘‘average normal’’.

Assume that the material quantity on the flat surface is projected to a free-form surface and the maximum deviation angle of the free-form surface relative to the normal of the flat surface is β_{th} . The maximum deviation angle is the maximum angle between the normal of the free-form surface and the flat surface. Then the

material quantity q_s of each point s on the free-form surface can satisfy the following inequality without considering the tool standoff variation:

$$\bar{q}_{\min} \cos(\beta_{th}) \leq q_s \leq \bar{q}_{\max} \quad (5)$$

Here \bar{q}_{\max} , \bar{q}_{\min} are maximum and minimum material quantity. Then the material quantity of each point q_s in the free-form surface satisfies the constraints:

$$|q_s - \bar{q}_d| \leq q_w \quad (6)$$

Here \bar{q}_d , q_w are average material quantity and the maximum material thickness deviation.

3 Trajectory Optimization Based on T-Bézier Curve

Bézier curves are widely used for constructing free-form curves and surfaces [15]. It is well known that the Bézier basis is a basis for the space of degree- n algebraic polynomials as:

$$T = \text{span}\{1, t, t^2, \dots, t^n\} \quad (7)$$

However, since this basis is rational and polynomial, it would be complicated to use for the tool trajectory of a spray painting robot. This is because each point is associated with six parameters which define the position coordinates and the orientation vector of the spray. In particular, repeated differentiation of Eq. 7 produces curves of very high degree [16]. In order to ensure computational efficiency, finding new bases of Bézier model in new spaces seems to be the only way.

In this paper, a new T-Bézier basis is presented in tool trajectory optimization problem of spray painting robot. We first give four initial functions:

$$\begin{aligned} B_{0,3}(t) &= (\cos t)^4 \\ B_{1,3}(t) &= 2(\cos t)^4(\sin t)^2 \\ B_{2,3}(t) &= 2(\sin t)^4(\cos t)^2 \\ B_{3,3}(t) &= (\sin t)^4 \end{aligned} \quad (8)$$

where $t \in [0, \frac{\pi}{2}]$. For $n > 3$, T-Bézier basis functions are defined as:

$$B_{i,n}(t) = (\cos t)^2 B_{i,n-1}(t) + (\sin t)^2 B_{i-1,n-1}(t) \quad (9)$$

where $B_{i,n}(t) = 0$ for $i > n$ or $i < 0$.

With this basis, the curves share most of the properties as those of the Bézier curves in polynomial space. The T-Bézier basis have the properties as follows:

(1) Partition of Unity:

$$\sum_{i=0}^n B_{i,n}(t) = 1 \tag{10}$$

(2) Positivity:

$$B_{i,n}(t) \geq 0 \tag{11}$$

So T-Bézier basis is a blending system.

(3) Properties at the endpoints:

$$\begin{aligned} B_{0,n}(0) &= B_{n,n}\left(\frac{\pi}{2}\right) = 1, \\ B_{0,n}\left(\frac{\pi}{2}\right) &= B_{n,n}(0) = 0, \\ B_{i,n}(0) &= B_{i,n}\left(\frac{\pi}{2}\right) = 0, 0 < i < n \end{aligned} \tag{12}$$

(4) Linear independence: $B_{0,n}(t), B_{1,n}(t), \dots, B_{n,n}(t)$ are linear independent.

(5) Symmetry:

$$B_{i,n}(t) = B_{n-i,n}\left(\frac{\pi}{2} - t\right) \tag{13}$$

(6) B-basis property: $\{B_{0,n}(t), B_{1,n}(t), \dots, B_{n,n}(t)\}$ is the normalized B-basis of the space $span\{1, \cos t, \dots, \cos nt\}$. By the properties (1) and (2), we have that T-Bézier Basis is a totally positives basis.

A T-Bézier curve $p(t)$ of order $n + 1$ is defined as follows:

$$p(t) = \sum_{i=0}^n B_{i,n}(t)V_i, t \in \left[0, \frac{\pi}{2}\right] \tag{14}$$

where $\{B_{i,n}(t)\}_{i=0}^n$ is the T-Bézier basis, V_i is the control point.

The geometric properties at the endpoints of the T-Bézier curves are obvious from those of the T-Bézier basis:

$$p(0) = V_0, p\left(\frac{\pi}{2}\right) = V_n \tag{15}$$

Especially for $n = 3$, suppose $V_0^{[1]}, V_1^{[1]}, V_2^{[1]}, V_3^{[1]}$ and $V_0^{[2]}, V_1^{[2]}, V_2^{[2]}, V_3^{[2]}$ are two adjacent sets of T-Bézier control points. The condition of position continuity

(C^0 continuity) is $V_3^{[1]} = V_0^{[2]}$, and the condition of target continuity (C^1 continuity) is that $V_2^{[1]}, V_3^{[1]}, V_0^{[2]}$ and $V_1^{[2]}$ are collinear, containing C^0 continuity.

The entire T-Bézier curve $P(t)$ must lie inside its control polygon spanned by $V_0, V_1 \dots V_n$. This property is a consequence of the property of the T-Bézier basis about partition of unity.

The control points of opposite order define the same curve in a different parameterization, just the opposite direction:

$$p(V_n, V_{n-1}, \dots, V_0; t) = p(V_0, V_1, \dots, V_n; \frac{\pi}{2} - t) \tag{16}$$

which can be checked by comparing the coefficients of V_0, V_1, \dots, V_n . on both sides of the equation. No plane intersects a T-Bézier curve more often than it intersects the corresponding control polygon.

The shape of a T-Bézier curve is independent of the choice of coordinates, i.e.

$p(V_0, V_1, \dots, V_n; t) = \sum_{i=0}^n B_{i,n}(t)V_i$ satisfies the following two equations:

$$\begin{aligned} p(V_0 + r, V_1 + r, \dots, V_n + r; t) &= p(V_0, V_1, \dots, V_n; t) + r \\ p(V_0 * T, V_1 * T, \dots, V_n * T; t) &= p(V_0, V_1, \dots, V_n; t) * T \end{aligned} \tag{17}$$

Here r is an arbitrary vector, and T is an arbitrary $(n + 1) * (n + 1)$ matrix. If the control polygon is convex, then the corresponding T-Bézier curve is also convex.

4 Simulation

The algorithms are implemented in C++, and a free-form surface, shown in Fig. 1, is used to test the trajectory optimization algorithm. The control points on the free-form surface are $i = 212$. The tool trajectory is formed through offsetting the

Fig. 1 The grids approximation of a free-form surface

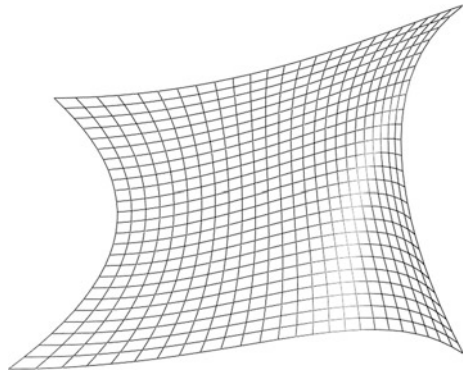


Fig. 2 The optimization trajectory on the free-form surface

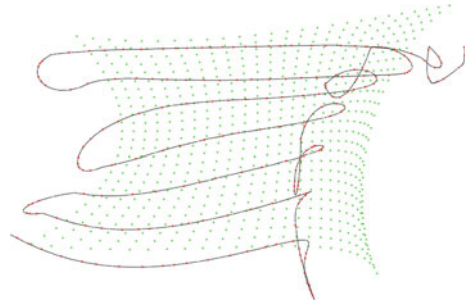


Table 1 The results for optimal trajectories planning of the simulation

	Material thickness of the 212 control points on the free-form surface
Average (μm)	49.7
Maximum (μm)	56.9
Minimum (μm)	48.1
Process time (s)	86

distance between spray tool and the free-form surface along the normal vectors. Then the optimization trajectory is generated using T-Bézier curve. The generated tool trajectory is shown in Fig. 2. Assuming that the shape of spray painting from the tool is a cone and the distribution model of spray is shown in [14]. Suppose the required average material thickness is $q_s = 50 \mu\text{m}$, and the max material thickness deviation is $q_w = 10 \mu\text{m}$. The spray radius is $R = 60 \text{ mm}$. The material deposition rate is:

$$f(r) = \frac{1}{15}(R^2 - r^2) \mu\text{m/s} \quad (18)$$

The results for optimal tool trajectories planning are summarized in Table 1.

5 Experimental Verification

Automotive body parts from a car company are tested. And the spraying parameter settings in experiment are the same to the simulation.

The first step is surfaces modeling. And a free-form surface model is approximated by a set of flat patches. The second step is trajectory optimization. We first determined the optimal movement patterns and sweeping directions, according to which the final trajectories are generated. Then the optimization trajectory is generated using the T-Bézier curves by the control points. The optimal tool trajectories

Fig. 3 The optimal tool trajectories on the car roof

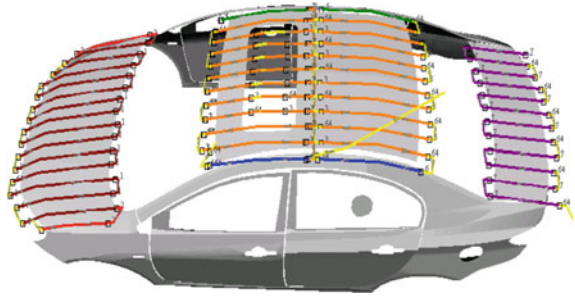


Fig. 4 The optimal tool trajectories on the car left body

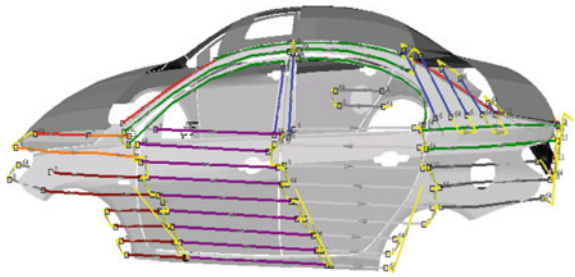


Fig. 5 Robotic spray painting experiment



on the car roof is shown in Fig. 3. The optimal tool trajectories on the car left body is shown in Fig. 4. And Fig. 5 shows the robotic spray painting experiment.

The thickness of the coating at randomly chosen points on the car body is measured using the Elcometer 456 automobile coating thickness gauge which resolution ratio is 1 μm . Figure 6 shows the results for material thickness of 200 chosen points along the direction of spray painting trajectory on the car body. The results for optimal tool path planning are summarized in Table 2.

Fig. 6 Material thickness of random chosen points on the car body

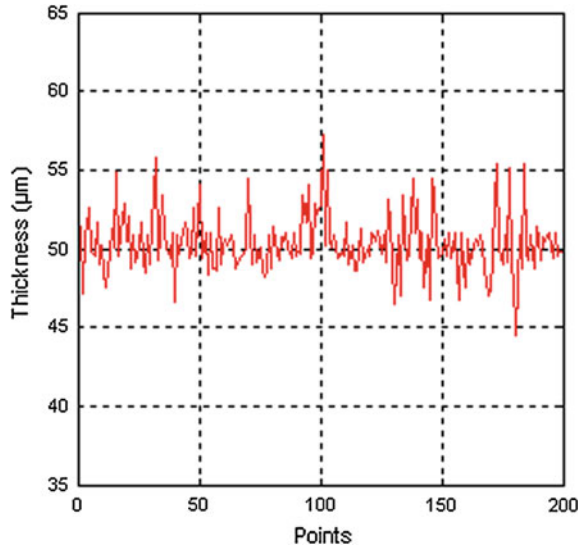


Table 2 The results for optimal trajectories planning of the experiment

	Material thickness
Average (µm)	49.5
Maximum (µm)	57.2
Minimum (µm)	44.5
Process time (s)	493

6 Conclusion

A grid approximation of a free-form surface is adopted in CAD modeling. And a free-form surface model is approximated by a set of flat patches. Each patch is treated individually to generate robot trajectories. A new trajectory optimization scheme based on T-Bézier curve is developed. Automotive body parts, which are free-form surfaces, are used to test the scheme. And the results demonstrate the advantages of the optimal trajectories planning algorithm. This algorithm can also be extended to other applications such as optimal tool path for free-form surface of cleaning robot or grinding robot, etc.

Acknowledgments This project is supported by University Science Foundation of Jiangsu province in China (Grant no. 14KJB510008), Senior talent Research Foundation of Jiangsu University (Grant no. 5503000046), Doctoral Scientific Research Foundation of Jiangsu University of science and technology (Grant No. 635031306) and National Natural Science Foundation Advance Research Project for Jiangsu University of science and technology (Grant No. 633031306)

References

1. Conner DC, Greenfield A, Atkar PN et al (2005) Paint deposition modeling for trajectory planning on automotive surfaces. *IEEE Trans Autom Sci Eng* 2(4):381–392
2. Antonio JK, Ramabhadran R, Ling TL (1997) A framework for trajectory planning for automated spray coating. *Int J Robot Autom* 12(4):124–134
3. Chen HP, Xi N, Sheng W et al (2005) Optimizing material distribution for tool trajectory generation in surface manufacturing. In: *Proceedings of the 2005 IEEE/ASME international conference on advanced intelligent mechatronics*, pp 1389–1394
4. Xia W, Wei CH, Liao XP (2009) Surface segmentation based intelligent trajectory planning and control modeling for spray painting. In: *Proceeding of the 2009 IEEE international conference on mechatronics and automation*, China, Changchun, pp 4958–4963
5. From PJ, Gunnar J, Gravdahl JT (2011) Optimal paint gun orientation in spray paint applications—experimental results. *IEEE Trans Autom Sci Eng* 8(2):438–442
6. Chen HP, Fuhlbrügge T (2008) Automated industrial robot path planning for spray painting process: a review. In: *4th IEEE conference on automation science and engineering*, USA, Washington DC, pp 522–527
7. Chen W, Zhao DA (2013) Path planning for spray painting robot of workpiece surfaces. *Math Probl Eng* 2013(8). doi:[10.1155/2013/659457](https://doi.org/10.1155/2013/659457)
8. Chen HP, Sheng WH (2011) Transformative industrial robot programming in surface manufacturing. In: *2011 IEEE international conference on robots and automation*, China, Shanghai, pp 6059–6064
9. Sheng WH, Chen HP, Xi N, Tan JD (2004) Optimal tool path planning for compound surfaces in spray forming processes. In: *IEEE international conference on robotics and automation*, USA, New Orleans, pp 45–50
10. Yu SR, Cao LG (2011) Modeling and prediction of paint film deposition rate for robotic spray painting. In: *Proceedings of the 2011 IEEE international conference on macaronis and automation*, China, Beijing, pp 1445–1450
11. Gasparetto A (2012) Automatic path and trajectory planning for robotic spray painting. In: *7th German conference on robotics*, German, Munich, pp 211–216
12. Li XZ, Landsnes OA, Chen HP (2010) Automatic trajectory generation for robotic painting application. In: *41st International symposium on robotics and 2010 6th German conference on robotics*, German, Berlin, pp 1–6
13. Li FA, Zhao DA, Xie GH (2009) Trajectory optimization of spray painting robot based on adapted genetic algorithm. In: *International conference on measuring technology and mechatronics automation*, ICMTMA 2009, China, Changsha, pp 907–910
14. Chen W, Zhao DA (2009) Tool trajectory optimization of robotic spray painting. In: *IEEE International conference on intelligent computation technology and automation*, China, ChangSha, pp 419–422
15. Juhász M, Ágoston R (2013) A class of generalized B-spline curves. *Comput Aided Geom Des* 30(1):85–115
16. Mainar E, Peña JM (2002) A basis of C-Bézier splines with optimal properties. *Comput Aided Geom Des* 19(10):291–295

The Research of Analysis Addiction of Online Game

Jason C. Hung, Min-Hui Ding, Wen-Hsing Kao,
Hui-Qian Chen, Guey-Shya Chen and Min-Feng Lee

Abstract The convenience and popularity of nowadays internet has more influenced as we know and the internet is actually essential on the daily life of individuals. Thus, internet addiction issues become topics as well as the main purpose of this study to identify and conclude how the generally problem—game addiction caused one’s interpersonal communication difficulties accompany disorder live and behavioral bias. This research refereed the remarkable internet addiction theory from S. Young Dr. Kimberly, and support from his contribution of Internet Addiction Rating Scale (Internet Addiction Test, IAT), we’ve turned it into a “game addiction questionnaire” continually, and via the questionnaire analysis to rank one’s addiction level to the hot games aims to find the time interval versus addiction for each hot game. The statistics of survey questionnaire questions are Likert 5 point questions, 20 questions to answer and calculate, and the value of average scores and time spent then be analysis to propose the Game Addiction interval with incremental increasing function, that we can recognize how long people will be addictive while they began to play and when they’ll appear to be such regular and continually addition, namely, we may hint the addition level of

J.C. Hung · M.-H. Ding · W.-H. Kao (✉)
Department of Information Technology, Overseas Chinese
University, Taichung, Taiwan
e-mail: star@ocu.edu.tw

J.C. Hung
e-mail: jhung@ocu.edu.tw

M.-H. Ding
e-mail: s10219203@ocu.edu.tw

H.-Q. Chen · G.-S. Chen · M.-F. Lee
Graduate Institute of Educational Measurement and Statistics,
National Taichung University of Education, Taichung, Taiwan
e-mail: gigi.baby@yahoo.com.tw

G.-S. Chen
e-mail: grace@mail.ntcu.edu.tw

M.-F. Lee
e-mail: antonio6561@gmail.com

games, then according these data analysis to specify different types of game addiction classification and provide better users notification whether it is hazardous, therefore prevent users to become the next victims of game addiction.

1 Background and Motivation

In 21st century, “Internet” is the most diverse and most growing technology in life. Internet which applies multiple application, constantly improved and innovation is a new technology impacts all the human’s aspects of life. Otaku (Otaku in meaning as it is used to refer to someone who stays at home all the time and doesn’t have a life) and Smartphone Addicts people’s behavior issues was discussed by news media, newspapers and magazines. Some people who died suddenly in playing game, teens addict internet in cyber cafe, use smartphone while walking to make the danger and even fantasize himself is the leading character in the online game to stabbed passengers on the metro, that reported in the newspaper society pages. Online game for the people is recreation and pass the time even though it can’t be proven the relation to the above all the issues with game addiction. It is a worth issue to investigate the game addiction that can transform human’s personality and life.

The game addiction model in this research can be input the name of game by user to acquire the gaming time addiction table, game addiction score and analysis of game content. Before the gaming, player will realize the condition of online game, and understand their game addiction from the analysis.

2 Specific Objective

“Game addiction” is a well-known term nevertheless less people to squarely indicate how big its harm and is a modern plagues with astonishing growth to become a health killer of all ages. It must be seriously discover “game addiction” influence, so draft these purpose for investigating in preventively.

- A. Explore “game addiction” forming reason and influence on body, mind, health and work effectiveness.
- B. Questionnaire survey to explore “game addiction” in the current situation, and analyze the characteristics of different games, according to their degree of analysis addictive draw of the table game time interval addiction, to find causes and definition of addiction range.
- C. Find a way to calculate the “game addiction” and set the degree of addiction.

3 Method

3.1 Schema

The process of the research is based on internet addiction theories of Dr. Kimberly S. Young in psychology and infrastructure, adapted for game addiction questionnaire research flow chart as shown in Fig. 1.

3.2 Steps and Analysis

This research proposal used questionnaires; the content of game addiction questionnaire contains “personal data”, “personal game use behaviors”, and “game addiction”. The process of questionnaire experienced collection of popular games, the preparation of questionnaires adapted subject and reliability analysis using SPSS, choose to 20 questions, take these five point to calculate the score, through the above to draw the chart, then get the time interval of game addiction table. The procedures step by step instructions are:

(A) collecting games

The questionnaire uses to top ten of the games for and analyzing the characteristics after collecting, fill in the other fields of the game, and the list will also analysis.

(B) Making the questionnaire

According to Internet Addiction Scale of Dr. Kimberly S. Young is adapted into a “game addiction” questionnaire.

1. The basic data survey include: name, school, department, grade, gender, birthdate, accommodation.
2. Behaviors of game using: average time to play games online in each day, last game played most often (multiple choices), based on the questions you most often preference, most attract you games, do you most often play games on one day to play the game of how much time?
3. Game addiction questionnaire shown in Table 1 that is according to 5 points to answer 20 questions. In order to assess the degree of addiction, the user needs to answer the following questions measuring stick.

(C) Analysis

There are 33 questions in questionnaire without filtered, and after reliability analysis, exclusion of the subject of reliability is less than 0.7, the remaining 20 questions is used to 5 point calculate the answer score. With reference to the

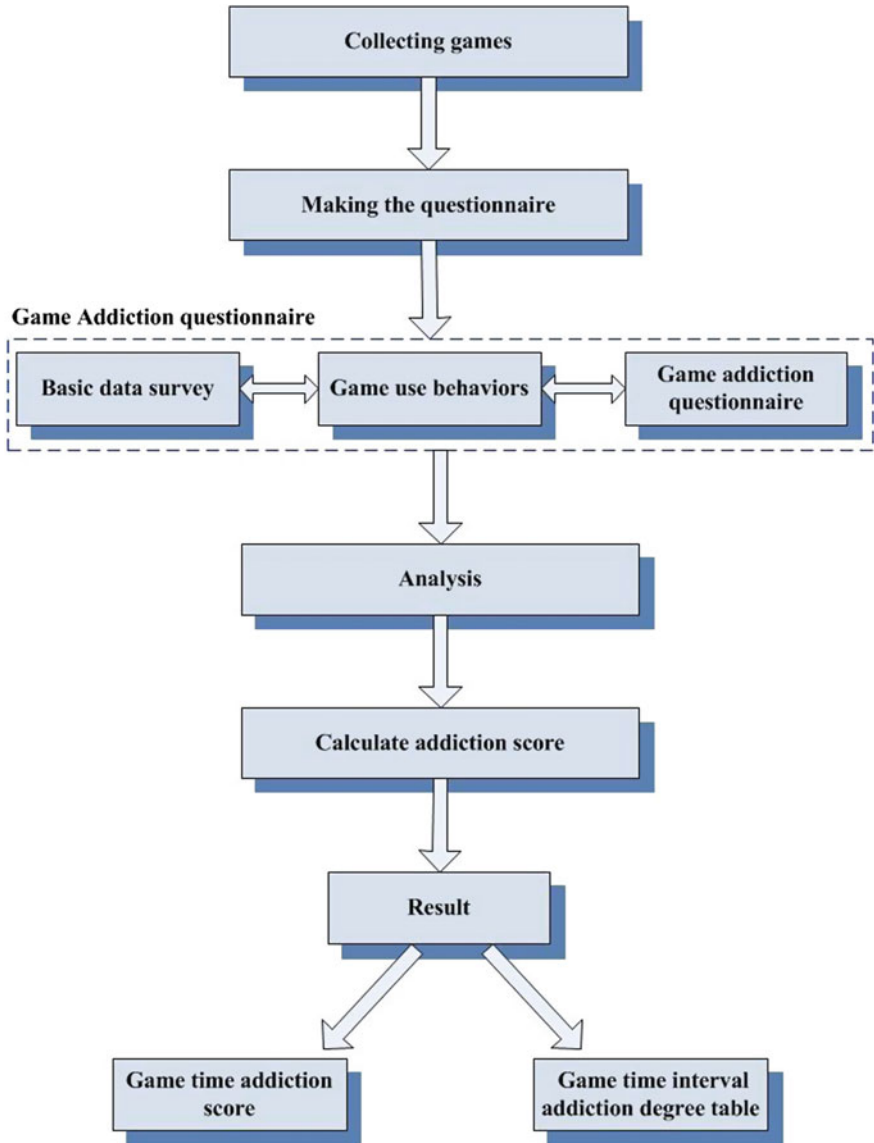


Fig. 1 Research flowchart

Internet Addiction Test (IAT) of Kimberly S. Young to provide addiction range, the range values are as follows:

- a. Normal range: 0–30 Score
- b. Mild: 31–49 Score
- c. Moderate: 50–79 Score
- d. Severe: 80–100 Score.

Table 1 Game addiction questionnaire

No.	Question
1	How often would you play game exceed to originally anticipated time?
2	How often do you put aside the completed or executed things and use time to play game?
3	How often do you play games get the excitement even interpersonal intimate interaction?
4	How often do you make new friends in the online game?
5	How often do you spend too much time to playing online game and being around people who complain or blame?
6	Do you spend too much time on the game which have begun to cause, academic setback?
7	How often do you have to do something else before opening the game?
8	How often do you have to open start the game before you do anything else?
9	How often do you play the game by recall pleasant thing to stop thinking troubled things?
10	How often do you expect to be able to start yet game to play?
11	How often do you fear less the game, life becomes boring, empty?
12	How often do you play the game sacrificing sleep at night?
13	When you playing the game, how often you tell yourself “just a few minutes”?
14	Have you ever ordered at bedtime will finish off the game before falling asleep?
15	Have you ever found yourself playing the game, in fact, don’t really feel interesting?
16	Do you feel like you have to spend more and more time in online game?
17	As long as there is free times will want to play the game?
18	When you finally have access to the game, feel happier and joyful; cheerful; delighted?
19	When I tried to cut down or stop using the internet, I would feel down, depressed or cranky
20	Suddenly you want to terminate the game, make you feel very bad

According to addiction factor type classified as: salience, mood changes, tolerance, and conflict, time limits, which have salient factor: Q4, Q11, Q12, Q14, mood changes: Q3, Q9, Q10, Q18, Q19, Q20, tolerance: Q2, Q7, Q15, Q16, Q17, conflict: Q5, Q6, Q8, time limit: Q1, Q13, as shown in Fig. 2.

(D) Calculate addiction score

The gaming time interval addiction degree table was drawn through the information collected and analyzed in each game of the degree addiction, and follows the incremental increasing function set addiction degree range.

(1) Each component defines its increasing value

If Average degree of addiction ≤ 30 then 0, Else Average degree of addiction $-30, = \Sigma \mid$ Average degree of addiction $-30 \mid$ (Table 2).

The tower of Saviors for example, the time interval addiction was shown in Fig. 3.

Fig. 2 Classification title game addiction

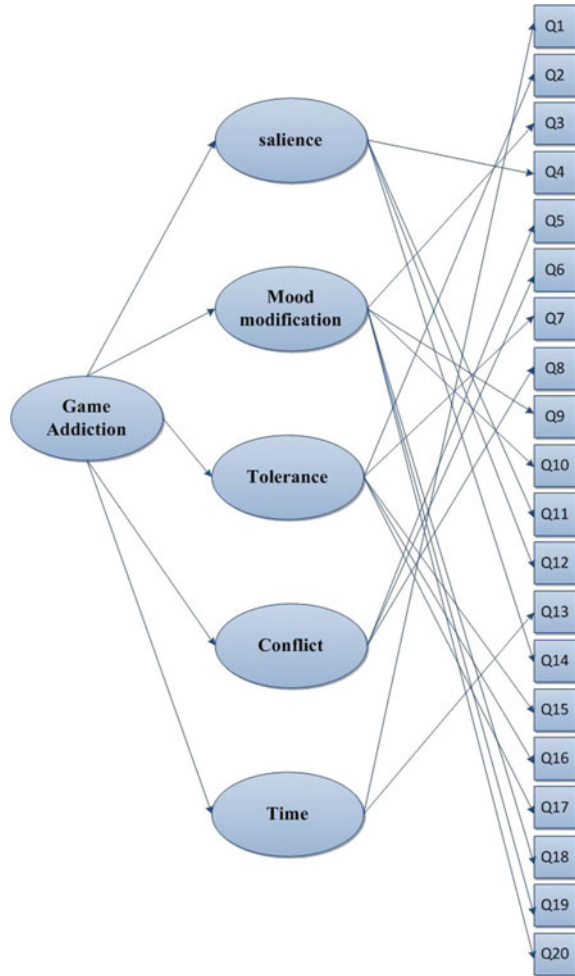


Table 2 Tower of saviors addiction time interval table

Time (h)	Addiction	Interval
1	38.4	8.4
2	43.9	13.9
3	53.0	23.0
4	53.0	23.0
5	56.0	26.0
6	59.5	29.5
7	59.5	29.5
8	60.0	30.0

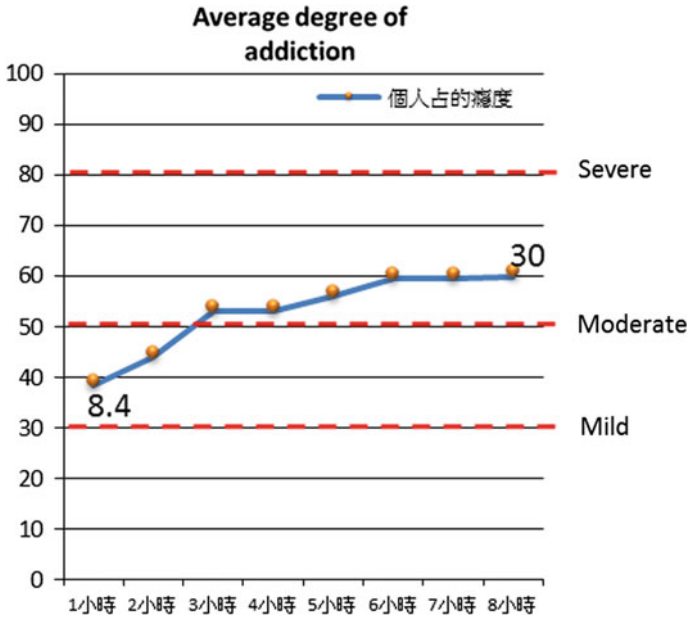


Fig. 3 Tower of saviors play average addiction graphs

To play 1 h in Tower of saviors the average degree of addiction degree is 38.4, 2 h is 43.9, 3 h is 53, 4 h is 53, 5 h is 56, 6 h is 59.5, 7 h is 59.5, 8 h is 60. The mild addiction between 31 and 49 degrees, the moderate between 50 and 79, and the severe between 80 and 100.

Figure 3 indicates the while playing 1 h in Tower of saviors, will be mild addiction, play more than 3 h will be the moderate addiction, and its addictive index showed a slope increasing, through Fig. 3 showed this game for playing longer will be more addictive, and this game is relatively difficult it may withdrawal.

(E) Addiction model present

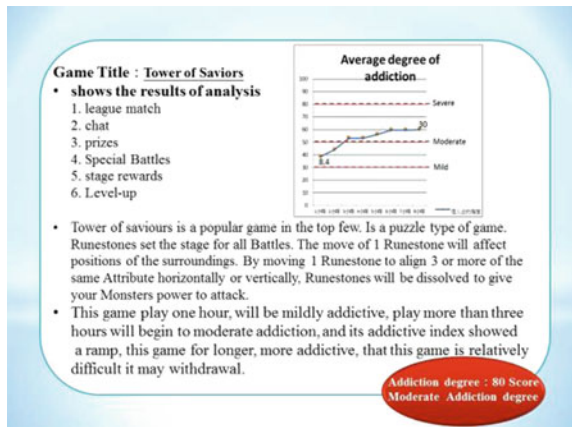
Figure 4 shows the search interface, the current analysis of games Tower of Saviors, LINE Running Gingerbread Man, national baseball 2014, Candy Crush, Poko Pang, League, Heaven, Puzzle & Dragons, the king of AVA Battlefield, tourism tycoon etc.

Figure 5 shows the results of analysis, accord with the user searching for a game, the content contains features, play, game time interval chart addiction, addiction scores, degree of addiction, addiction through time interval graph can hourly addiction know, the most important inform users of this addictive game play will be how long, and how high degree of addiction.

Fig. 4 Search interface



Fig. 5 Analysis of interface results



4 Conclusion

This model is divided into two kinds of “game addiction grade” and “personal game addiction grade”, the difference is that the game belongs to the public of addiction scores, a personal addiction classification is in accordance with the user’s data do their own analysis of addiction.

Game’s rating should not only ordinary level, protection level, parental guidance, restricted, but also play this game by the degree of addiction do classification.

In this paper, expect to establish a model and practical application in the game to prevent a starting point, the contribution of their paper as the following:

- (A) Model in which users can play this game is addictive, or warn users the game play more than how long it must to become addicted.
- (B) Let parents know their children play the game content through inquiry system.

- (C) The game ratings, according to their degree of addiction do classification (mild, moderate, severe addiction).

5 Future Research

Future research will focus on the content of individual game addiction, according to the type of subject classification: time constraints, salience, tolerance, emotion changes, and conflict, analyze the impact of this game due to the user and to find addiction factor, expectations you can achieve the following objectives:

- (A) The model is just the idea, hoping to make this idea to reality, and developed into the web page or APP, APP directly to the record of the use of mobile phone users to play games of the time, frequency, will be able to more accurately calculate the degree of addiction for users use at any time.
- (B) The model of addiction grading only do the follow-up hoping to be the type of personal information and subject classification, thrust reversers find addiction factor, that is able to find a way of its treatment.

Parameter Estimation of Trailing Suction Hopper Dredger Dredging Model by GA

Zhen Su and Wei Yuan

Abstract The trailing suction hopper dredger dredging model contains many parameters related to the soil types. And the parameters have big difference under different soil conditions. In this paper, genetic algorithms are used to estimate the parameters associated with soil type in a hopper dredger dredging model. The results were compared with the actual monitoring data, and the optimal parameter estimating values were obtained. The example showed that this approach of parameter estimation, based on genetic algorithms, is applicable.

1 Introduction: Genetic Algorithm

A genetic algorithm (GA) is a search and optimization tool, which works differently compared to classical search and optimization methods. Because of its broad applicability, ease of use, and global perspective, GA has been increasingly applied to various search and optimization problems in the recent past.

Over the last decade, a genetic algorithm (GA) has been extensively used as a search and optimization tools in various problem domains, including sciences, commerce, and engineering. The primary reason for its success is its broad applicability, ease of use, and global perspective. The concept of a genetic algorithm was first conceived by John Holland of the University of Michigan, Ann Arbor. Thereafter, he and his students have contributed much to the development of the field.

GA is a procedure used to find approximate solutions to search problems through application of the principles of evolutionary biology. Genetic algorithms use biologically inspired techniques such as genetic inheritance, natural selection,

Z. Su (✉) · W. Yuan

School of Electronics and Information, Jiangsu University of Science and Technology, The 2nd Mengxi Road, Zhenjiang 212003, China
e-mail: suzhen415@126.com

mutation, and sexual reproduction (recombination, or crossover). Along with genetic programming (GP), they are one of the main classes of genetic and evolutionary computation (GEC) methodologies.

2 Problem Formulation: Parameter Estimation of TSHD Dredging Model

Trailing suction hopper dredgers are primarily used in land reclamation projects. In such projects, the operators of the hopper dredger aim to achieve the highest production possible. In order to obtain the minimization of the integral dredging costs per m³ of sand or the maximization of the production per time unit, a control-oriented dynamic model of the hopper dredger has been developed and calibrated by using recorded process data.

Based on this model, a suitable control strategy can be derived, for instance, by using model-predictive control. Since the model contains many uncertain parameters related to soil types, genetic algorithms are used in dredging data from different vessels for estimation of parameters and model calibration.

A dynamic state-space model of the sedimentation process has been developed, based on [1–5].

$$\begin{cases} \dot{V}_t = Q_i - Q_o \\ \dot{m}_t = Q_i \rho_i - Q_o \rho_o \end{cases} \tag{1}$$

where V_t is the total volume of the mixture in the hopper, m_t is the total mass in the hopper, Q_i is the flow-rate of the incoming mixture, ρ_i is the incoming mixture density, Q_o is the overflow rate, ρ_o is the overflow density.

This flow rate is modeled as follows:

$$Q_o = k_o \max(h_t - h_o, 0)^{\frac{3}{2}} \tag{2}$$

where k_o is a parameter depending on the overflow weir shape and circumference, h_t is total height of mixture, h_o is the height of the overflow.

This flow density is modeled as follows:

$$\begin{cases} Q_w = A(1 - \mu)v_{so} \frac{\rho_m - \rho_w}{\rho_q - \rho_w} \left(\frac{\rho_q - \rho_m}{\rho_q - \rho_w}\right)^\beta \\ \mu = \min\left(\frac{Q_o^2}{(k_e h_m)^2}, 1\right) \\ Q_s = A(1 - \mu)v_{so} \frac{\rho_m - \rho_w}{\rho_s - \rho_w} \left(\frac{\rho_q - \rho_m}{\rho_q - \rho_w}\right)^\beta \\ Q_{ms} = \max(Q_o - Q_w, 0) \\ \rho_o = \frac{(\rho_m - \rho_w)Q_{ms}}{Q_{ms} + Q_w} + \rho_w \end{cases} \tag{3}$$

where the overflow rate (Q_o) is the sum of the water flow (Q_w) and the mixture soup flow (Q_{ms}), v_{so} is undisturbed settling velocity, ρ_m is Average density of mixture in hopper, ρ_w is water density, ρ_q is sand quartz density, β is exponent in the hopper settling velocity, k_e is erosion parameter, h_m is height of the mixture layer.

2.1 Soil-Type-Dependent Parameters

The hopper model has four parameters that depend on the type of soil and, therefore, are variable (see Table 1). The other parameters can be obtained off-line.

3 Proposed Solution

GAs are search and optimization procedures that are motivated by the principles of natural genetics and natural selection. Some fundamental ideas of genetics are borrowed and used artificially to construct search algorithms that are robust and require minimal problem information [6, 7].

According to the theory of genetic algorithms, the optimal solution for the function has found by matlab software based on genetic algorithms. Genetic algorithm parameters: population size is 100, the evolution of the number is 500, crossover probability is 0.6, mutation probability is 0.1. The average value and the best value of each generation function are shown in Fig. 1.

4 Results and Discussion

In this article, we use two groups of dredging data which were respectively collected from A area and B area. A and B area are have different soil particle size. And then we estimate 50 times for each group data. The final estimated parameters in the following Table 2:

Table 1 Summary of parameters that need to be calibrated with data

Parameter	Description
ρ_s	Sand-bed density in hopper
v_{so}	Undisturbed settling velocity
k_e	Erosion pickup flux coefficient
β	Exponent in settling equation based on particle Reynolds number

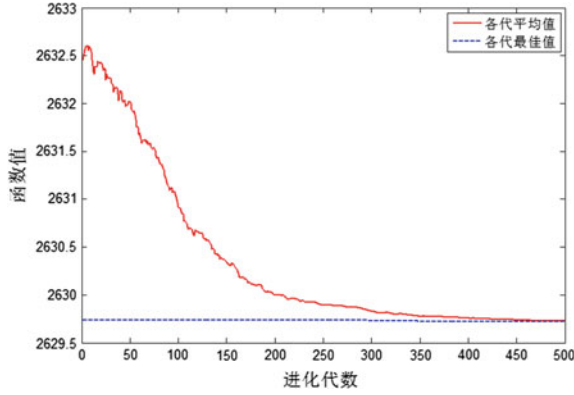


Fig. 1 Optimize process of genetic algorithm

Table 2 Results of different soil parameters

Soil parameters	ρ_s	v_{so}	k_e	\sqrt{J}	VAF
A	1984	7.3	1.4	582.3	0.90
B	1780	6.5	1.3	450.6	0.98

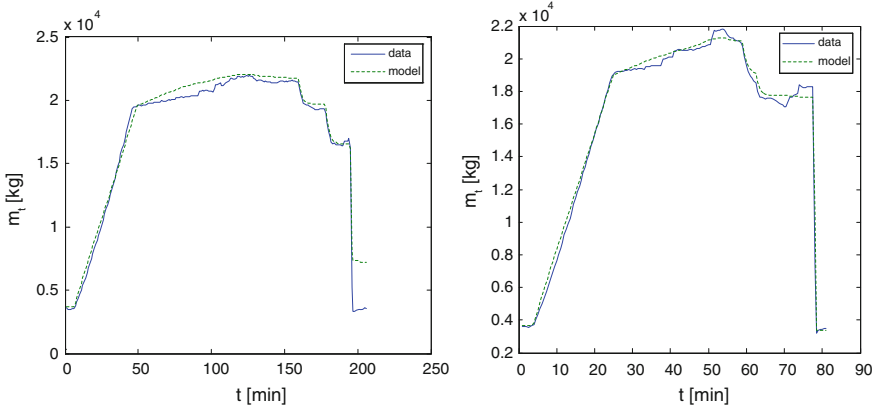


Fig. 2 Validation results of total mass m_t in hopper: *left* A data; *right* B data

The next step is that the parameters of estimation are used for dredging model, In order to estimate m_t . Figure 2 shows the simulation results of the model compared with data.

These results show that in different conditions, soil parameter estimated are applied to the hopper model, and the output value and the measured data showed good fit, indicating that the soil parameter estimates genetic algorithms has given high accuracy.

5 Conclusion

In this paper, we made some attempts in terms of parameter estimation based on Genetic Algorithms. The results proved that the method is effective and useable. In our future research, genetic algorithms will be used to improve parameter estimation in complex mechanistic models of the hopper sedimentation process. It will also be integrated in a decision support tool for the future use on board of the hopper dredger.

References

1. Braaksmā J (2008) Model-based control of hopper dredgers. Delft University of Technology, Netherlands
2. Yagi T (1970) Sedimentation effects of soil in hopper. In: Proceedings of WODCON world dredging conference
3. Ooijens S (1992) Adding dynamics to the camp model for the calculation of overflow losses. *Terra et Aqua*, pp 12–21
4. van Rhee C (2002) On the sedimentation process in a suction hopper dredger, Ph.D. dissertation, TU Delft
5. Ooijens S, de Gruijter A, Nieuwenhuijzen A, Vandycke S (2001) Research on hopper settlement using large-scale modelling. In: Proceedings of the CEDA dredging days 2001, pp 1–11
6. Todd DS, Sen PA (2007) Multiple criteria genetic algorithm for containership loading. In: Back T (ed) Proceedings of the seventh international conference on genetic algorithms. Michigan State University, Morgan Kaufmann Publishers, pp 674–681
7. Grefenstette JJ (1986) Optimization of control parameters for genetic algorithms. *IEEE Trans Syst Man Cybern*, pp 122–128

CPP Control System Design of Ship Based on Siemens PLC

Liang Qi and Shengjian Huang

Abstract Control system of controllable pitch propeller (CPP) is the key device of ship propulsion system. Along with better requests of the response speed and handling quality in ship propulsion system, the control system has been designed based on Siemens PLC and PROFIBUS, in order to enhance the performance of real-time, maneuverability, stability and reliability.

Keywords Ship · Controllable pitch propeller (CPP) · Siemens PLC · PROFIBUS

1 Introduction

Controllable pitch propeller (CPP) is an important part of ship's propulsion system. When ship is sailing with any load, CPP can adjust pitch to realize better propulsion efficiency and lower fuel consumption under the condition of constant rotating speed and direction of main engine. So the control system of CPP can effectively resolve the contradiction between multiple working conditions of ship and single main engine, improve ship propulsion efficiency under the complex voyage working condition [1].

Along with development of shipbuilding industry and automation, the propulsion device with CPP have a wider application whose performance straightly affects control level of overall ship. Then programmable logic controller (PLC) of Siemens

L. Qi (✉) · S. Huang
School of Electronics and Information, Jiangsu University of Science and Technology,
Zhenjiang, China
e-mail: alfred_02030210@163.com

S. Huang
e-mail: huang1989just@163.com

is applied to realize control system of CPP which can enhance ship's performance of real-time, maneuverability, stability and reliability, according to wider and better application of PLC in industrial control field.

2 The Composition of CPP

CPP control system is composed of console in wheelhouse, console in command control room, console in centralized control room, console near the steering gear, master control cabinet (including master PLC), CPP electronics servo device and speed adjusting system of main engine, which is shown as Fig. 1.

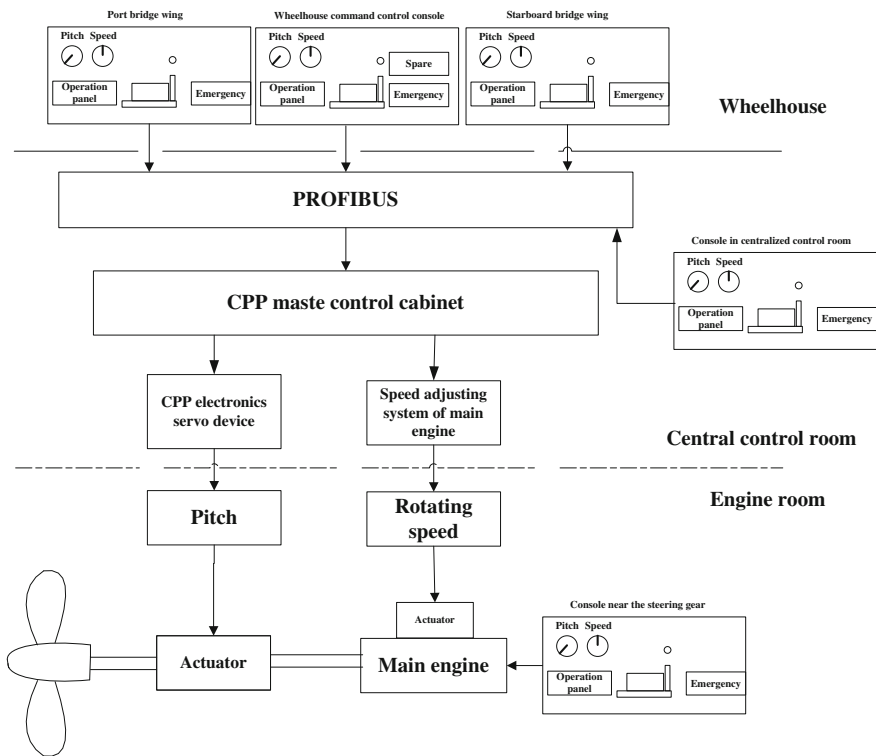


Fig. 1 The block diagram system

3 The Hardware Design of CPP

The hardware of CPP system applies PLC of Siemens as master controller. PLC receives many signals including switching value, rotating speed of main engine and feedback value of pitch from console in wheelhouse and console in command control room. These signals are calculated or processed by PLC, whose results are delivered to three devices shown as below:

- (1) operation panels of console in wheelhouse and console in command control room;
- (2) electronics speed controller near the steering gear;
- (3) electro-hydraulic proportional valve in hydraulic engine device.

The PLC configuration is shown as Tables 1, 2, 3 and 4.

The hardware structure is shown as Fig. 2.

Table 1 Wheelhouse PLC(S7-200)configuration

Module	Name	Configuration	Quantity
CPU 224	CPU module	14 input/10 output	1
EM 223	Digital extension module	16 input/16 output	1
EM 231	Analog extension module	4 input	1

Table 2 Command control room PLC(S7-200)configuration

Module	Name	Configuration	Quantity
CPU 224	CPU module	14 input/10 output	1
EM 223	Digital extension module	16 input/16 output	1
EM 231	Analog extension module	4 input	1

Table 3 Near the steering gear PLC(S7-200)configuration

Module	Name	Configuration	Quantity
CPU 224	CPU module	14 input/10 output	1
EM 223	Digital extension module	16 input/16 output	1
EM 231	Analog extension module	4 input	1
EM 232	Analog extension module	2 output	1

Table 4 Master PLC(S7-300)configuration

Module	Name	Configuration	Quantity
PS 307 5A	Power module	5A, 24VDC	1
CPU 314	CPU module	\	1
CP 342-5	Serial communication module	\	1
AI4/AO2×8/8Bit	Digital input/output module	4 input, 2 output	1
DI32×DC24V	Digital input module	32 inout, 24VDC	1
DO8×DC24V/0.5A	Digital output module	8 output, 24VDC	1

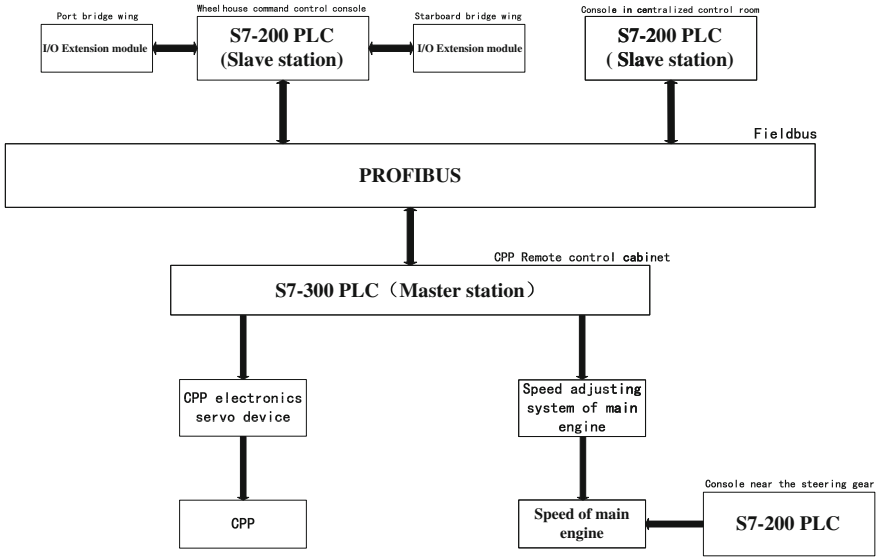


Fig. 2 Hardware structure

4 The Software Design of CPP

4.1 The Overall Design

Software design of CPP control system mainly included design of system main program, rotating speed of main engine, pitch control, load control and other performance, which is shown as Fig. 3.

4.2 Pitch Control Algorithm Design

Voyage course of ships has time-variable, non-linear and big time-lag properties. Common PID control hasn't frequently obtained desired control effects. In order to improve control quality and avoid frequent actions of hydraulic pitch controller, PID control algorithm having dead-zone is applied in pitch control system.

The equation of PID algorithm having dead-zone is depicted as below [2]:

$$u(k) = K_p e(k) + K_I \sum_{j=0}^k e(j) + K_D [e(k) - e(k - 1)] \quad (1)$$

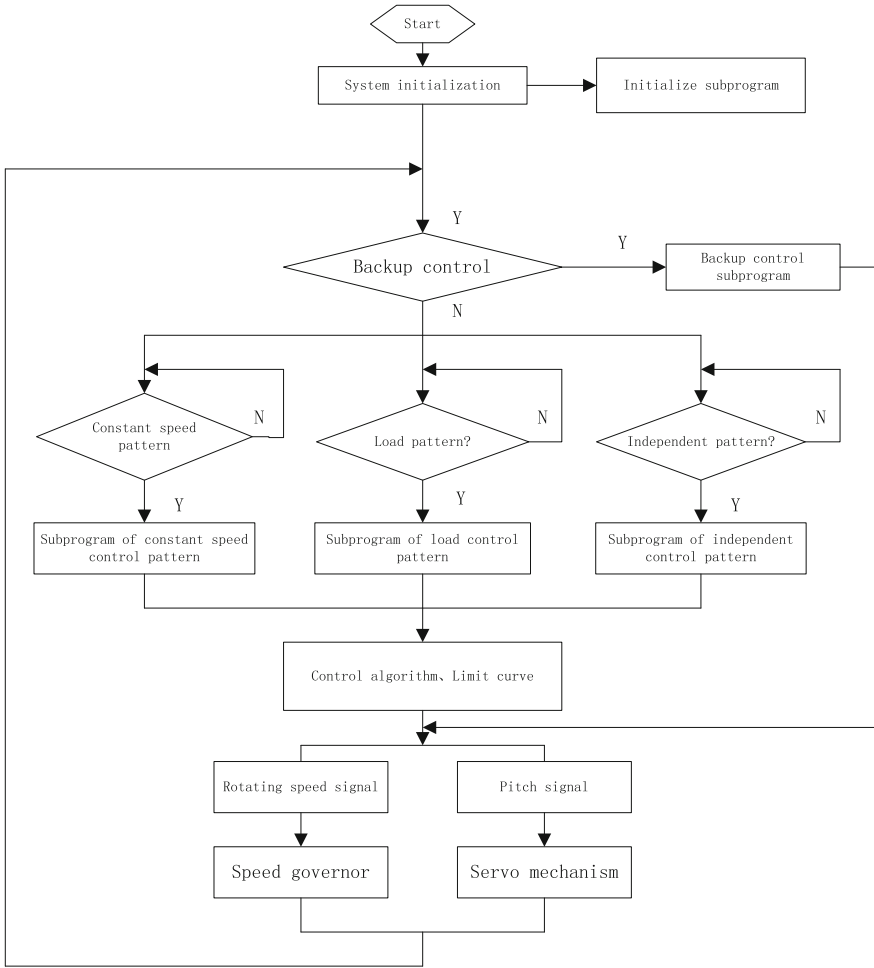


Fig. 3 The overall control system flowchart

$$e(k) = \begin{cases} 0 & |e(k)| \leq |e_0| \\ e(k) & |e(k)| > |e_0| \end{cases} \quad (2)$$

where k is sampling number; K_p is proportional coefficient; K_I is integral coefficient; K_D is differential coefficient; $u(k)$ is output of controller with k sampling time; $e(k)$ is input error with k sampling time; $e(k - 1)$ is input error with $k - 1$ sampling time; e_0 is width value of dead-zone.

4.3 Realization of Control Software of STEP7

Hardware configuration should be done at first before setting up a program when control software of STEP7 designs automation system.

4.3.1 Hardware Configuration

Hardware configuration is defined as hardware allocation and parameters distribution of SIMATIC work station which is designed by STEP7. The control system of CPP in the paper is composed of a set of PLC (S7-300) and three sets of PLC (S7-200). CPP system control cabinet is controlled by the one set of PLC (S7-300) as one mater station and consoles in wheelhouse, centralized control room and near the steering gear are controlled by the three sets of PLC (S7-200) respectively as two slave stations. The hardware configuration is accomplished in PLC (S7-300) which is shown as Fig. 4.

PLC (S7-300) is communicated with PLC (S7-200) by means of PROFIBUS FieldBus which is done by EM277 (communication module). Thus four sets of PLC constitute a network of FieldBus which is shown as Fig. 5.

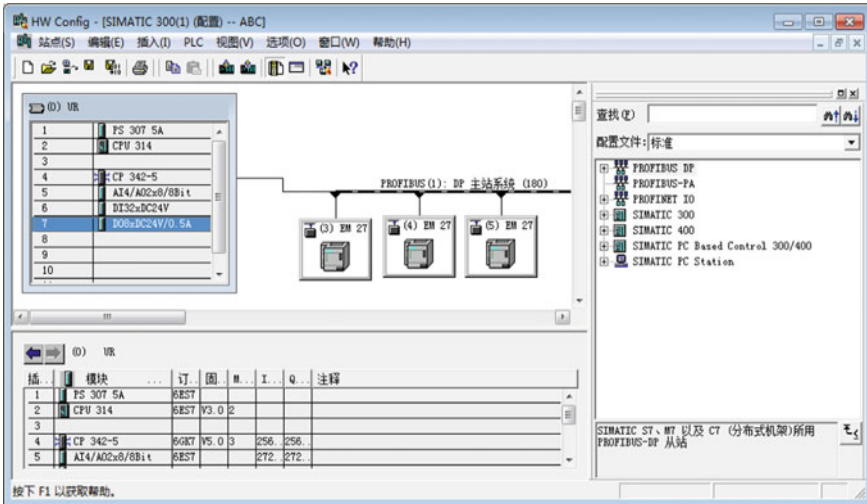


Fig. 4 Hardware configuration

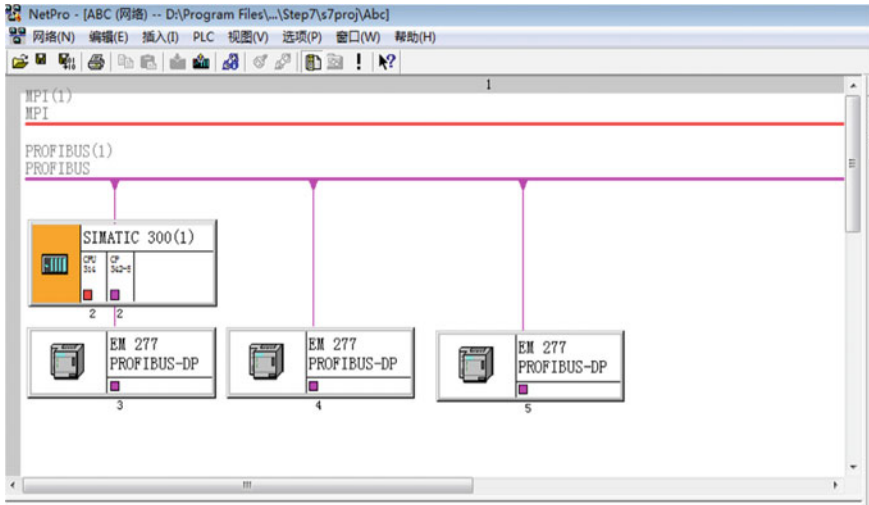


Fig. 5 Network configuration

4.3.2 Modularization Programming

(1) Divisions of program modules

According to control system requests and performances of every operating device, the overall program is divided into five function modules which is shown as Fig. 6, where OB1 is main program; FB1 is subprogram of load control pattern; FB2 is subprogram of constant speed control pattern; FB3 is subprogram of backup control pattern; FC1 is subprogram of speed control of main engine; FC2 is subprogram of pitch control; DB1-3 are universal global data blocks.

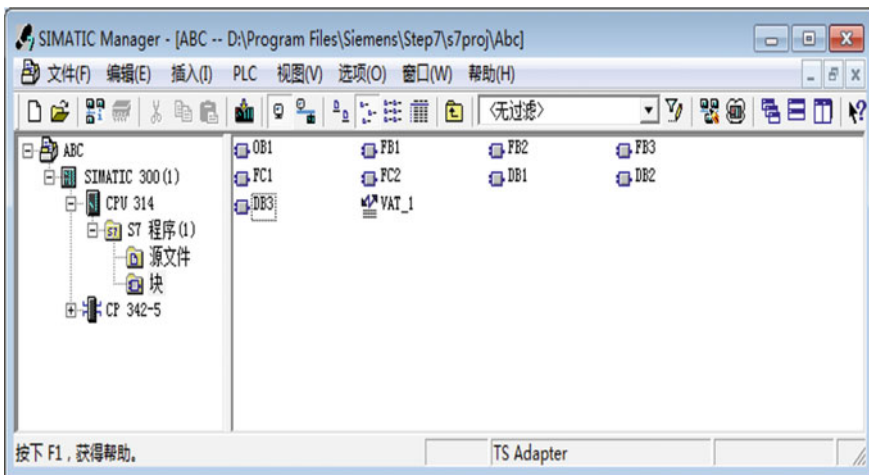



Fig. 6 Module partition

	状态	符号 /	地址		数据类型	注释
1		0 Pitch	I	8.4	BOOL	
2		0 Pitch /0 Load request	I	9.0	BOOL	
3		0 Pitch switch signal	I	8.7	BOOL	
4		1# Pump running	I	9.7	BOOL	
5		2# Pump running	I	10.0	BOOL	
6		PTO combination	Q	12.5	BOOL	
7		Local /Remote	I	8.1	BOOL	
8		Load /Pitch deceleration	I	9.1	BOOL	
9		Load decrease reset	I	9.3	BOOL	
10		Pitch control signal	PQW	290	WORD	
11		Propeller pitch feedback	PIW	294	WORD	
12		1# Pump start-up	Q	12.6	BOOL	
13		2# Pump start-up	Q	12.7	BOOL	
14		Speed control signal	PQW	288	WORD	
15		Emergency stop	I	9.5	BOOL	
16		Boost pressure	PIW	290	WORD	
17		Main engine overload	I	8.2	BOOL	
18		Main engine start-up	Q	12.0	BOOL	
19		Main engine stop	Q	12.1	BOOL	
20		Main engine running	I	8.3	BOOL	
21						

Fig. 7 Design of symbol table

(2) Design of symbol table

Symbol table is designed by inputting after clicking the icon of  Symbols. Symbol table after being designed is shown as Fig. 7.

(3) Pitch control-PID control ladder diagram with dead-zone.

The pins are defined as follows:

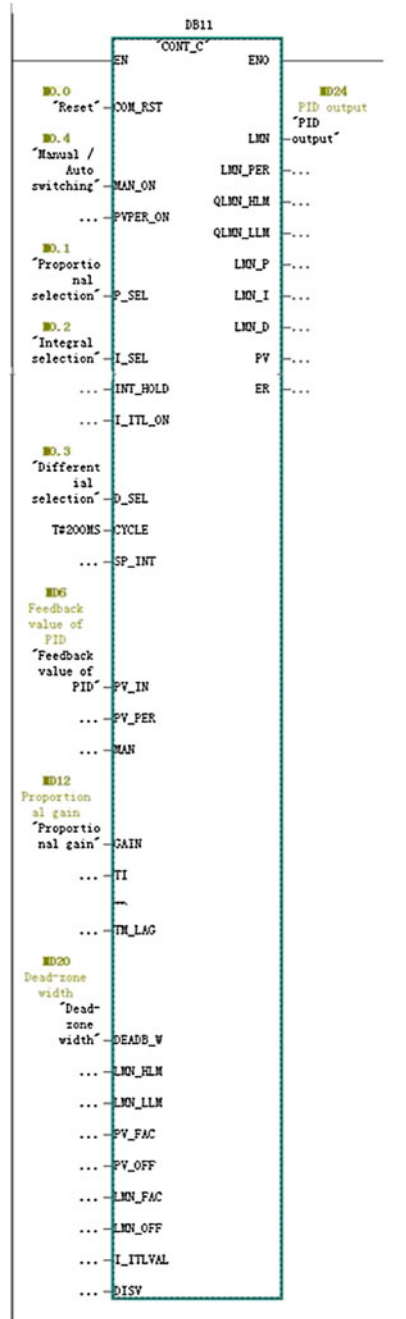
A: Partial input parameters:

- COM_RST: Restart PID;
- MAN_ON: Manual/Auto switching bit;
- PEPER_ON: Process variable;
- P_SEL: Proportional selection bit;
- I_SEL: Integral selection bit;
- D_SEL: Differential selection bit;
- CYCLE: PID Sampling period, Generally set to 200MS;
- SP_INT; PID Values;
- PV_IN: The feedback value of PID (Also known as process variable);
- PV_PER: The feedback value without normalized, By PEPER-ON selection effective (Not recommend);
- GAIN: Proportional gain;
- TI: Integral time;
- TD: Differential time;
- DEADB_W: Dead-zone width;

B: Partial output parameters:

LMN: REAL: PID output (Fig. 8).

Fig. 8 PID control ladder diagram with dead-zone



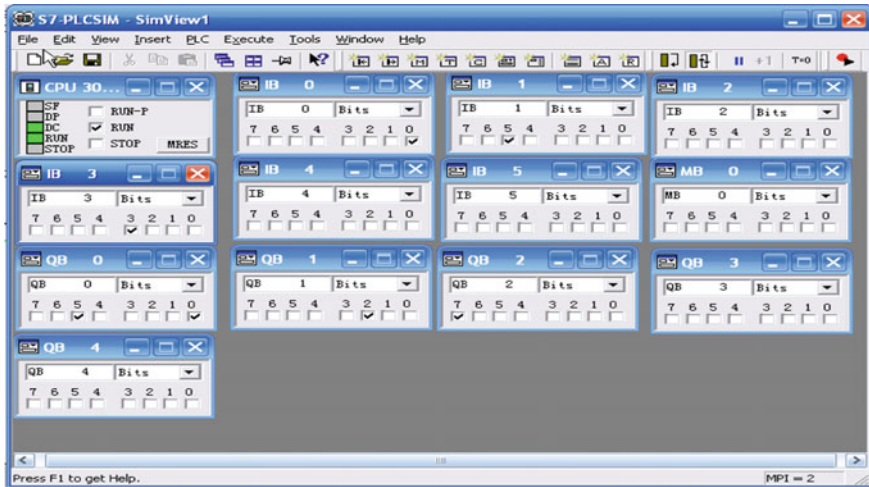


Fig. 9 Debugging interface

4.3.3 Debugging of Simulator

Simulator of S7-PLCSIM is operated. Overall program designed by STEP7 is downloaded to PLC in order to verify the accuracy of overall program. The debugging interface of overall program is shown as Fig. 9.

5 Conclusion

A control system of CPP which based on Siemens PLC and PROFIBUS is designed in this paper. Then the design of hardware structure and the software framework are illustrated in detail. The selected hardware is practical and reliable. The design of algorithm structure is simple. So this scheme can be a very good application in the actual control of the CPP.

References

1. Wang YQ (2004) Simulation of CPP propulsion system and designing of teaching software. Dalian Maritime University, Dalian
2. Liu JK (2004) Advanced PID control and its MATLAB simulation. Electronics Industry Press, Beijing

The Surface Deformation Prediction of Ship-Hull Plate for Line Heating

Liang Qi, Feng Yu, Junjie Song and Xian Zhao

Abstract Line Heating (LH) is the main method for forming ship-hull plates. And it's mainly operated by skilled workers manually, so the accuracy of final shape and the productivity solely depend on the experience of the workers. In order to predict the surface deformation of plates, a new method is developed to determine the processing parameters and improve the productivity. Firstly, LH process is simulated by Finite Element Analysis (FEA) according to the complexity of LH mechanism. Secondly, a model of Artificial Neural Network (ANN) is established. Finally, the computation results of simulation by FEA are applied to train the ANN model. In the way, a method of surface deformation prediction is proposed for real time analysis.

Keywords Ship-hull plate · Line heating (LH) · Finite element analysis (FEA) · Surface deformation · Artificial neural network (ANN)

1 Introduction

The hull surface is mainly composed of three-dimensional, complex and undevelopable curved plates. The technique of Line Heating (LH) is a popular, efficient and economic process of forming curved hull surface shapes by means of mouldless

L. Qi (✉) · X. Zhao

School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, China
e-mail: alfred_02030210@163.com

X. Zhao

e-mail: zhao37082@163.com

F. Yu · J. Song

Jiangnan Shipyard (Group) Co., LTD, Shanghai, China
e-mail: yufeng200807@sina.com

J. Song

e-mail: junjie-song@163.com

forming, which is extensively applied in shipbuilding industry all over the world. Traditionally, LH process is mainly used to form large and double-curved hull plates in most shipyards, and depends on the abundant experiences of skilled workers who typically accomplish the task manually. Thus, the quality and efficiency of forming curved plates relies on the skills and experiences of workers. However, these manual and experiential techniques of producing patterns restrict the cycle and quality of shipbuilding and have become a bottleneck in the modern production line, as shipbuilding's development and mode change. Over the past decades, increasing productivity of the line-heating process is one of the goals in the shipbuilding industry. As an important content, many scholars and technicians have carried on fruitful research on predicting surface deformation rapidly and efficiently.

As unceasing development and improvement of artificial intelligence, especially Artificial Neural Network (ANN), the efficient approach can be presented to describe the complicated relationship between many processing parameters and residual deformation precisely. ANN has the superior property of solving high-order nonlinear problems and ability of self-learning, whose computing speed is fairly higher than Finite Element Analysis (FEA). So it is feasible for ANN model to describe high-order coupled nonlinear relationship. Besides, the performance of real-time and precision of ANN model can meet the request of practical producing.

2 Brainstorming

It is the central content of LH process that the mathematics model between processing parameters and residual deformation is established based on obtaining abundant data of LH process. Firstly LH process is numerically simulated by FEA. Secondly, the relationship between LH processing parameters and residual deformation is modeled by ANN, and the ANN model is trained by FEA results. Finally, the ANN model can be applied to predict surface deformation after efficient training repeatedly, and the method is highly efficient and precise. During the course of ANN modeling, the paper sets two parameters describing surface importing energy of heat source as ANN's input variables, and residual deformation as output variables. The relationship between input and output variables can satisfy deformation of LH process under the condition of series thickness, varies of heat flux, length of heating line and moving velocity of oxyacetylene torch. ANN model is fairly established concisely and practically which embodies the combined effect of above processing parameters to residual deformation. The flow of surface deformation prediction based on FEA and ANN is shown in Fig. 1.

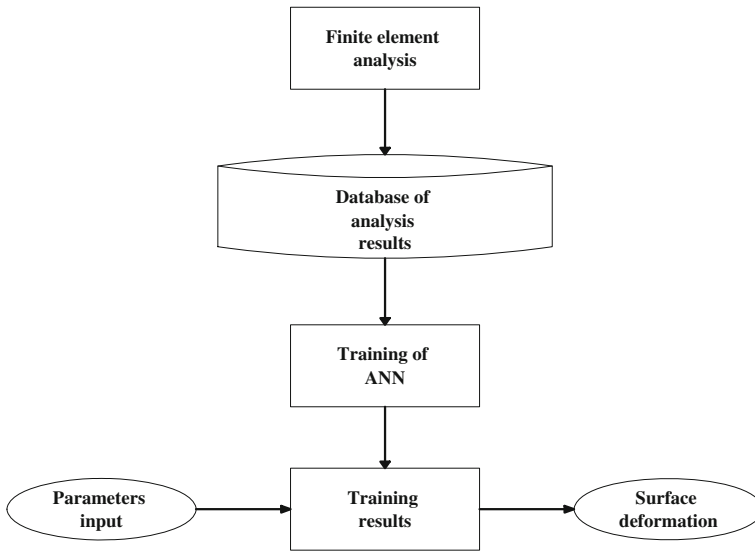


Fig. 1 Flow of surface deformation prediction

3 FEA Simulation of LH Process

As the heat source moves along the heating line designed in advance on a steel plate, the metal of heated zones expands rapidly, then shrinks confined by unheated zones and water cooling compulsively. The effects of heat expansion and cold contraction in the plate result in residual deformation of the plate because of much thermal strain. The three-dimensional transient thermal- elastic-plastic model for FEA computation was established by the way of mode of sequential coupling. The moving heat source (in Rg1 region) is divided into a plurality of step loads which are applied on plate, the results of thermal analysis are taken as initial load in mechanical analysis. Then mechanical analysis includes thermal elastic-plastic modeling under bilinear yield criterion of Von Mises, also considering the regards of material properties with time.

3.1 Heated Zones Division and Measuring Points Distribution

Suppose the size of a steel plate expressed by $L \times W \times h$. The heating line, the domain region Rg1 which is close to the heating line and Rg2 which is away from

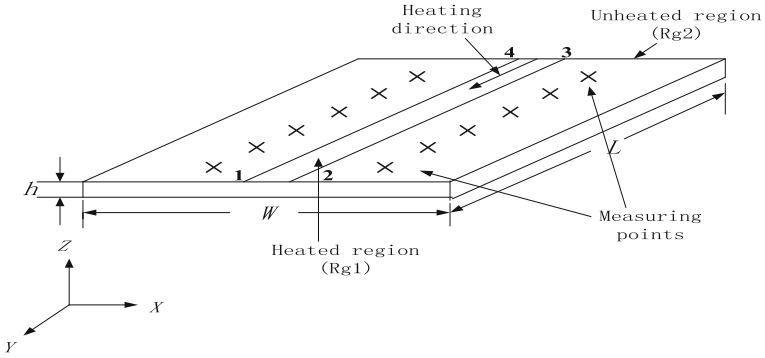


Fig. 2 Various regions of plate and distribution of measuring points

heating line are shown in Fig. 1. The region which is bounded by the points 1–2–3–4 is considered as Rg1. The other regions are considered as Rg2. Six points which is away from heating line at the distance of 50 mm are used to measure the residual deformation (Fig. 2).

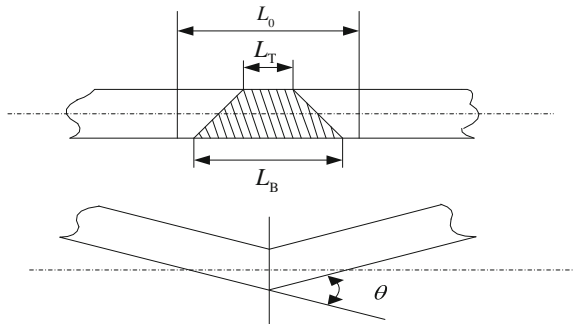
3.2 Description of Steel Plates Surface Deformation

During the course of LH process, the shrinkage may be produced perpendicular to moving direction of heat source, which is called transverse shrinkage, namely line deformation. The steel plate may be bended for different line deformation of top and bottom surface, namely angular deformation. Commonly the description method adopts local line deformation and angular deformation, which is shown in Fig. 3.

Computing method of line deformation and angular deformation is shown as below:

$$\Delta_T = L_0 - L_T, \quad \Delta_B = L_0 - L_B \tag{1}$$

Fig. 3 Definition of line deformation and angular deformation



$$\Delta L = \frac{\Delta_T + \Delta_B}{2} \tag{2}$$

$$\tan(\theta) = \frac{L_B - L_T}{h} \tag{3}$$

where Δ_T is line deformation on top surface; Δ_B is line deformation on bottom surface; ΔL is line deformation and θ is angular deformation.

3.3 FEA Numerical Simulation of LH Process

1. The dimension of test steel plates:

$$L = 1000 \text{ mm}, \quad W = 1000 \text{ mm}, \quad h = 14 \text{ mm}; \tag{4}$$

2. Material property:

The test plate in shipbuilding is low-carbon steel. Some material properties with temperature are shown in Table 1.

3. The producing parameters in LH process:

The range of the parameters are shown as below confined under the condition of maximum surface temperature (850 °C) and hardware of motion organization in numerical device.

$$\text{Acetylene flow } Q_{C_2H_2} = (800 \sim 2500) \text{ l/h}; \tag{5}$$

$$\text{Moving velocity } v_{HL} = (0 \sim 8) \text{ mm/s}; \tag{6}$$

$$\text{Length of heating line of three specification } L_{HL} = (200, 500, 1000) \text{ mm} \tag{7}$$

Table 1 Temperature dependent material properties of mild steel

Temperature °C	Modulus of elasticity GPa	Poisson ratio	Coefficient of thermal expansion $10^{-6}/^{\circ}\text{C}$	Coefficient of heat conduction $\text{W}(\text{mK})^{-1}$	Specific heat $\text{J}(\text{kg K})^{-1}$
0	200	0.2786	10	51.9	450
100	200	0.3095	11	51.1	499.2
300	200	0.331	12	46.1	565.5
450	150	0.338	13	41.05	630.5
550	110	0.3575	14	37.5	705.5
600	88	0.3738	14	35.6	773.3
720	20	0.3738	14	30.64	1080.4
800	20	0.4238	14	26	931
1450	2	0.4738	15	29.45	437.93

4. Meshing:

The temperature of the node in top surface decreases rapidly as the distance between the node and the heating line increases. The meshing is not symmetrical: the size of units in heated region is small; the other regions are big size. The FEA model consists of heated region, unheated region and transitional region. It is divided to 3 ~ 5 units along the direction of thickness. Considering the units of heated region and unheated region with regular shape, units of heated region are close and temperature gradient of unheated region is small, so SOLID70 which is the hexahedron unit with 8 nodes is chosen for meshing these regions. SOLID90 which is the tetrahedron unit with 20 nodes is chosen for meshing transitional region, this type of unit is appropriate for simulate curved edge, and it supports the function of grid transition. The SOLID95 is the equivalent of SOLID90 for carrying on the mechanical analysis, but the distribution of nodes is unchanged. For the analysis of temperature field, the heated region is covered with SURF152 unit in order to accomplish the loads of heat flux and convection on the same surface for one kind of unit in simulation. These surface effect units are removed while mechanical analysis carries on. The meshing results are shown in Fig. 4.

5. Results of FEA simulation:

Temperature distribution. In LH process, the temperature field changes as the heat source moving on the plate. As the heating source moves in a stable velocity, it'll reach a steady-state temperature field relatively around the torch soon. Figure 5 shows the isothermal map of the numerically simulated temperature distribution on the top face of the steel plate 327 s after commencement of the experiment.

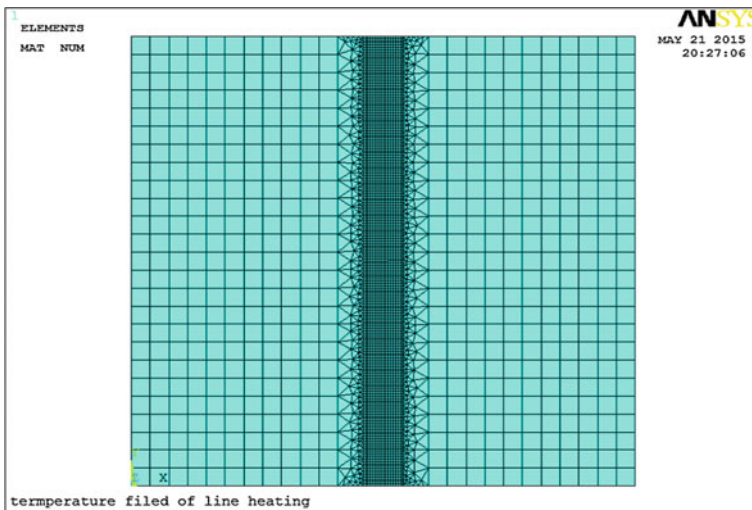


Fig. 4 Finite element mesh of a plate for LH

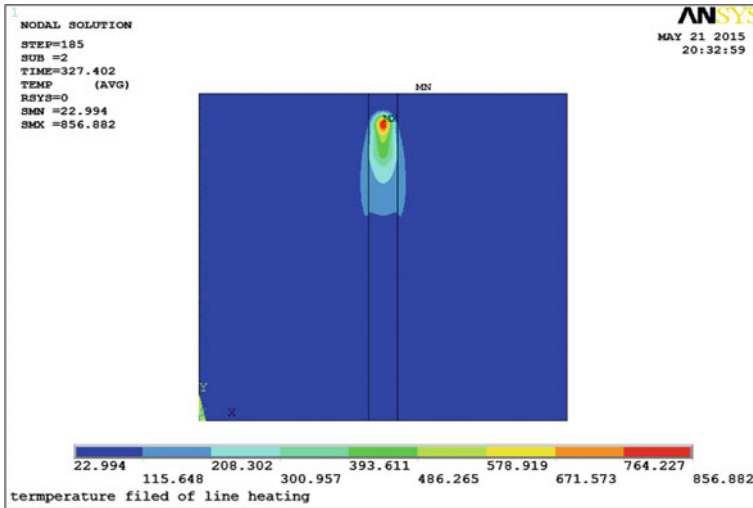


Fig. 5 Temperature field distribution

Residual stress distribution. The critical temperature region can be considered as the plastic strain generation region because although plastic strains might be created below the critical temperature, the bulk of plastic strains are produced above this critical temperature. Figures 6, 7 and 8 depicts the distribution of residual stress in the (X, Y, Z)-direction on the top surface as measured by using the impact-indentation method respectively and the simulated results using the finite element method.

Distribution of residual deformation (Table 2).

4 Modeling of ANN

4.1 ANN Model

The above model is identified by v-SVM [1] with fast convergence speed and hybrid kernel function [2]. Sample set given is $\{(x_i, y_i), i = 1, 2, \dots, l\}$, where $x_i \in R^N$, $y_i \in R$ and l are input, corresponding objective value, number of samples, respectively. The requested form of fitting function is:

$$f(x) = w \cdot \phi(x) + b \quad w, \phi(x) \in R^N, b \in R \tag{8}$$

where w and $\phi(\cdot)$ are parameter and function array vector, by which input samples space can be mapped into characteristic space.

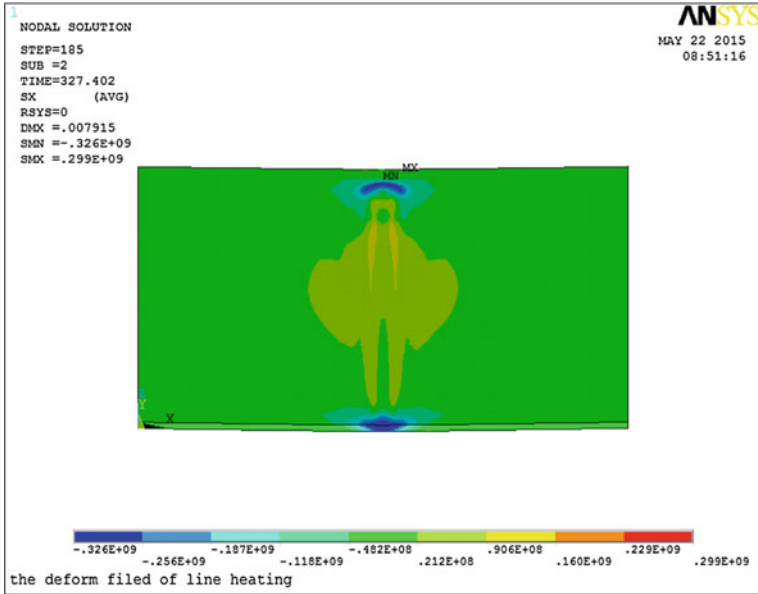


Fig. 6 Residual stress in the X-direction

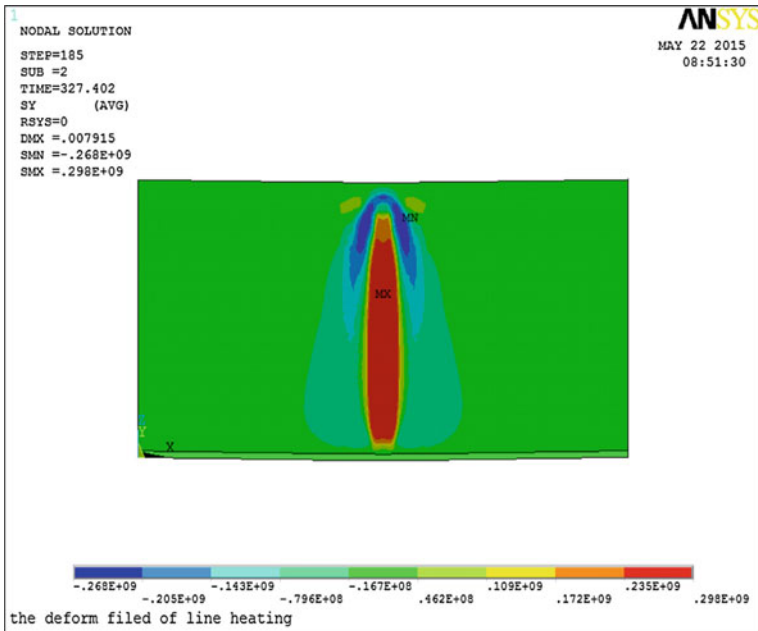


Fig. 7 Residual stress in the Y-direction

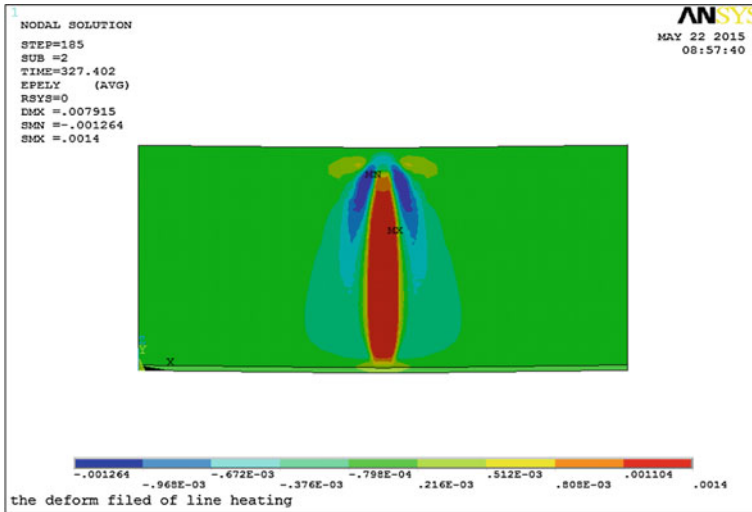


Fig. 8 Residual stress in the Z-direction

Table 2 Provides results of line deformation and angular deformation

h (mm)	$Q_{C_2H_2}$ (l/h)	v_{HL} (mm/s)	η	r_0 (mm)	L_{HL} (mm)	ΔL (mm)	θ ($^\circ$)
14	1356	2.81	0.28	40	1000	0.24026984	1.82604840
					500	0.23554649	1.22207741
14	1370	2.87	0.28	40	1000	0.23279370	1.81519077
					500	0.23033942	1.22252900
14	1375	2.89	0.28	41	1000	0.23308319	1.82397869
					500	0.22883284	1.22311833
14	1582	3.52	0.28	41	1000	0.19451438	1.80711968
					500	0.19626141	1.26390446
14	1596	3.57	0.28	41	1000	0.19616579	1.83460386
					500	0.19571617	1.25603674
14	1710	3.82	0.27	41	1000	0.17828399	1.75802019
					500	0.18356012	1.24509492
14	1722	3.86	0.27	41	1000	0.17482037	1.73854632
					500	0.18263492	1.24670716
14	1736	3.92	0.27	41	1000	0.17607474	1.76319931
					500	0.18030090	1.24538809
14	1748	3.97	0.27	41	1000	0.17175466	1.73921062
					500	0.17852781	1.24476924
14	1806	4.21	0.27	41	1000	0.16359669	1.72082320
					500	0.16869797	1.25736690

Representative Polynomial kernel function (globe kernel function) and RBF kernel function (local kernel function) constitute a kind of hybrid kernel function. CPP control system is identified by v-SVM with the hybrid kernel function, which has good model fitting precision and effectively restrains predicting output fluctuation aroused by local kernel function. The kernel function is shown as follow:

$$\text{Polynomial : } K(x, x_i) = [(xx_i) + 1]^q; \text{RBF : } K(x, x_i) = \exp\left[-\frac{|x - x_i|^2}{2\sigma^2}\right];$$

$$\text{Hybrid kernel function : } K_{mix} = \rho K_{poly} + (1 - \rho)K_{RBF} \tag{9}$$

where K_{poly} , K_{RBF} and $\rho(0 \leq \rho \leq 1)$ are Polynomial function, RBF function, and constant adjusting effect of this two functions.

Finally, the fitting output function is:

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i)K(x, x_i) + b \tag{10}$$

where α, α^* are **Lagrange** factors.

4.2 The Determination of Input and Output Parameters in ANN Model

Input parameters. Many factors may influence the deformation of steel plates in LH process, mainly including material and shape of plates, processing parameters and cooling. According to the article [3], main factors affecting deformation of plates were integrated to two new parameters (q_s, q_{st}). It's aiming to indicate the effect of surface input. The variable q_s is defined as the valid heating energy input per unit area, and it indicates the macroscopic effects of surface input. The variable q_{st} is defined as the valid heat energy input per unit area and per unit time, and it indicates the instantaneous effects of surface heating. Mathematical models of new variables are shown as below:

$$q_s = \frac{\text{effective energy (J)}}{\text{heating area (mm}^2)} = \frac{\eta \cdot Q_{C_2H_2} \cdot \frac{L_{HL}}{v_{HL}} q_{\text{calorific value}}}{L_{HL} \cdot 2r_0} = \frac{\eta \cdot Q_{C_2H_2} \cdot q_{\text{calorific value}}}{v_{HL} \cdot 2r_0} \tag{11}$$

$$q_{st} = \frac{\text{effective energy (J)}}{\text{heating area (mm}^2) \cdot \text{time (s)}} = \frac{\eta \cdot Q_{C_2H_2} \cdot q_{\text{calorific value}}}{L_{HL} \cdot 2r_0} \tag{12}$$

Residual deformation of steel plate has relation with as $Q_{C_2H_2}, v_{HL}, \eta, r_0, L_{HL}$ and so. q_s and q_{st} can reflect the comprehensive effect of the heating parameters,

Fig. 9 Input and output parameters of SVM

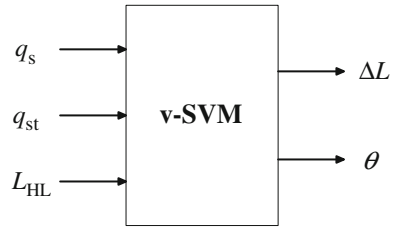


Table 3 Results of training SVM

				Simulation results by ANSYS		Training results of LS-SVM		Error	
h	L_{HL}	q_s	q_{st}	ΔL	θ	ΔL	θ	ΔL	θ
(mm)	(mm)	(J/mm ²)	(J/mm ² s)	(mm)	(°)	(mm)	(°)	(%)	(%)
14	200	24.86	0.36	0.14	0.68	2.39	2.01	3.02	2.43
14	200	21.93	0.42	0.11	0.70	3.97	2.66	1.00	4.72
14	300	24.78	0.24	0.18	0.91	4.20	3.12	0.24	1.27
14	300	21.57	0.29	0.14	0.93	2.07	1.47	0.48	2.80
14	500	22.84	0.16	0.20	1.25	2.63	1.80	0.75	0.55
14	500	21.01	0.18	0.17	1.26	4.31	2.45	1.60	6.06

different heating parameters resulting in different residual deformation can be comparable.

Output parameters. Output parameters are confirmed as ΔL and θ which are mentioned above. So the input and output variables is determined shown as Fig. 9.

4.3 Training Results of SVM

Table 3 provides partial results of SVM.

5 Prediction Experiments by SVM

In order to verify the effectiveness of SVM for surface deformation prediction, the experiments were tried by the numerical devices, shown as Fig. 10. Figure 10a shows the executing mechanism actuator whose control system applies PLC of Siemens and servo motor. Figure 10b shows the measuring device for surface deformation which realizes non-contacted, rapid and high-accuracy measurement of three-dimensional curve steel plates by the means of computer visual technology.

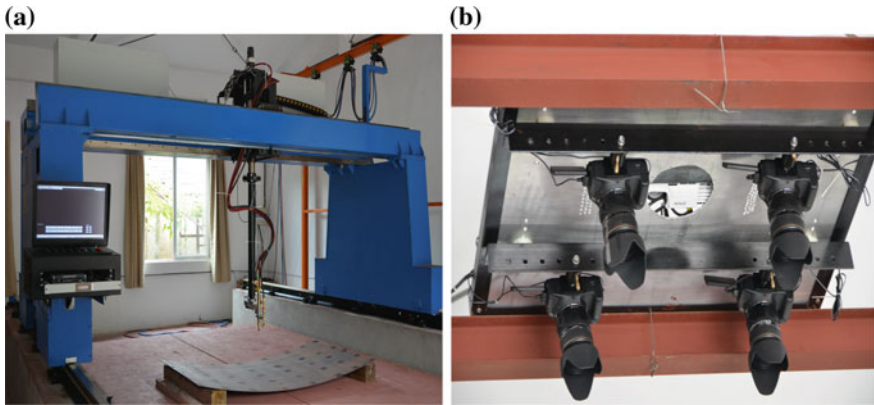
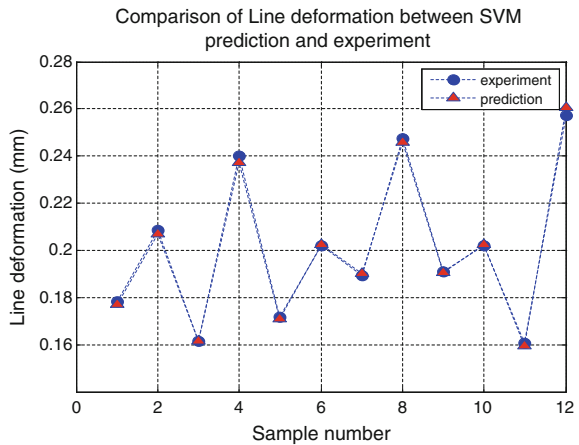


Fig. 10 a Experiment Device. b Measuring device

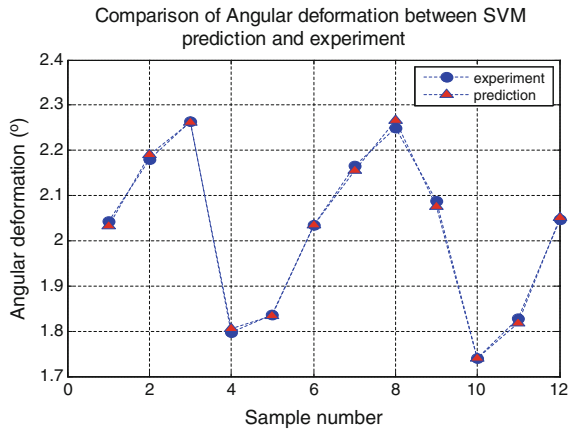
Fig. 11 Comparison of line deformation between SVM prediction and experiment



The simulation of prediction model of this paper is firstly normalized to the sample data, and the ν -SVM model is established, Let control factor $C = 125$, parameter of Polynomial kernel function $q = 3$, parameter of RBF kernel function $\sigma = 325$, adjusting factor of kernel function $\rho = 0.86$.

As can be seen from the Figs. 11 and 12, ν -SVM prediction model errors compared to experiment results are within the range of allowable error. And it is proved that the model has better performance to predict the effect and can be applied in real production.

Fig. 12 Comparison of angular deformation between SVM prediction and experiment



6 Conclusion

To predict the plate deformation precisely and rapidly, a new method which is based on FEA and ANN is developed to determine the processing parameters and improve the productivity. Prediction model based on v-SVM applies two coupling parameters (q_s, q_{st}) which can depict integrated effects of multiple factors which may affect surface deformation of ship-hull plates. Prediction results and experiments results show that the method can be applied in real production guidance of LH process better.

Acknowledgements The authors would like to acknowledge the support of Industry-Academia Cooperation Innovation Fund Projects of Jiangsu Province (BY2011143), the support of the Special Natural Science Foundation for Innovative Group of Jiangsu University during the course of this work.

References

1. Zhang HR (2003) SVM based on nonlinear system identification. *J Syst Simul* 15(1):119–121
2. Smits GF, Jordan EM (2002) Improved SVM regression using mixtures of kernels. In: *Proceedings of the 2002 international joint conference on neural networks*, vol 3(3). Honolulu, USA, pp 2785–2790
3. Qi L, Zhang CL (2013) Effect of forming factors on surface temperature and residual deformation of the plate in line heating. *Int J Mater Struct Integrity* 7(1):171–181

The Framework Research of the Internet of Things in Dispatching Emergency Supplies

Tongjuan Liu, Yanlin Duan and Yingqi Liu

Abstract Through the researches of emergency supplies dispatching problem have formed a certain scale. However, the sudden and diversity of events can easily lead to the lack of relevant statistical data and errors. Thus, it's easy to lead the mistakes on analysis and decision-making. The Internet of Things technology in the field of emergency supplies scheduling, it can provide more effective real time information to dispatch emergency supplies acquisition and analysis process, assist decision makers draft the scheduling plan, improve the scheduling efficiency. So that it can reduce the social and economic losses, the maximum time to reach maximum efficiency and cost minimization emergency rescue loss goal.

1 Introduction

The dispatching emergency supplies is a major decision problem in govern emergency supplies. Under normal circumstances, the cost minimization is a basic principle of materials scheduling. However, the dispatch of emergency supplies to emphasize its urgency, weak economy and dynamic characteristics, Therefore, the shortest time will be the first principle of dispatching emergency supplies. So, in the application of networking technology support, is able to obtain more information and data in real-time effective than traditional dispatching emergency supplies. By processing of these data to develop a highly efficient emergency dispatch plan.

Contra pose the problems and deficiencies of the traditional emergency supplies scheduling, combined the application and the architecture of the networking technology, establish a emergency supplies scheduling platform architecture based on networking technology.

T. Liu (✉) · Y. Duan · Y. Liu
Beijing Wuzi University, 101149 Beijing, China
e-mail: ltjjiaoxue@163.com

2 The Internet of Things and Related Technologies

2.1 The Basic Concept of the Internet of Things

Modern society is the world of the Internet and it promote the development of modern technology, network sharing applications, making the level of information continues to improve daily life; things came into being, the integration of the physical world and the world of information has become increasingly closer.

Definition of the concept of the Internet of thing according to the scope of the narrow and broad. The ‘Internet of Things’ narrowly referring to things and things are interconnected Internet including the connection of goods and goods and the connection of goods and radio. Things broadly refers to the integration of physical space and information space, which is the integration of real and virtual, the Internet will combine all things digital and networked, achieving the intelligent management of things.

2.2 The Basic Features of the Internet of Things

Things technology compared to the traditional internet industry, has the following unique features:

- (1) Widely used of the sensor. Application of sensor technology is widely used in the premise of things, provide real-time data and effective information system for the whole of things.
- (2) On the basis of the establishment of the Internet, combined with the telecommunications network technology.
- (3) The use of cloud computing technology to improve the treatment of Things intelligent level. Things to sensor technology and cloud computing, intelligent technology, the sensor mass data collected in accordance with certain rules and algorithms for analysis, processing and handling, enabling data classification, extraction of useful information, meet with different levels user’s demand.

2.3 The Key Technology of the Internet of Things

- (1) Sensor technology. Sensor technology is the basis for operation of the system of things, is the source of data throughout the system. Currently, toward the development of intelligent sensor technology, smart sensors will be an important indicator of the future development of intelligent things.
- (2) Radio frequency identification technology. RFID technology (Radio Frequency Identification, RFID) is an emerging automatic identification technology, it’s a use of the radio frequency signals through space coupling.

- (3) Network communication technology. Network communication technology is a key technologies to achieve things, the perception of the world, is a true premise of sensors and RFID technology and the Internet of things linked together, are the basis for data collection and information sharing.
- (4) Cloud computing technology. Cloud computing operating system to provide an efficient intelligent computing model to the operation of the internet of thing, provide reliable and efficient data storage center so that it can achieve information sharing data between different devices, making information between different levels of fast, efficient and accurate transmission technology.

3 The Application Framework of the Internet of Things in Emergency Dispatch Architecture

The Internet of Things technology in the field of emergency dispatch, emergency management can become integrated, comprehensive, real-time. Comprehensive application process things related technologies, the application architecture research is one of the very important part.

3.1 The Architecture of the Internet of Things

The Architecture of the internet of things usually includes perception layer, network layer, data layer and application layer. Specific architecture shown in Fig. 1.

(1) Perception layer

More emphasis on emergency supplies scheduling process timeliness and accuracy of materials dynamic information collection and transmission, and therefore, the perception layer sensor technology applications require a very high perceived level of precision emergency supplies. Accurate perception layer of emergency supplies and emergency vehicles, such as information collection and transmission is the premise and foundation of the entire emergency supplies scheduling process. The main information collection equipment perception layer include: EPC RFID tags, RFID readers, global positioning system (GPS), cameras, sensors, geographic information systems (GIS) and so on.

EPC RFID tags to uniquely identify each piece of emergency supplies, read the corresponding EPC code read by the reader and other devices to obtain the appropriate information on emergency supplies, emergency supplies for the realization of real-time tracking and tracing provides protection of the premise; EPC RFID tag reader for reading material stored in the appropriate information, via a wired or wireless network system to achieve emergency supplies information and links to the system of things; Sensor technology will deliver

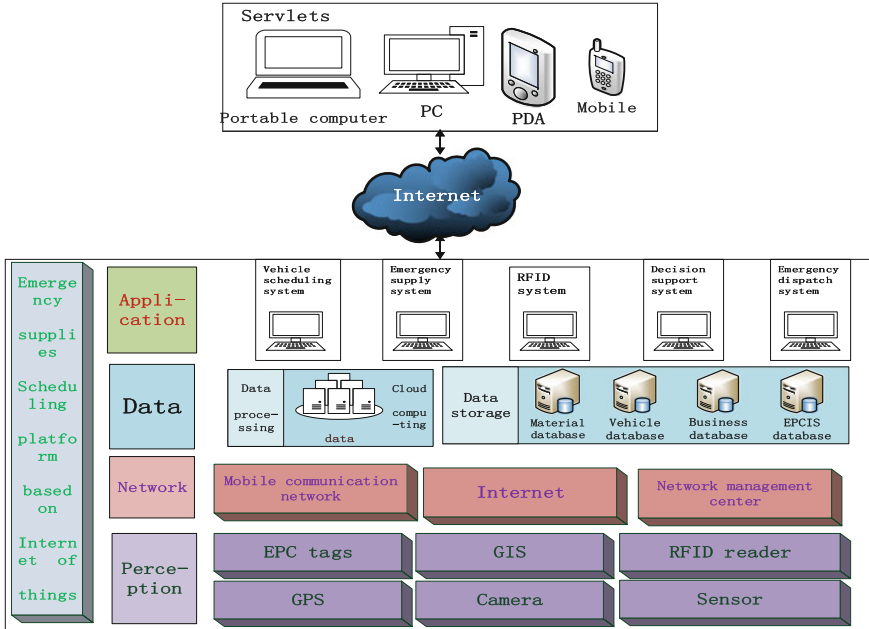


Fig. 1 Internet of things structure chart

real-time information to emergency supplies Things system, it's the main source of raw material things in information systems; camera technology enables security monitoring and out of the warehouse storage operations monitoring, enabling emergency supplies secure storage. GPS and GIS technology to locate emergency dispatch during emergency vehicles and emergency supplies, combined with computer systems and emergency supplies physical device management platform system, dispatching emergency supplies to remote control and command purposes and auxiliary scheduling decisions, so emergency vehicles can accurately and timely supplies will reach the destination, emergency supplies for transport, vehicle scheduling and decision support can play an irreplaceable role.

(2) Network layer

The network layer is responsible for converting the data format between different protocols and the transmission of data between the lower and super stratum. The equivalent of a bridge across things architectures, data link layer and the perception layer and application layer, making the physical world and the world to achieve the seamless information. Throughout the architecture system, all data will through the network layer, the mobile communication network, the Internet and network management center will perceive each network node link layer throughout the emergency supplies scheduling system, making emergency supplies to mobilize all sectors of the nodes in the system can get the appropriate information according to their needs.

(3) Data layer

Emergency supplies scheduling more attention to be completed within the shortest possible time, thereby reducing the social and economic losses, so emergency supplies scheduling information on the timeliness and accuracy requirements are very high, while the inability to forecast and EPC labels applied emergencies cause the original data redundancy, complicated, so we introduce the data layer, through the data processing and classification.

At the time of dispatch of emergency supplies related data analysis and processing, so it's necessary to introduce the cloud computing technology. Cloud computing speed data processing core, based on the internet of things dispatch emergency supplies to provide an efficient platform for intelligent computing mode operation, providing reliable and efficient data storage center, achieve the information sharing data between different devices, making information between different levels of fast, efficient and accurate transmission. Making all kinds of emergency supplies dispatching emergency supplies real time location and management, intelligent data analysis becomes possible, to minimize the processing time of emergency scheduling information, analysis and decision making to improve the efficiency of emergency command in order to achieve maximum benefit period.

Data storage layer includes a database of emergency supplies, emergency vehicle database and business databases. Emergency supplies database and emergency vehicle is mainly used to store the corresponding point of the emergency contingency reserve data, and business databases to store emergency supplies emergency scheduling process business data.

(4) Application layer

Application layer is the networking technology and related industries technology, the final presentation to the user's smart application solutions, but also the entire networking platform architecture can only interact with the user. This article is the application layer which networking technology and dispatch emergency supplies combined dispatching emergency supplies. Through the mode of Things environment, networking infrastructure, research, EPC encoding scheme, etc. So that it can initially set up under emergency supplies networking environment scheduling platform. To achieve the emergency response point location queries, efficient dispatch emergency supplies, emergency vehicle location monitoring, auxiliary functions scheduling decisions functions to provide technical support. In addition to these basic four-layer structure, there is a client tier, client layer provide different application modes for emergency supplies scheduling systems provide decision makers. Such as smart phones, PDA, mobile or PC terminals and other applications mode, making decision-makers can be anywhere on emergency supplies scheduling more timely and accurate decisions.

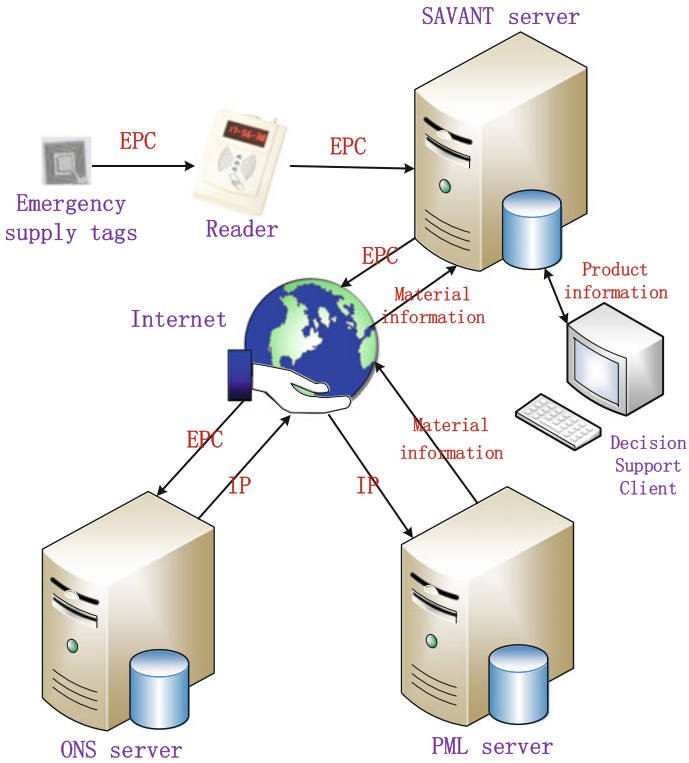


Fig. 2 Work flow of EPC networking system

3.2 The EPC IOT System in Emergency Supplies Scheduling

EPC IOT system is a very advanced integrated systems. It consists of three parts: EPC electronic coding systems, RFID systems and information network system. Entities and emergency supplies is closely linked to the Internet network, in order to achieve “connected” transitional system, played the entire scheduling system information communicated over the role. In the scheduling system, the work flow of EPC IOT system shown in Fig. 2.

The EPC IOT system in emergency supplies scheduling system is composed of EPC tags, readers, EPC middleware (Savant server), ONS server, PML server, emergency dispatch decision support systems and many database systems. The reader reads the tag encoding each on emergency supplies, but a corresponding EPC code, Things EPC system is to use this EPC code to get the corresponding IP address from the ONS server, access to the IP address corresponding to the PML server to obtain information related items stored in the server, Savant software and uses a distributed information system for processing a series of EPC read by the

reader. Since only one EPC code on the label, the computer needs to know additional information that matches the EPC, which requires ONS to provide an automated network database services. Savant will pass EPC ONS, ONS instruction savant holds a PML file look up server products, the file can be copied Savant, so the file can be transmitted to the emergency dispatcher product information application client.

3.2.1 EPC Electronic Coding System

EPC electronic coding is a global unified coding system to expand and extend the new generation of coding system, it's the prerequisite and basis for EPC encoding system. It can dispatch system for emergency supplies and emergency supplies were all one code, all the emergency supplies in accordance with specific data formats to store emergency supplies in database management platform, to provide emergency supplies scheduling and tracking of data and information assurance.

EPC electronic code is a regular digital collection consisting of a group of four regular of four: the header, vendor identification code, object classification code and serial number field. Depending on the code length and the length of each field, EPC electronic coding can be divided into three type: EPC-64, EPC-96 and EPC-256.

In real emergency supplies scheduling process, from the number of firms, the type of emergency supplies and emergency supplies reserve amount scheduling considerations, choose EPC-96 bit-type coding structure, the field is sufficient to identify every piece of emergency supplies, and do not have too much empty field.

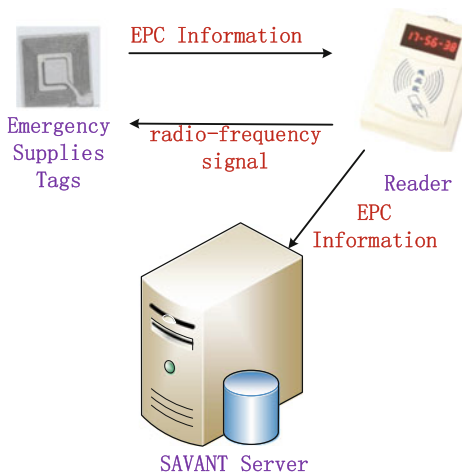
3.2.2 RFID System

RFID system is a section to identify and read EPC electronic coding, is a part of the first link to emergency supplies into the scheduling system. It mainly consists of EPC code tags, readers and antennas. Entire tag reading process as shown in Fig. 3. Certain frequency reader sends a radio frequency signal through the antenna, corresponding EPC tag is activated, and transmits its coded information to the reader; reader received signal is demodulated and decoded, and then sends the data to the background system.

3.2.3 Information Network System

Information management and interactive information network system is responsible for the system and the Internet of Things EPC system. By EPC middleware, Object Naming Service (ONS), EPC Information Service, four-part entity Markup Language (PML) collaborative complete.

Fig. 3 Work flow of RFID system



(1) Middleware of EPC

EPC middleware is known as “Savant”, a transition section connecting RFID radio frequency identification systems and enterprise applications, the main effect lies in the reader to read data to the disorder clutter preliminary processing and filtering, to reduce the amount of data transferred, and transfer read process appears to reduce redundant and erroneous data.

EPC is a series of message-oriented middleware “program module” or “service” integrated. It’s a interface to provide different information for different customers, in order to meet the information needs of different customers, specifically the principle shown in Fig. 4.

If the emergency dispatch client can obtain information through the intuitive interface material EPC middleware, etc. And other procedures such as ONS module system, the interface provided by the middleware needs to obtain information of the corresponding data format.

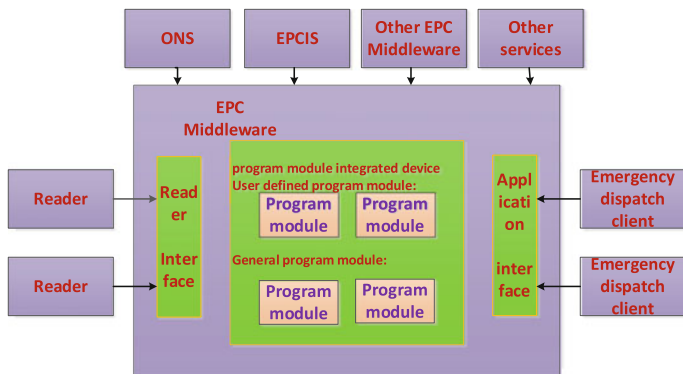


Fig. 4 EPC middleware functional diagram

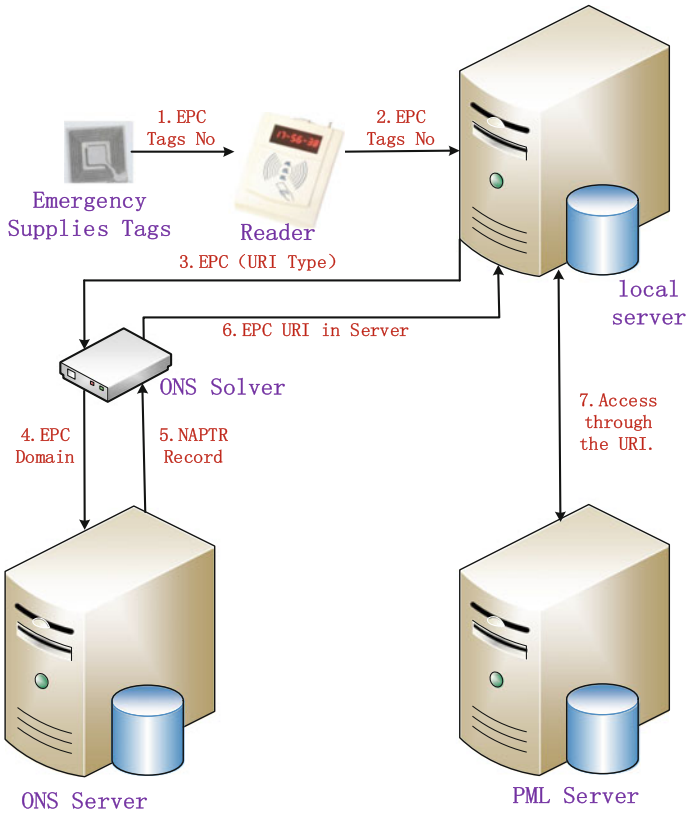


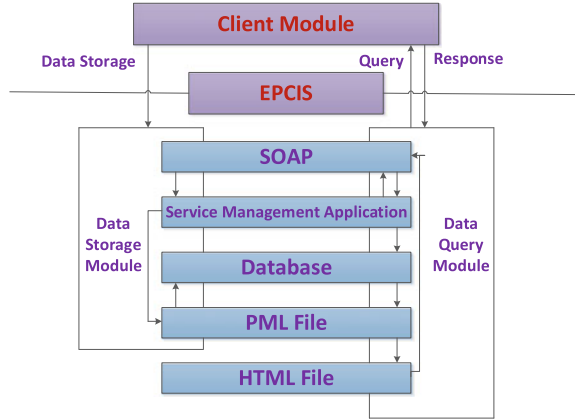
Fig. 5 Work flow of ONS

(2) Object Name Service (ONS)

ONS service for locating emergency supplies corresponding web service with EPC tags, by positioning the corresponding EPC encoding PML server storage material information to help achieve information service client applications. ONS is a distributed architecture, is able to use a maximum extent Internet architecture, the compatibility can be increased to some extent program. ONS detailed work flow system as shown in Fig. 5.

In the Object Name Service, ONS work flow details are as follows:
First, RFID system reads EPC tags emergency supplies through the reader, and the label will read the content delivered to the local server;
Second, the local server based on specific standards EPC tag data encoded into the corresponding URI format, and sent to the ONS solver;
Third, ONS solvers use format string conversion received EPC URI converted to a domain name, the domain name will have the authority pointer (NAPTR) query;

Fig. 6 EPCIS system function module diagram



Fourth, ONS server receives a pointer returned by the appropriate answer a series of NAPTR records, which contain one or more related point server URI; Fifth, local ONS extracted from the NAPTR record parser returned URI PML server needs to return to the local server; Sixth, according to the local server URI link corresponding PML server, obtain the required EPC information.

(3) EPC Information Services (EPCIS)

EPC Information Services are things in the system core module, which implements the EPC Things codes and emergency supplies information stored separately, only the storage of materials that EPC tag code, while all information is stored in the EPCIS emergency supplies server, thereby reducing the cost of the tag.

EPCIS functional modules include client module, data storage module and a data query module of three parts, as shown in Fig. 6.

Client module EPC tag information transmitted to the designated EPCIS server; The universal data storage module data stored in the database, invoke generic data attribute information generated for each product, and stored in the product information document PML initialization process; Data query module based on the query requirements and client privilege, access to the appropriate documentation PML, generate HTML documents, returned to the client.

(4) PML

When EPC commodity information identification, all the information about the product are standard with a new computer language—entity Markup Language (PML) writing, PML is based on widely accepted Extensible Markup Language (XML) evolved, PML via a standard, generic way to describe the physical world where people, it has an extensive hierarchy. PML’s goal is to provide a simple remote monitoring and environmental monitoring for physical entities, generic description language.

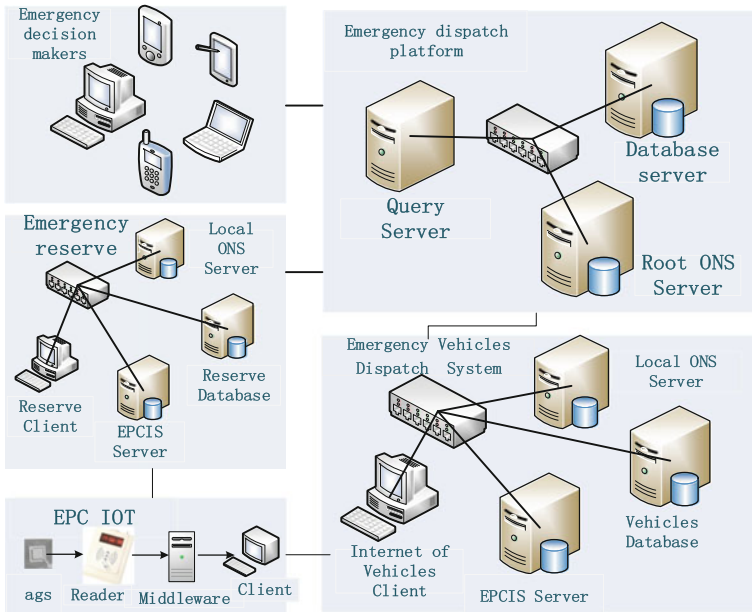


Fig. 7 The application of the logistics network application framework for emergency mater

3.3 Things Application Architecture for Dispatching Emergence Supplies

In this paper, the Internet of Things technology used in emergency supplies scheduling can improve the efficiency and accuracy of data collection capacity transmission, improve decision-making aid emergency dispatch. Things is a collection of perception, the Internet, computing and control in one of the intelligent system will be applied to the field of emergency dispatch, realize emergency data collecting, transfer and analysis. Finally, to achieve the positioning of emergency supplies and aid decision-making, the overall architecture of Things applications as shown in Fig. 7.

Things environment under emergency supplies scheduling system is an integrated information perception, messaging, integrated data processing systems, emergency locator and auxiliary functions such as scheduling decisions. Including emergency supplies reserve system, emergency vehicles and emergency dispatch scheduling system.

(1) Emergency supplies reserve system

Emergency supplies reserve system is the material basis for the operation of the emergency scheduling process, is the entity object emergency dispatch through stockpiles of emergency supplies for all points within the EPC tags were bound, and the reserve point information within the server all the details

are stored emergency supplies. When positioning supplies and information tracing by scheduling platform, through ONS resolution services, call the inquiry within the corresponding material reserves corresponding point database and information server supplies the information, enabling visualization of emergency supplies and tracing, improve decision-making.

(2) Emergency vehicle dispatch system

Emergency vehicle scheduling system is a necessary means to achieve scheduling emergency supplies, emergency supplies transported from stockpiles point to point in the process affected the need for adequate freight vehicles involved in the distribution process to ensure that the affected point timely and adequate emergency supplies and distribution. Throughout the scheduling process, emergency vehicle dispatch system can locate all of the queries in transit vehicles and vehicle load, so that timely dispatch vehicles to ensure timely supply of emergency supplies, ancillary scheduling decisions.

(3) Emergency management platform

Through the emergency supplies scheduling of Things mode, networking infrastructure, research EPC encoding scheme, set up under emergency supplies networking environment scheduling platform. By building the platform, Internet of Things technology links under each emergency scheduling module, reserve position location within the reserve point queries supplies, emergency vehicles query and location, assisted scheduling decisions and other functions. Emergency dispatch decision makers can through wired and wireless networks, the use of computers or mobile smart devices clients timely link into platform system. As long as there is a network of local, you can make scheduling decisions, improve the efficiency of emergency supplies scheduling, dispatching process to minimize the time needed to achieve the maximum benefit period.

4 Conclusion

Based on networking technology and application architecture studies in the field of emergency supplies scheduling, networking systems for key technologies, networking architecture and application architecture has been focused on.

Things of the physical world and the world of information seamlessly links, emergency dispatch system supplies all the emergency supplies and emergency vehicles links in the system can be in things, through sensors, GIS, cloud computing networking technology applications, improve emergency scheduling decisions efficiency, reduce emergency dispatch time. The Internet of Things technology used in emergency dispatch area is the inevitable trend of future development, not only for the development of networking technology is important, but also for the further developments in the field of emergency, rescue the affected point in time, reduce the social and economic losses, to protect the lives and property of the people have a far-reaching practical significance.

Acknowledgments This work was supported by funding project for Youth Talent Cultivation Plan of Beijing City University Under the grant number (CIT&TCD201504051) and this work was supported by Beijing Wuzi University Cultivation Fund Project (GJB20143006).

Simulation and Optimization of the AS/RS Based on Flexsim

Tongjuan Liu, Yanlin Duan and Yingqi Liu

Abstract As the development of China, The logistics industry has also made great strides, and there is lots of software to give the simulation of this system. Flexsim is one of them, which can be used to build up the discrete system and an ideal choice for the simulation of the Automatic Storage and Retrieval System (AS/RS). In this passage, taking the AS/RS as an example, by building the simulation of this system using Flexsim, we make some analysis about and then give some measures for optimization.

1 Introduction

Automatic stereoscopic warehouse is also called high-rise warehouse, AS/RS (Automatic Storage and Retrieval System). It is a kind of high-level stereo rack and a warehouse that make the operation using computer control and automatic stacking truck. The stereoscopic warehouse is consisted of high-level storage and mechanical equipment, buildings, facilities for controlling and other equipments.

An automated stereoscopic warehouse, a complex system that has many factors and many goals, usually contains kinds of logistic subsystems such as automated storage system, automatic guide vehicle system, automatic sorting system and automatic conveying system. Through the traditional analysis method, it is often difficult to get the optimal solution and usually needs a long time, also the cost is high, on the opposite, Flexsim is better in this aspect.

This paper takes an automatic stereoscopic warehouse for example. It makes the analysis using the system modeling and simulation technology, building a model of the stereo warehouse in the environment of Flexsim, and then give the result of this simulation. Then, We analyze the result and make the optimization to the allocation of resources in this automatic warehouse.

T. Liu (✉) · Y. Duan · Y. Liu
Beijing Wuzi University, 101149 Beijing, China
e-mail: ltjjiaoxue@163.com; ltj7905@163.com

2 Flexsim

Flexsim is developed by U.S. Flexsim Software Production Company; it is the first simulation software which is integrated C++ compiler in the graphical environment. In this situation, C++ can not only define the model directly, but also can have few problems. So, we no longer need the traditional dynamic link library and some other complex links.

Flexsim is simulation software based on objects, which used for dynamic discrete event system modeling and visual simulation and monitoring of some system. It set computer 3d image processing technology, simulation technology, artificial intelligence technology, data processing technology, specialized for manufacturing, logistics, etc. Using Flexsim system simulation software, it can establish the 3d model of the research objects in computer, and then analyze the data in the report, finally gain the optimal design or retrofit scheme.

3 Simulation Steps

The correct simulation step is an important guarantee for the success of system simulation, the automatic stereoscopic warehouse simulation generally has the following several steps:

- A. The system research:
System research is to understand the system operation condition and collect system data in order to further understand of parameters in overall process, so as to establish system model.
- B. Determine the target of simulation:
According to the different conditions, it has different targets and models, so it is necessary to determine the goal.
- C. Establish system model:
The system model is the description of the system, which consists of model and model parameters. The form of system model can be multiple, with written narrative type, flow pattern, the chart type, mathematical expression type.
- D. Determine the simulation algorithm:
Event scheduling method, activity scanning method and processes interaction method are the most commonly used algorithms at present.
- E. To establish the simulation model:
Simulation model is the process of standardization and digitalization. At the same time, it also needs some necessary parts according to the characteristics of computer operation such as input module, the simulation clock, random number generator and so on.
- F. The operation of simulation model:
It needs to determine the end time. Generally there are two kinds of termination method. A kind of method is to give a simulation time length, such as 8 h.

Another method is to determine the number of the event. According the different conditions, we choose the different ways.

G. The output of the simulation result:

It has two different measures: one is real-time online output, namely, a result of the simulation stage. The other is to give the value of all variables at the end of the simulation.

H. Analysis of results, forming simulation report:

The simulation results are analysed by statistical method, mainly to the accuracy and reliability of this system.

4 Target

Establish a model should first make clear the purpose of simulation, so that we can avoid the entwining of unnecessary details in the simulation process and emphasize key point of the problem.

The objective of this passage is to analyze the rationality of the design of the system, the feasibility and the system efficiency, which provides evidence for the selection of the control strategy such as system design, equipment configuration and management. It can also predict the handling capacity of the existing system, the work efficiency of conveyor, rail cars, stacking machine, etc. In this paper, our purpose is to find the deficiencies of this system and optimize the warehouse resource configuration, at the same time, meeting the requirements for the number of input goods 800, output 533 each day.

5 The Simulation of Stereoscopic Warehouse

A. System description:

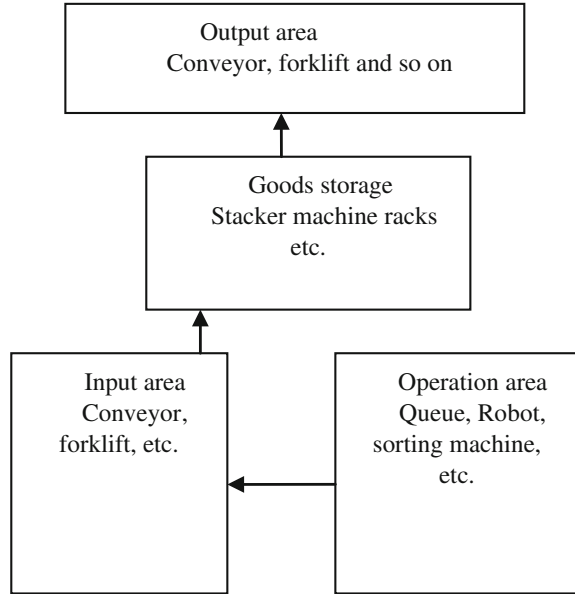
A company invests to build an automatic stereoscopic warehouse, which mainly for the storage of pallet materials. It has the limited cost and a high requirement of the efficiency of equipment as well as a fixed number of input and output goods.

Details will be described in the back of the article. We need to establish a model and analyze whether it can meet the demand of the whole system and obtain the best configuration of the resources.

B. The overall structure:

This stereoscopic warehouse is mainly composed of operation area, input area, goods storage area and the output area, its overall structure is shown as Fig. 1. The operation area is for the acceptance of the goods, sorting, packaging, labelling, check work, grasping the categories of goods, quantity, packing conditions and so on. Then the pallet goods are delivered out from the operation

Fig. 1 The overall structure of the warehouse



area to the platform, waiting for storage. The occupation of goods storage is definitely to offer a place for storage. Finally, Corresponding goods are carried from storage to the right place according to the outbound order, which is the duty of output area.

(1) The main equipment and functions in this model:

- a. Queue: pile up goods and pallets.
- b. Robot: carrying the goods that are separated to the place for output.

Processor: processing according the different categories of goods.

Transport: delivering goods in pallets from the combiner to the processor for input processing.

Combiner: after goods were sorted, we can use combiner to packing them into pallets, which makes it more convenient to taking them into storage.

Conveyor: carrying goods to the next procedure in the model, generally for the long distance relatively.

Rack: used for goods storage.

AS/RS vehicle: put pallet goods into appointed place, or taking the goods out of storage for outbound.

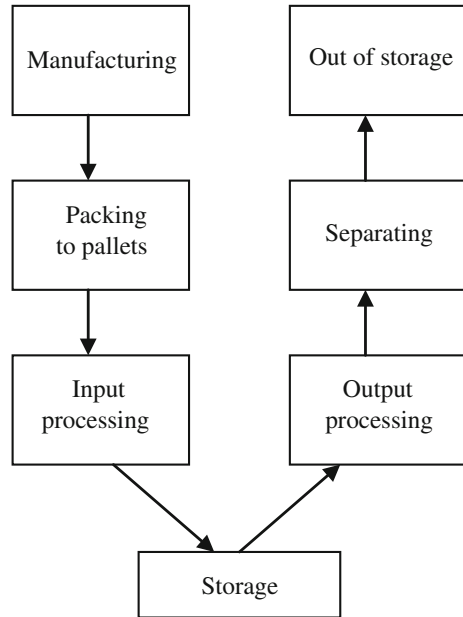
Separator: separate goods in pallets.

Operator: for goods delivering.

Dispatcher: assign goods to moving equipment such as operator or transporter.

The main working flow is shown as Fig. 2.

Fig. 2 The main working flow



(2) Assumptions and parameter setting

Establishing the model of simulation

Taking the actual requirements into consideration, we must put forward some assumptions for the simulation of this system, mainly has these:

We assume that the manufacturing rate obey index distribution exponential (0, 10, 1);

The minimum dwell time in shelves also obey exponential distribution whose mean for 2000;

Every pallet load 4 goods; one transport can carry one cargo each time the same as robot, and the capacity of AS/RS vehicle is 10;

There are six racks in this model; each rack is constituted by 10 levels by 10 bays.

The transporters, robots and AS/RS vehicles all have a maximum speed of 2 m/s, the speed of conveyors are 1 m/s, the processing time of one pallet is 5 s.

The simulation time is 8 h.

At last, we can set some other parameters to find a reasonable solution that can meet the requirements of the stereoscopic warehouse and rational utilization of equipment.

The whole model is shown as Fig. 3.

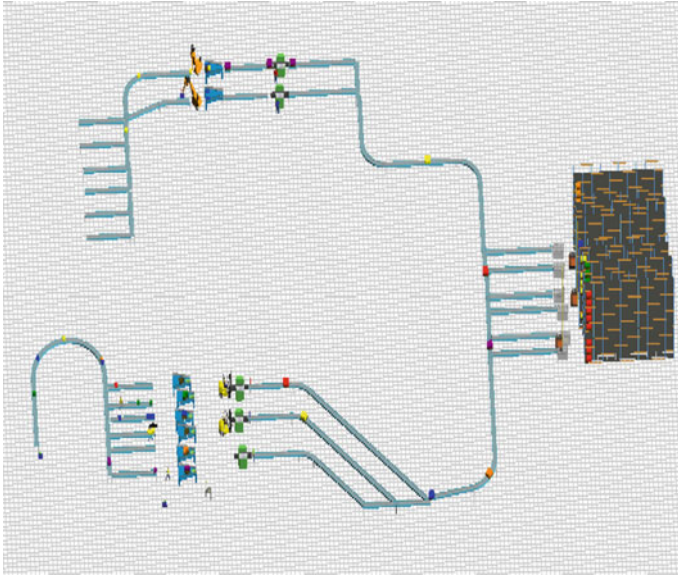


Fig. 3 The whole layout of the warehouse

Analysis and optimization of the simulation results.

First, we can watch the system operation process through the intuitive animation demo. It is easy to grasp and understand the overall situation of the system, and then we can find the bottlenecks. The main links of those combiners with other equipment are seen as Fig. 4.

When goods arrive, to whom they are carried by operators; goods are packed together here for pallets which can carry four single goods; and then they are processed before storing, such as recording the number of different categories of goods or scanning bar code. After this procedure, goods are delivered by conveyors to the warehouse, where there are queues for input and output areas, goods are delayed here for some time representing the storage. When an order is made by downstream distributors, goods begin to make preparation for output, which are separated by robots, out of pallets. At this time, the empty pallets are absorbed by the module sink.

The conveyors are shown as Fig. 5.

At the end of the system, the separated goods are delivered into trucks according the different orders.

In the current system, we know the ratio of some main equipment in Table 1.

Here the queues are for entering the warehouse; from the table we can clearly see that the efficiency of all these mobile equipment is very low when the warehouse is designed as mentioned above. And at this time, the number of input is 730, output 678, and also doesn't match with the requirements of the stereoscopic warehouse.

Through the analysis of data, I think the bottleneck of this system is the amount of transporters, which make the low efficiency and also the waste of this facility.

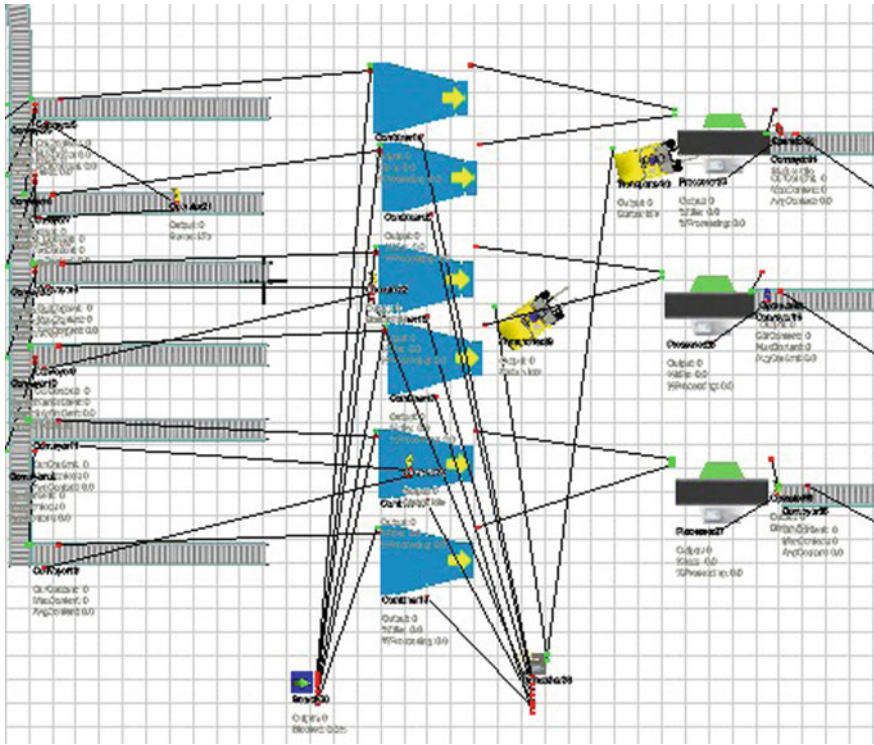


Fig. 4 The links in main equipment

So, we can take an improvement at this aspect. We decrease a transport in this node and increase the lift speed of the transport left probably, then we find its efficiency increased, the state can be seen as Fig. 6.

We can see the efficiency is about twice than before. In this system I think there is also another bottleneck which is caused by the queue for goods entering warehouse that make the low utilization of AS/RS. Namely, the capacity of queue is low, which has a bad influence of the whole system. So if we want to optimize the system, we must increase the capacity of queue, we can change the maximum content of the input queues from 10 to 30. Then when goods come, they can pile up at queues waiting for carrying, which can make the AS/RS vehicles busy, increasing the efficiency, after the improvement, we can see the results in Table 2.

From the comparison of these two tables, we can see all the efficiency of this equipment improved, and at this time the amount of input goods is 800, output for 533, consistent with the requirements of this warehouse. This is the simulation that we have to simulate each part of the warehouse to guarantee the integrity of the system, actually, when we establish a real warehouse, taking multiple elements into consideration, especially for low efficiency and large cover of the land, we can put queues of input and output areas together, which means a whole area but be

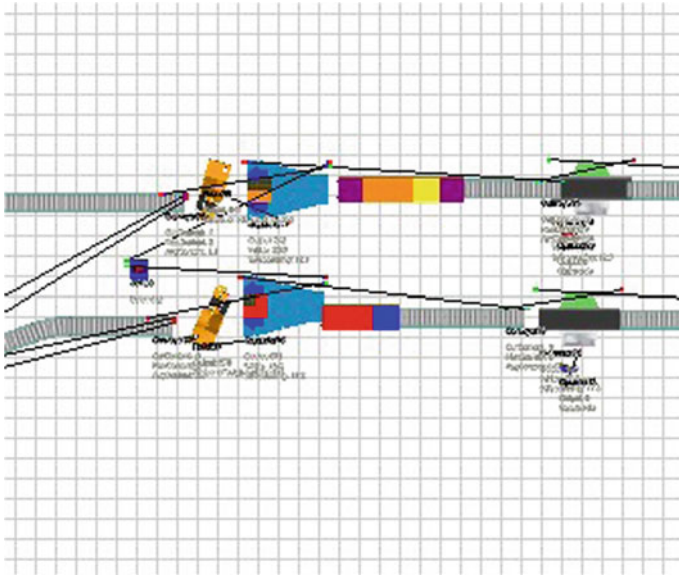


Fig. 5 Plate distribution figure

Table 1 The efficiency of some main equipment

Equipment	Efficiency (%)
Transportor1	5.2
Transportor2	20.4
AS/RS1vehicle	15.3
AS/RS2vehicle	14.0
AS/RS3vehicle	13.6
Separator1	84.5
Separator2	76.2
Robot1	77.4
Robot2	69.1
Queue1	3.0
Queue3	2.8
Queue5	2.3

Fig. 6 The state of transporter

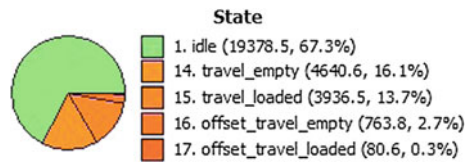


Table 2 The equipment efficiency after improvement

Equipment	Efficiency (%)
Transportor1	32.7
AS/RS1vehicle	45.0
AS/RS2vehicle	41.3
AS/RS3vehicle	58.4
Separator1	66.5
Separator2	61.3
Robot1	60.1
Robot2	55.6
Queue1	14.4
Queue3	13.6
Queue5	28.4

separated for two different functions, not the mixture of goods. On the one hand, we can satisfy the requirements of the warehouse, on the other hand, it can save a large space for other use, reducing the cost indirectly.

6 Conclusion

With Flexsim for automated warehouse modeling and simulation, we can get different data in manufacturing and management, this provide a theoretical basis for the actual operation of this warehouse. According to the characteristics of the automated stereoscopic warehouse, in this paper, we take an actual event for example, firstly, we analyze the overall structure and operation process of this automated warehouse; secondly, we establish the simulation of this system, and give out the simulation analysis of the process, finally, we give an analysis to the rationality of the parameters in the warehouse and make an optimization of the resource configuration, which provide decision-making basis for improvement of the whole system.

Acknowledgments This work was supported by funding project for Youth Talent Cultivation Plan of Beijing City University Under the grant number (CIT&TCD201504051) and this work was supported by Beijing Wuzi University Cultivation Fund Project (GJB20143006).

Design and Experiment of Control System for Underwater Ocean Engineering Structure Inspection and Cleaning Remotely Operated Vehicle

Haijian Liu, Zhenwen Song, Song Liang, Lu Chang, Renyi Lin, Wei Chen and Qingjun Zeng

Abstract The abstract should summarize the contents of the paper and the control system is designed for the new multi-functional and model-switched remotely operated vehicle, which is developed for detection and decontamination of underwater structure in Ocean Engineering. The dynamic propulsion system and controller unit was designed by analyzing the underwater motion force of robot. The surface console consist of power supply, control system, communication interface and PC software. The console can control the robot to complete medium-scale searching and point-fixed detection under remote control or monitoring. And the robot can switch modes each other between floating or climbing. Tests including posture, navigation, depth, monitoring and crawling clean-up verify that the ROV control system has reliable performances and it can satisfy the tasks in underwater complex environment.

Keywords ROV · Control system · Model-switched · PC software

H. Liu · Z. Song · S. Liang · L. Chang · R. Lin · W. Chen · Q. Zeng (✉)
School of Electronics and Information, Jiangsu University of Science
and Technology, Mengxi Road 2, Zhenjiang, China
e-mail: zheng28501@163.com

H. Liu
e-mail: kedaxiaoliu@163.com

Z. Song
e-mail: zhenwensong@yeah.net

S. Liang
e-mail: just_ls@163.com

L. Chang
e-mail: 921005971@qq.com

R. Lin
e-mail: 78621251@qq.com

W. Chen
e-mail: cwchenwei@aliyun.com

1 Introduction

For a long time, the land natural resources has been a lot of exploitation until exhaustion, while the ocean which contains abundant resources remains to be developed [1]. The underwater vehicle is a safe, economic, efficient tool for exploitation of marine resources, which can complete high strength and high work load in high depth and dangerous environment instead of divers [2]. The Remotely Operated Vehicle (ROV, Remotely Operated Vehicle) is a pioneer in exploration and exploitation of ocean, which has become a kind of important underwater operation equipment. The ROV is widely used in the field of detection of river and dam [3], marine engineering installation and repair [4], deep sea resources exploration [5], marine pipeline maintenance [6], underwater rescue [7]. Especially the underwater operation technology combined the ROV with divers has become an important mode in the technology of the detection and feculence-clearing of underwater structure. Therefore, it has an important significance to the efficient and safe operation of offshore platform by carrying out the research of ROV.

In recent years, the majority of the underwater robot only has floating or creeping ability, while the underwater robot with ability of floating and creeping is very rare [8], at home and abroad. The American LBC underwater robot is a kind of underwater monitoring robot with ability of floating and creeping. The robot can crawl with four-wheel drive, and product negative pressure adsorption force by using vortex generators [9]. But the robot has numerous dynamic driver device (9 motors), the power supply system with big power rating and power consumption, too much controlled object, high complexity, motors with lower utilization, high manufacturing cost and difficult processing.

This paper design a control system for the MC-ROV based on the research of the key technology of the domestic and foreign typical ROV control system. The water control system is designed for the new multi-functional and model-switched ROV (referred to as MC-ROV), which is developed for the detection and decontamination of the underwater structure in Ocean Engineering. The MC-ROV can complete medium-scale searching and point-fixed detection under remote control or monitoring in the complex environment by communicating with the mother ship through cable. On the other hand, MC-ROV can realize the floating investigation and crawling clean-up operation and switch mode between floating and crawling mode.

2 System Design

The sensors will transmit the real-time underwater information to the PC, after the multi-function and mode-switched robot MC-ROV was placed into water. Referring to the information of sensors, operators use PC software or operating handle to control the underwater operation of MC-ROV. The underwater body can realize floating movement and use observation equipment to do underwater

reconnaissance operations by its own thrusters. At the same time, the robot can do underwater reconnaissance operations by observation equipment. The robot can switch mode between floating and crawling mode by changing the action point of vertical motor power, after accessing to the target surface. When the longitudinal motor and propeller coming to gear mesh, the vertical motor can realize 4 degrees of freedom floating movement. At the same time, when the longitudinal motor and longitudinal propeller coming to gear mesh, the vertical motor can realize crawling clean-up operations. Gear box can transmit power to rear two wheels and cleaner roller. Adsorption thrusters can provide adsorption capacity when the robot are crawling. The design of the adsorption thrusters of MC-ROV is multi-functional and modularized, which contribute to update and maintenance of components. The entire system is low power, convenient and simple, energy-saving. The robot can crawl and clean through the wall and do floating investigation, which makes it has a wide range of applications.

According to the function and characteristics of the above, the control system of MC-ROV is divided into the water control system and the underwater control system. The two subsystems transmit signal and energy through the umbilical cable. The general structure of the control system of MC-ROV as shown in Fig. 1.

The water control system includes power supply, joysticks and PC software. The system can control the motion of underwater robot and display real-time sensor information and video information when the robot under water. The underwater control system is divided into four parts: power propulsion unit, underwater controller, visual lighting module, underwater sensing devices. Among of them, the power propulsion unit and the design of underwater controller are the emphasis and difficulty. The following will be discussed in detail. The vision lighting module consists of a HD underwater cameras and two LED underwater lamps. The underwater sensing device is composed of the inertial navigation unit and depth transducer, which can capture the real-time information. The underwater controller will handle the information and then upload them to the water control system. The operators can refer to that information to control the floating movements and crawling clean-up operations.

3 Design of Underwater Control System

3.1 The Design of Dynamic Propulsion Unit

Five thrusters of the MC-ROV provide floating power. When the two longitudinal thrusters has same speed, the robot can do the forward or backward actions. The robot also can do turning actions when the thrusters have different speed. Three vertical thrusters with same speed can do ascending or descending. On both side of them, the two thrusters with different speed can adjust pitching attitude, which makes early preparations for the crawling. This layout scheme of the thrusters can

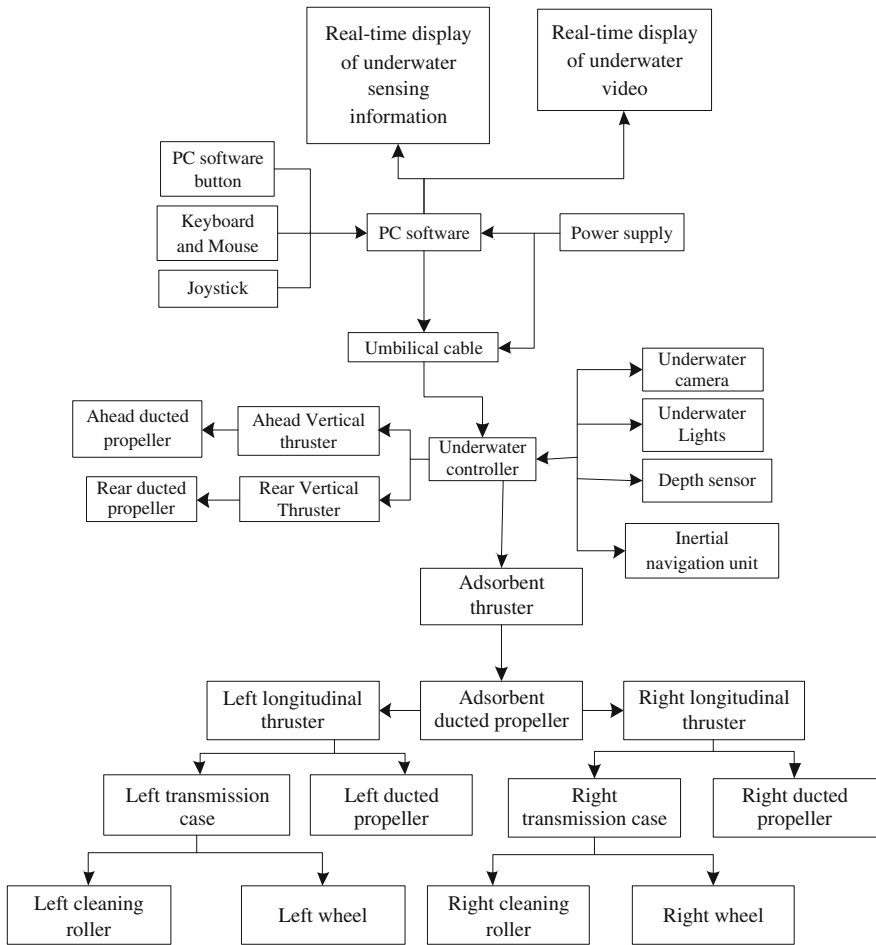


Fig. 1 The system structure of the MC-ROV

satisfy the movement of advance, retreat, Ascent, submersible and turning, when the robot in floating mode. At the same time, that scheme can adjust to the pitch attitude of underwater robot. The arrangements of horizontal thrusters and vertical thrusters are shown in Figs. 2 and 3.

The underwater motion force equation of six freedom degrees robot as follows:

$$\begin{bmatrix} F_{Tx} \\ F_{Ty} \\ F_{Tz} \\ M_{Tx} \\ M_{Ty} \\ M_{Tz} \end{bmatrix} = \begin{bmatrix} T_1 + T_2 \\ 0 \\ T_3 + T_4 + T_5 \\ 0 \\ b(T_3 - T_5) \\ a(T_1 - T_2) \end{bmatrix} \tag{1}$$

Fig. 2 The arrangement of horizontal thrusters

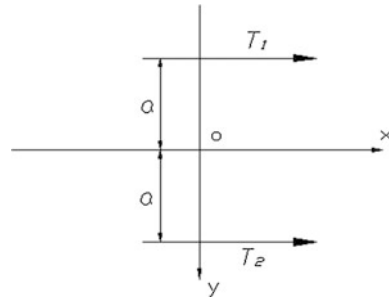
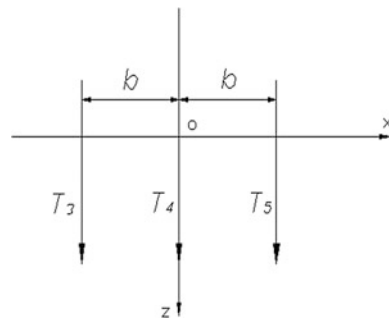


Fig. 3 The arrangement of vertical thrusters



Among of them, T_1, T_2, T_3, T_4, T_5 respectively represent for the five thrusters. F_{Tx}, F_{Ty}, F_{Tz} respectively represent the propeller of thrusters in the longitudinal, lateral and vertical direction. M_{Tx}, M_{Ty}, M_{Tz} respectively represent the torque of thrusters in the motion of roll, pitch and rotating. a represents the distance between the longitudinal thrusters and axis. b represents the distance between the vertical thrusters and axis.

Due to the motion of advance and retreat of the underwater robot is the main motions, we can estimate the propeller of thrusters according to the longitudinal resistance [10, 11]. The resistance of underwater robot includes two parts: exercise resistance and cable resistance. The resistance equation of the MC-ROV is shown as follows:

$$F = \frac{1}{2} C_d \rho V^2 L^2 \tag{2}$$

Among of them, C_d is the drag coefficient. We take the value of which between 0.1 and 0.2. Here we take 0.12. V is the movement speed of underwater robot; ρ is the density of water, L is the characteristic length. Here we take the longitudinal length of MC-ROV at 1 m. We estimate the cable resistance by the following equation:

$$R_d = \frac{1}{2} C_d \rho V^2 A \quad (3)$$

Among of them, A is the characteristic area. A is equal to the diameter of the cable multiplied by the length that vertical to the flow direction. Here we take 0.38. Thus we can deduce the output power of a single thruster according to the Eq. (1):

$$P = (F + R_d) \cdot V/2 \quad (4)$$

Considering the calculation results of the output power and the state of the low-speed navigation of underwater robot, we select the HP three-blade thruster, 200 W. The underwater dynamic propulsion unit is divided into four parts: vertical propulsion module, longitudinal propulsion module, mode-switched module, crawling clean-up module. The design scheme of the dynamic propulsion system as follows: Five thrusters can realize the four degrees of freedom motions and operations in ascending, descending, advance, retreat, yawing, pitching and the crawling clean-up operation. In the vertical propulsion module, two vertical thrusters can realize the motions of ascending and descending. The adsorptive thruster of mode-switched module can achieve adsorption for wall. When the robot in the state of floating, two longitudinal thrusters drive two longitudinal propeller to do the motions of advance, retreat and yawing. When the robot in the state of crawling clean-up, two longitudinal thrusters, meanwhile, drive two rear small wheels and two cleaning roller brush to achieve the crawling clean-up operation synchronously. The adsorptive thruster can also drive mode-switched module to do vertical motions to switch mode. The entity of mode-switched module as shown in Fig. 4. The entity of crawling clean-up module as shown in Fig. 5.



Fig. 4 The mode converted device

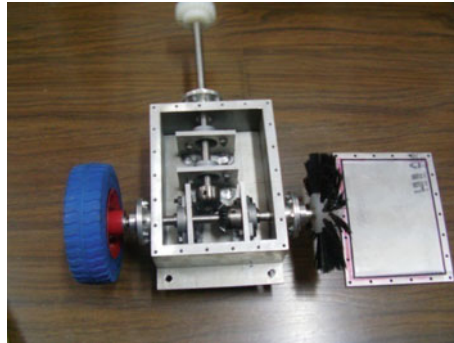


Fig. 5 Interior structure of the module

3.2 The Design of the Underwater Controller

The underwater controller unified administrate and control the dynamic and behavior of the robot based on the core of ARM microcontroller K60 (Cortex-M4). In Fig. 6, the inertial navigation module use Mini IMU inertial navigation module to collect the information of acceleration and gyroscope. Meanwhile, the module can collect the real-time attitude information of ROV by calculating the attitude date. We choose silicon pressure resistor sensor to capture the depth information of

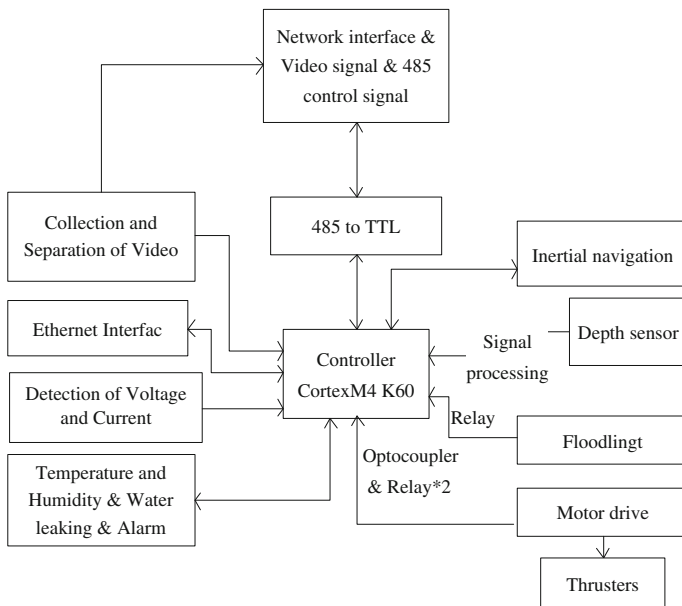


Fig. 6 The hardware structure of underwater vehicle

underwater robot in ascending and descending. We choose 12 V/200 W brushed DC motors as thrusters, which are driven by full-bridge drive. When the controller sends the PWM signal, it can control the speed of motors. The ROV owns a high-definition simulative cameras which can capture video. The design of the separation module for capturing video is useful for the transmission of the information of underwater video. Due to the robot need four kinds of power: 24, 12, 5, 3.3 V, we supply power by adopting combination of DC/DC regulated power supply and DC/DC isolation power. Otherwise, we design the module for detecting current and voltage. The communication of system uses 485 bus and TTL-485 communication module, in half-duplex mode.

4 The Design of Water Surface Console

The water surface console contains supply systems, control systems, and communication interfaces. The control system uses 220 V AC power supply to power the computer of water surface console. On the other hand, the 24 V DC power produced by high-power AC-DC power can power the body of MC-ROV. The water surface console also contains a PC computer, a set of 485 communication system, a video acquisition module and a rocker. The PC needs to timely display the data of inertial navigation equipment (three axis gyro, three shaft acceleration, three axis magnetometer), manometer (measuring depth), speed of thrusters (target, actual velocity). At the same time, the PC will show the calculated attitude information. Through the inertial navigation algorithm, the PC can display the 3D motion curve of MC-ROV. If allowed, we can do some further processing for the picture captured by camera. The console of water surface as shown in Fig. 7. The main interface of PC control as shown in Fig. 8.

The upper area of control interface display the state information of underwater robot (including depth values, the state of lamp, pressure value, the state of electronic lock, the speed of the motors, the state of switch, etc.). We can set the path of data saving, start or stop of data saving, communication rate by controlling the button. The middle part of the control interface is the area that display simulated state which

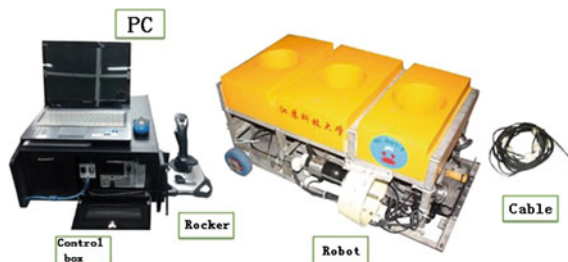


Fig. 7 The console of water surface

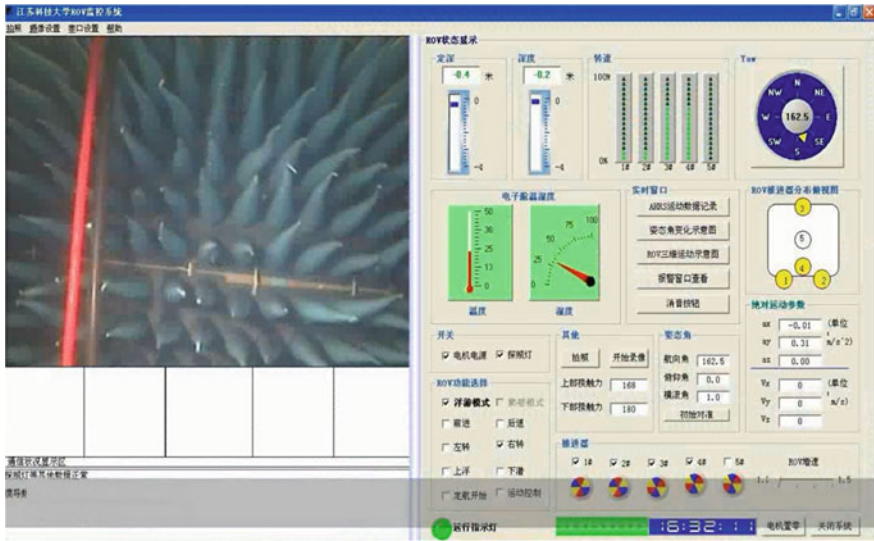


Fig. 8 The main control interface

contains the temperature and humidity of electronic cabin, real-time attitude data, the distribution of the thrusters. The following part of the control interface is the area for selecting model which can switch the modes of the MC-ROV. That part can control the robot to complete medium-scale searching and point-fixed detection under remote control or monitoring. Meanwhile the robot can switch modes each other in between floating and crawling. The system supports joystick operation which control the most of the functions of the control interface and simplifies the operation steps. The left side of the main interface for the display of video. The right side of the main interface multifunctional areas where contains the functions of control, display and storage of information, alarm, child windows, etc.

5 Experimental Study

The experiment of MC-ROV control system is based on the body of robot and the human-computer interface which is designed by using VC programming. The underwater test includes attitude measurement test, navigation test, depth test and monitoring and crawling clean-up test.

5.1 Attitude Measurement Test

First of all, we need to test communication, before the experiment of attitude measurement. That is to say, we need to test the correctness of communication

protocol and online of upper-lower computer. After preparation work, first of all, we open the serial port settings of the main interface of PC. Then, according to the communication protocol, we set the correct date of serial port, baud rate, data bit. After click the “opening ports” TAB, if the indicator light is bright, the PC begin to communicate and exchange date between the upper and lower computer. If the communication is successful, we will see the tips that communication is normal in the lower left corner of the main interface.

When the MC-ROV in the underwater navigation, we need to grasp its attitude angle in real time, in order to adjust the requirements. The attitude angles of underwater robot include the roll angle, pitch angle and yaw angle. The main purpose of attitude measurement experiment is to test the accuracy of the attitude angle and find out the error data. In the experiment, the attitude data of MC-ROV is automatically stored in the database. We can see the attitude changes and the trend of underwater robot visual in the interface by means of the posture curve and 3D motion diagram of MC-ROV.

5.2 Underwater Navigation Test

The directional navigation is not only one of the main functions of underwater robots but also one of the main index to measure the performance of underwater robot. After the desired angle was set, the MC-ROV system will record the real navigation data of the MC-ROV by the navigation sensor. After importing the experimental data into MATLAB software, the software will map navigation curve on the basis of navigation solution. The graph as shown in Fig. 9. According to the navigation curve, we find that the system of directional navigation has the following features: rapid response, low overshoot, stability time less than 20 s, steady error range under 3° , which meet the requirements of directional navigation. The causes

Fig. 9 Navigation test curve of MC-ROV

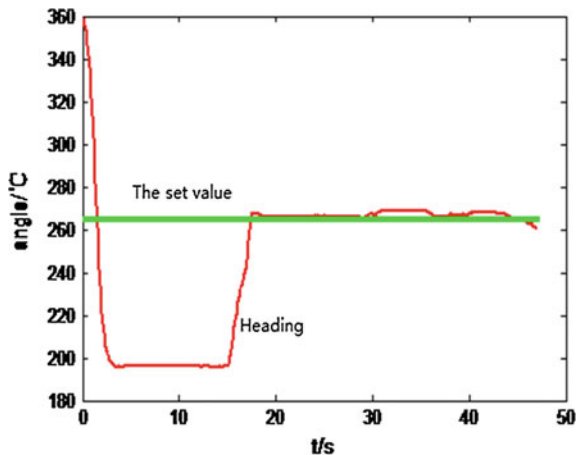
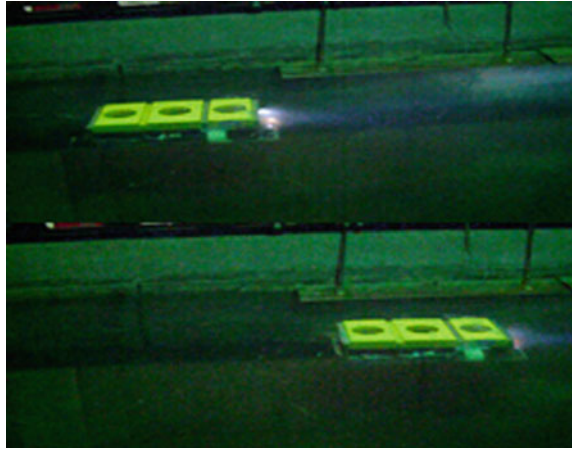


Fig. 10 Navigation test of MC-ROV



of errors are mainly due to the cumulative error of inertial device over time, which lead to the effect becoming poor. The underwater navigation test of MC-ROV is shown in Fig. 10.

5.3 Underwater Depth Test

The same as directional navigation, depth test is one of the main index to measure the performance of underwater robot. After the desired angle was set, the MC-ROV system will record the real navigation data of the MC-ROV by using the depth sensor. After the desired angle was set in the MC-ROV system, the system will record the real navigation data of the MC-ROV by the navigation sensor. Depth test curve and depth error curve are respectively shown in Figs. 11 and 12.

Fig. 11 Depth test curve of MC-ROV

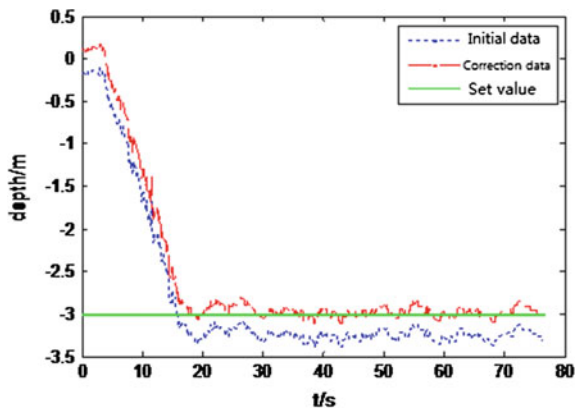
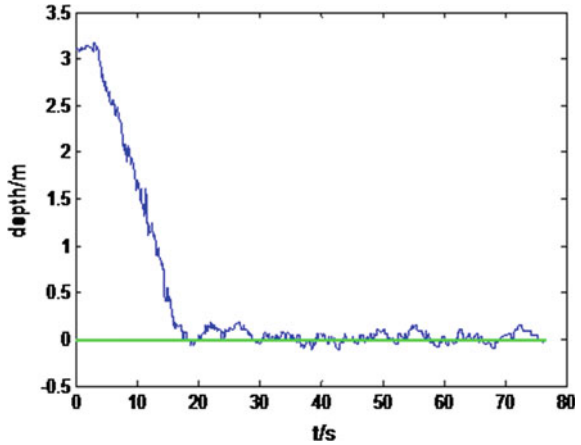


Fig. 12 Depth error curve of MC-ROV



In Fig. 11, the uploaded data of depth is sensor initial data. Revised data is the depth values which is obtained by filtering constant error and biggish error. The set point is the desired depth that the robot need to reach. After analyzing of the above two figures, we find that the MC-ROV takes about 15 s or so to reach desired depth. The errors basically fluctuate within 0–0.2 m. There are three main reasons: First, the power of thrusters is slightly smaller, which lead to the impetus is not enough. The second is that the error of depth device is large. Finally, the randomness of the water pressure cause an uncertain influence on the depth sensor. The depth test of MC-ROV is shown in Figs. 13 and 14.

Fig. 13 MC-ROV is doing depth-keeping

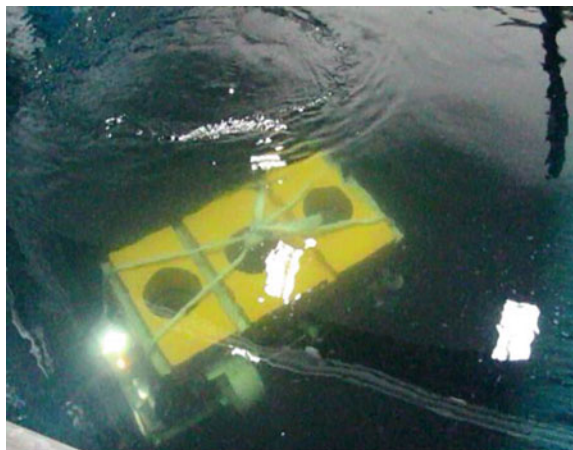
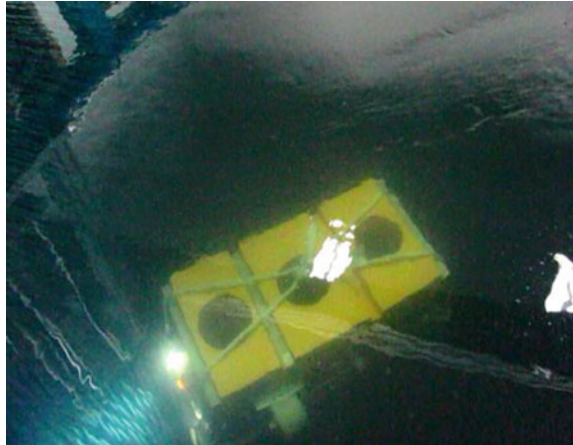


Fig. 14 MC-ROV has completed depth-keeping



5.4 *Monitoring and Crawling Clean-up Test*

The underwater video monitoring and crawling clean-up are the basic functions of MC-ROV. In order to realize underwater video monitoring function, we need to equip underwater camera and underwater lamps in MC-ROV. As shown in Fig. 15, in conjunction with two underwater lamps, the camera begin to work in monitoring mode. By switching mode, the MC-ROV work in crawling clean-up mode. As shown in Fig. 16, two rear wheels and two cleaning brush roller rotate synchronously to achieve underwater crawling clean-up operation.

Fig. 15 The video monitoring of MC-ROV

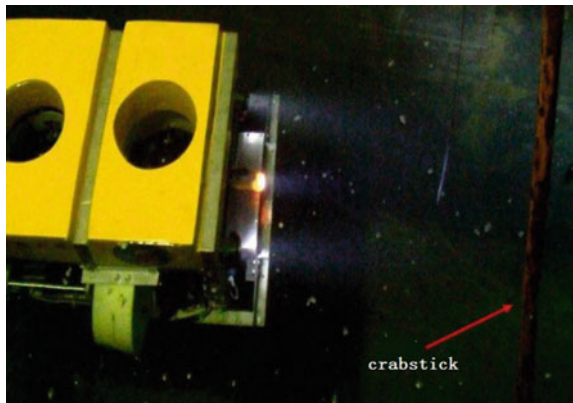
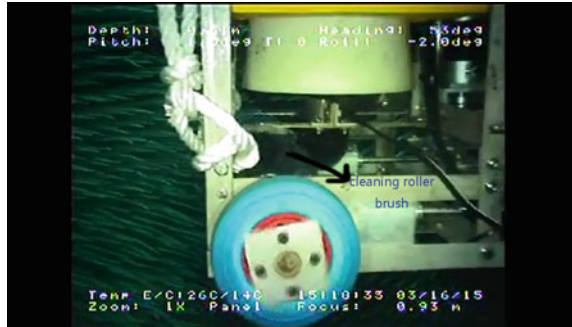


Fig. 16 The crawling clean-up test



5.5 The Summary of this Chapter

Through the experiments, the results show that the mechanical, software hardware of the control system and communication system are normal. The robot can better achieve monitoring and crawling clean-up function under water. Meanwhile, the curve of navigation and depth can prove that the stability of the system is better, which satisfies the requirements of use.

6 Conclusion

In this paper, we develop the control system for the new multi-functional model-switched remotely operated vehicle. The water surface control system consists of power supply, control system, communication interface and PC software. The system can control the motion of MC-ROV and display real-time sensor information and video information when the robot under water. The underwater control system is divided into power propulsion unit, underwater controller, visual lighting module, underwater sensing devices. We design the dynamic propulsion system and controller unit on the basis of the analysis for the underwater motion force of robot. The design of the water surface console consist of power supply, control system, communication interface. The PC software can control the robot to complete medium-scale searching and point-fixed detection under remote control or monitoring. The robot can switch modes each other in between floating and crawling. The results of underwater experiments have shown that the control system achieves satisfactory performance and it can satisfy the tasks in underwater complex environment.

Acknowledgment This project is supported by the National Natural Science Foundation of China (Grant No. 11204109), University Science Foundation of Jiangsu province in China (Grant No. 14KJB510008).

References

1. Liu ZY, Wang L, Cui WC (2011) State-of-the-art development of the foreign unmanned submersibles. *J Ship Mech* 15(10):1182–1193
2. Lionel L, Didik S (2007) Nonlinear path-following control of an AUV. *Ocean Eng* 34(2):1734–1744
3. Yong G, Yushan S, Lei W (2005) Research on the motion control system of GDROV. *J Harbin Eng Univ* 5:26–32
4. Rust IC, Asada HH (2011) The eyeball ROV: design and control of a spherical underwater vehicle steered by an internal eccentric mass. In: *IEEE international conference on robotics and automation*, pp 5855–5862
5. Zhu DQ, Liu Q, Hu Z (2009) Reliability control technology of unmanned underwater vehicles. *Shipbuild China* 50(2):75–80
6. Ma H, Jiang J, Du Y et al (2009) Subsea pipeline replacement repair technology and engineering practice. *Shipbuild China* 50(S):991–995
7. Azis FA, Aras MSM, Rashid MZA (2012) Problem identification for underwater remotely operated vehicle (ROV): a case study. *Proc Eng* 41:554–560
8. Yao F (2012) The basic motion control system of underwater robot structure and motion control technology research. Harbin Engineering University, Harbin
9. Rodocker D, Rodocker J (2007) Underwater crawler vehicle having search and identification capabilities and methods of use. United States: US 2007/0276552 A1, 29 Nov 2007
10. Wu JM, Yu M, Zhu LL (2011) A hydrodynamic model for a tethered underwater robot and dynamic analysis of the robot in turning motion. *J Ship Mech* 15(8):827–836
11. Wang T, Ye X, Wang L, Zhang C (2011) Hydrodynamic analysis and optimization for dish shaped underwater robot. In: *2011 international conference on mechatronics and automation*, pp 1406–1411

Using Experiment on Social Learning Environment Based on an Open Source Social Platform

Jing-De Weng, Martin M. Weng, Chun-Hong Huang
and Jason C. Hung

Abstract As we know, social platform has been very popular for many years, such as Facebook, Twitter, Youtube and so on. User on social platform may retrieve lots of information (breaking news, comments, sharing article...). Those information lead user to learn as their new knowledge, it's one kind of social learning. Otherwise, the popularity of portable devices (cell phone, Tablet) also help user to access internet and social platform in anytime and anywhere, user can get knowledge from social platform. In this paper, we conduct an experiment to compare the students learning on Elgg social platform on portable devices and the students learning in traditional methods. The results show us the students learning with social learning environment have better learning efficiency.

Keywords Social learning · Elgg · Learning experiment · Social platform

1 Introduction

E-learning, as a possible complement to traditional learning, delivers training or academic instructions electronically, in a structuralized and systematic form. Although e-learning has been developed for a long period, lots of efforts of

J.-D. Weng
College of Management, Yang-En University, Quanzhou City, China

M.M. Weng
Department of Computer Science and Information Engineering, Tamkang University,
New Taipei City, Taiwan

C.-H. Huang
Department of Computer Information and Network Engineering,
Lunghwa University of Science and Technology, Taoyuan City, Taiwan

J.C. Hung (✉)
Department of Information Technology, Overseas Chinese University,
Taichung City, Taiwan
e-mail: jhungc.hung@gmail.com

e-learning are on creating multimedia courseware, encourage interactions, and automatic assessment. Those implemented services really help students on learning performance that was proved by many study. For now, the social learning also integrated the advantages of traditional e-learning and provided several advantages that traditional E-learning may missed. In this study, we implemented several learning services on an open source social learning platform, and compare two groups of traditional learning and social learning on Elgg. The structure of this paper as below: Related works are discussed in Sect. 2. An experimental process and analysis are revealed in Sect. 3. The last Sect. 4 is conclusion.

2 Related Work

To have a better understanding of social learning, a wide range of related researches related to e-learning and social learning were studied. The differences between e-learning and social learning were summarized. We look at seven major perspectives which are general definitions, resources for learning, methods to learn, environment or channel to proceed learning, motivation of student. Conclusions in each column were given based on the experiences of previous researchers and us (Table 1).

Table 1 The comparison of social learning and traditional E-learning

	E-learning	S-learning
Definition	Delivering instructions to students at anytime and any place	Issues and solutions are discovered by students themselves
Learning resources	Pre-defined or prepared by specific instructors or domain experts	Everything shared by participants
Learning methods	Self-study with pre-defined curriculum	Through reading, sharing, and discussing with other participants
Environment/channel	Learning management system	Social network services
Motivation	With specific purpose (e.g., self-achievement, enforced by instructor, etc.)	With specific purpose and often enforced by student himself/herself
Interaction	With instructor or teaching assistant but interaction seldom takes place	Frequently-interacted with participants in specific groups

3 Experiment Conduction

In this study, we want to ensure the comprehensive of this experiment. We arranged two experiments, the first experiment (the comparison of experiment group and control group) conducted in a private university in Japan. The other one (Pre-Post experiment and social interaction with specific group) conducted in a private university in Taiwan. The Fig. 1 is the flowchart of this experiment.

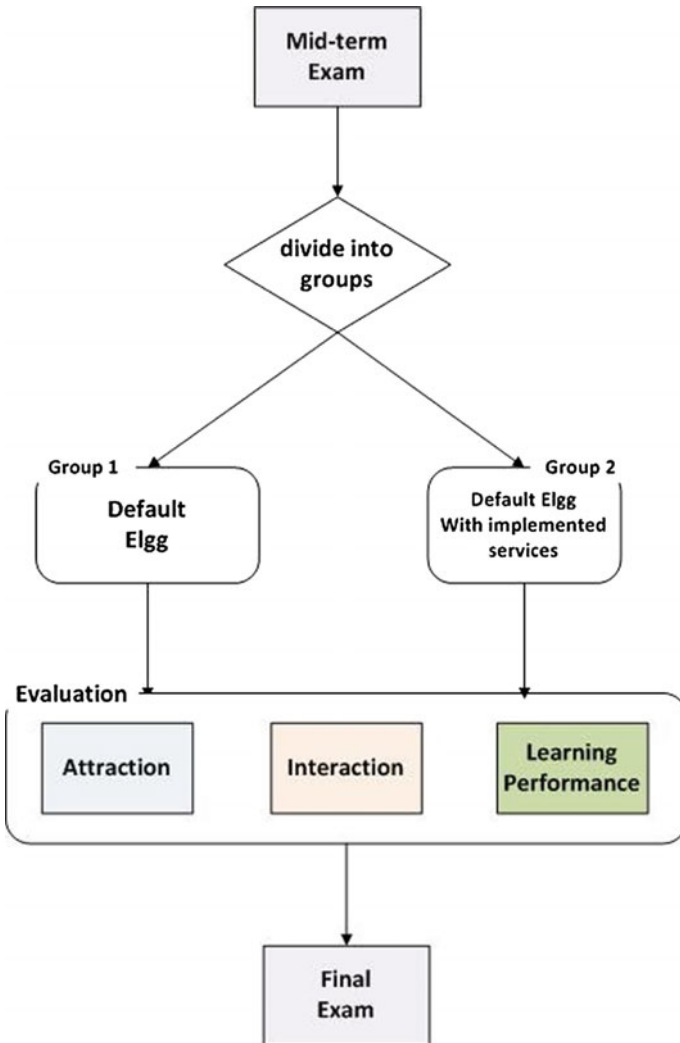


Fig. 1 The flowchart of first experiment

The Comparison of Experiment Group and Control Group

In this experiment, our research want to evaluate the following three items of our implemented services on Elgg, (1) attraction, (2) interaction (reply, post content, give comment, evaluate the content) and (3) learning performance. The first item, attraction, is the implemented services attract user to enhance their using frequency and using time? Because one of the most important factor in social learning is interaction, and we believe that the more time you spend on social platform, the more content you will retrieve from platform. The second item, interaction, since the interaction is the most evaluate factor in social environment, so the research made this evaluation for our implemented services. The third item, learning performance, this research want to know is the implemented services affect the learning performance? Positive affect or negative affect?

The subjects of this experiment consist of 138 freshmen from a private university in Taiwan. Those 138 students are from the department of German and department of Russian, over 90 % of subjects are freshman. We conduct this experiment in the class “Introduction of Computer Science” for six weeks that started after middle exam and finished before final exam. This study divided those 138 students into two groups (group 1 and group 2, each group have 69 people) according to their score of middle exam in average. Students in group 1 with the average score 76.3 of their middle exam and student in group 2 with the average score 76.8 of their middle exam.

The important factors that may affect social behavior are the attraction form services and the interaction on social platform. First, this study want to evaluate the attraction of implemented services, hence this study compare the average using frequency and average using time per week that recorded by system between group 1 (Elgg platform only with default services) and group 2 (Elgg platform both with the default services and our implemented services). As the results in Figs. 2 and 3, the average using frequency and average using time in group 1 became improved after week 2, since the average using frequency and average using time in week 1 are most than in week 2. Especially from week 5 to week 6 (just before final exam), the using frequency and using time enhanced obviously. We believed this is not

Fig. 2 The comparison of average using frequency per week

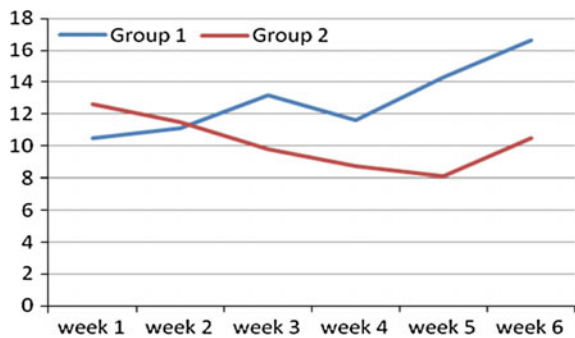
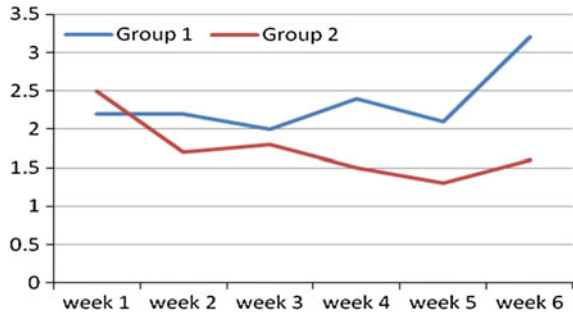


Fig. 3 The comparison of average using time (hours) per week



only the time period that approached final exam but also the helpful services that this study implemented, just because the implemented services are useful and attractive.

Second, this study want to evaluate the interaction between group 1 and group 2. Hence, this study counts all the related interaction on Elgg that including reply comment, post content and evaluate the content (like and dislike). In the Elgg, the system records every interaction we defined. As the results on the comparison of the average interaction times by each student can be seen in Fig. 4. The result indicates that the average interaction times in group 1 is more than the average interaction times in group 2 in every week. We believe the most important reason to affect the results is the implemented services. Because the lecture generation service will recommend more proper content to user that may cause user have more interaction with this content or other user. Otherwise, the influencing domain service will provide the *influencing topics* and *influenced users* and enhance the interaction with those influencing topics and influenced users by user (Fig. 5).

With the figure in Figs. 6 and 7, we can find the comparison of final exam, it is obviously the students in group 1 have better performance on final exam, especially on the score from 70 to 100.

Fig. 4 The comparison of the average interaction times by each student

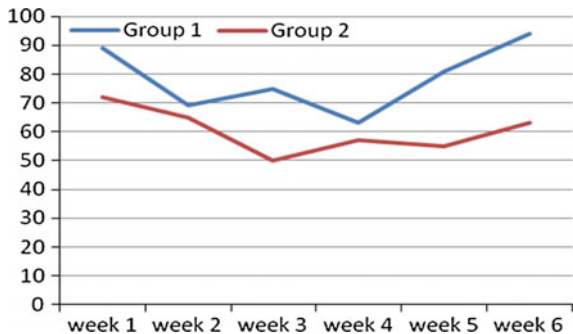


Fig. 5 The comparison of the average download times of learning content by each student

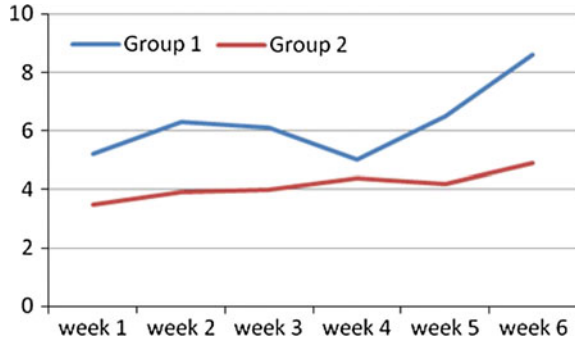


Fig. 6 The comparison of final exam by different score interval-1

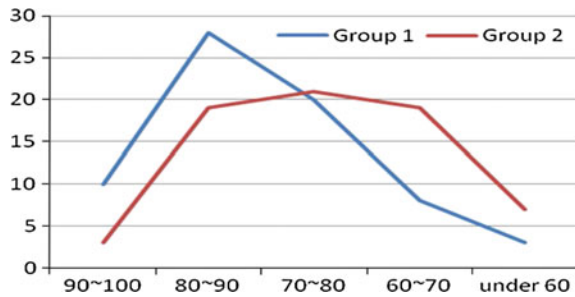
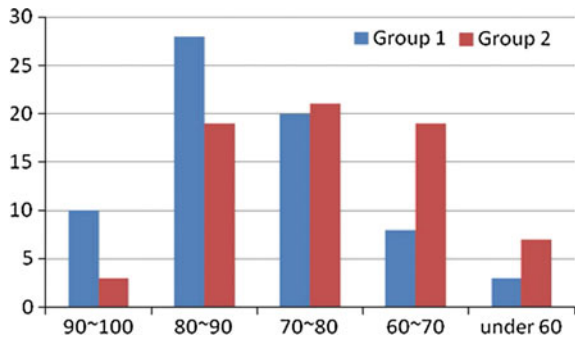


Fig. 7 The comparison of final exam by different score interval-2



4 Conclusion

In this study, we made a using experiment on social learning environment base on Elgg social platform, because we want to prove that the social learning has better efficiency that traditional learning with several implemented services. The outcome of this research indicates that the students group interacted with social learning have better learning outcome and learning efficiency. In the future, our research team want to implement more social learning mechanism and learning service that help user have better learning efficiency and learning outcome.

User Authentication Mechanism on Wireless Medical Sensor Networks

Wei-Chen Wu and Horng-Twu Liaw

Abstract This research can offer the function for users with different limits of authentication to access the system of control list; it includes the user identification, group identification and authorization. The authentication information stores in the smart cards in user authentication phase and adopts the Elliptic Curve Cryptography (ECC), hash function and symmetric key algorithm to get the authentication. Besides, this scheme contents forward security and backward security in order to protect the information security of patients. Compares with the previous authentication mechanisms which were proposed by other papers, the scheme of this research can provide more functions and higher safety on the application of WMSN without increasing the costs of communication and operation.

Keywords User authentication · ECC · Smart card · WMSN

1 Introduction

In recent years, the usage of WSN on medical care is mainly hope to improve the quality of medical care and reduce the costs through such techniques, and then, solve the problem of long-term care personnel shortage. WMSN is the network composed of a large number of sensor nodes and data collection machine. Sensor node usually consists of the processing unit, the sensing unit, the transceiver unit

W.-C. Wu (✉)

Computer Center, Hsin Sheng Junior College of Medical Care and Management, Taoyuan, Taiwan
e-mail: wwu@hsc.edu.tw

H.-T. Liaw

Department of Information Management, Shih Hsin University, Taipei, Taiwan
e-mail: htliaw@cc.shu.edu.tw

and a power unit with the limited computing capability. Apply WSN to medical treatment needs to pay some particular attentions as the following points [1]:

- Security

When deploying sensor nodes on the extensive and unrestricted area, it leads WSN much easier to be eavesdropped, copied and attacked since the open characteristic of wireless communication, and the hardware limitation for sensor node processor, memory, power and bandwidth etc. [2]. So, WSN must need to ensure that the authentication mechanisms can resist the relevant attacks of security in order to make sure that the information of sensor nodes can be detected accurately and provide uninterrupted service.

- Privacy

It must need to ensure that the patients' sensitive health information is protected under the environment of WMSN and only authorized users can access it. Therefore, this characteristic is also related to patients' privacy issues.

- Reliability

WMSN must ensure that the sensed data is completely correct and can be used. All of the data detected by WMSN is related patients' security of life, such as heart rate, blood pressure and temperature. If the information is incorrect or unavailable, it will affect the physician to determine the status of the patients, leading doctors to make the mistakes on determining the patients' conditions and even affect patient' safety.

Therefore, in order to increase the safety and efficiency of the network in recent years, WSN began to use the mutual authentication mechanism to achieve higher requirement of the security strength [3]. According to the reasons, this paper aims at providing user authentication mechanisms with mutual authentication to be applied in WMSN. Only authorized users can access patient's sensitive health information and to ensure the legitimacy of information sources.

2 Wireless Medical Sensor Network (WMSN)

WMSN is mainly composed of the large number of sensor nodes and wireless network [4]. According to different purposes of measurement, sensor nodes may contain the function of detecting the data of heart rate, blood pressure and blood sugar, then, the sensed data can be sent back to the servers or the devices of user through WSN, and the sensing information also can be instantly displayed on the user's device or transmitted to the server to do the follow-up treatment. When the sensing device reads the abnormal statistic, the system will automatically issue a warning by the sound through the relevant mechanisms, and sends the abnormal

statistic to the relevant medical personnel for informing them to make the appropriate response measures.

For example, Harvard University of America worked together with other research institutes to develop the platform, CoreBlue [5], this system is the use of WSN technology to achieve the effect of the medical care. The system uses wireless sensor node to detect the data and integrate the medical equipment of patient monitoring. Then, through the WSN, the collected information can be sent back to the hospital's server, so that, the patient can own the complete medical records to get the data collection of patient and make observation of patient medical status. It makes WSN become easier to be attacked and get the threats of security than other types of networks because of its simple, low cost, wide distribution and hardware characteristics of the resource constraints. When apply WSN to medical using, it especially needs to pay more attentions on the requirements of security as the following [1]:

- Data Confidentiality

The data on the sensor networks typically contains patient's information and even some unspeakable disease conditions. These data must need to be ensured to be accessed by the authorized users, referring physicians and healthcare personnel. So, there will be a very important security requirement of WMSN about how to maintain the confidentiality of information and won't let intentional people capture packets through the network eavesdropping to use them on some unlawful purposes.

- Data Authentication

To ensure that the data comes from the end of legitimate source and is transmitted to a legitimate destination by confirming the identity of each sensor nodes.

- Mutual Authentication

The biggest problem encountered by WSN used in the medical aspects is the issue of mutual authentication. If mutual authentication mechanism is not provided when the information is transmitted to the illegal nodes or personnel due to the broadcast characteristics of WSN, it may lead to the problem about leaking out the personal and pathogenic information of the patient. Therefore, users, servers and sensor node requires mutual authentication mechanism to verify the legality of the two sides for each other in order to ensure the source end and the destination nodes of WMSN are all correct and lawful.

- Data Integrity

Because of the broadcast characteristics of WSN, patient's data can easily be tampered by attacker during the transmission, it leads medical personnel to make the wrong judgment of patient's condition as the result of getting incorrect information. So, the WMSN needs to provide a mechanism to ensure that the

information has not been tampered with during the transmission and make sure to protect the security of patient.

- **Key Distribution**

On the WMSN, safe and effective key exchange mechanism is also very important, because the security of the symmetric key can ensure the safety of the subsequent exchange of information and the protection of patient's data privacy.

- **Access Control**

In the medical applications, there are various role of users need to access the physical information from the patient, but it is different for each role to access the data from the area. So, the WMSN, also require the role-based control mechanisms for users to access the data.

- **Data Availability and Freshness**

In order to provide uninterrupted physiological information as the foundation for related medical personnel and systems to determine the patient's condition, the patient needs to be continuously monitored. As a result, it must need to ensure that the sensor node can provide continuous operation and won't be ruined by intentional people, get caught in the entire network or transfer patient's old physiological signal information by attacks in order to avoid errors in judgment and even endanger to the patient's safety.

So, in order to avoid errors in judgment and even endanger to the patient's safety, it must need to ensure that the sensor node can provide continuous operation and won't be ruined by intentional people, get caught in the entire network or transfer patient's old physiological signal information by attacks.

- **Secure Localization**

In the medical applications, some of them need to know the exact location of patients, especially to take care of people with dementia section. So, the WMSN also need to make sure how to properly sense the patient's position and transport the information into the relevant systems and medical staff safely.

- **Communication and Computation Cost**

It is not suitable for using some security mechanisms which need to be processed by a lot of additional hardware resources on the WMSN due to its restrictions. So, how to provide an effective security mechanism to achieve a balance between security and the use of hardware resources is very important.

3 Our Scheme

The study assumed the WMSN of the hospitals as the system's environment. This network consists of the following three roles:

- Data collection layer

It contains the sensor node (Sensor) and data collection node (Sink). Sensor nodes are responsible for collecting patient's physiological signals, such as body temperature, blood pressure, blood oxygen levels and heart rate, and some of them are also responsible for collecting the related information of the sickroom's environment, such as the temperature and humidity. Data collection node will be primarily responsible for resending the data collected by the sensor node to the backend server or user's handheld device.

- Data access layer

The main users of this WMSN are physicians, nurses and other users need to access the information in the medical institutions. They can use a handheld device or a terminal device to access the relevant information for patients. The authentication is mainly based on the smart card which is allocated by the medical institutions. At the beginning, the user needs to set an identification as the user's login account. Each user can access the patient's information of allowed sensor node or server through a terminal device or a handheld device according to their privilege.

- Data processing layer

In addition to provide the subsequent processing of sensing data, the backend server for medical institutions is also responsible for generating the related parameters of the authentication mechanism in our scheme and store access control lists for each user's belonging group and various groups with access control list. This server also contains information of each patient owns by sensor node. When the user login successfully and complete the authentication, the server will notify the relevant sensor nodes to transmit data to the user according to their privilege.

Before the deployment of WSN, the server S generates the required parameters in advance, the parameters are as follows:

Server chooses a finite field F_p , p is a very large prime number and elliptic curve $E_p(a, b)$ coincide to equation $y^2 = (x^2 + ax + b) \bmod p$, and take a reference point G in $E_p(a, b)$. Then, server selects a random number X_S to be the private key of server S and calculates the public key P_S of server S through the following formula.

$$P_S = X_S \times G$$

Server generates necessary access control list ACL and GID in advance through each user's access right which is already owns by them and assigns the only identity SN_j , private key X_j and shared key MK_{js} with server S to each sensor node. Finally, selects one-way hash function $H(\bullet)$.

$$MK_{js} = (X_j \times G) \bmod p$$

Before deploying the sensor nodes to WMSN, server S will pre-load the following parameters into the memory of each sensor node.

- The identity of sensor node SN_j .
- The secret key X_j of sensor node SN_j .
- The shared key MK_{j_s} with server S .
- Elliptic curve $E_p(a, b)$.
- Reference point G .
- The public key P_S of server S .
- One-way hash function $H(\bullet)$.

Our scheme consists of three steps: registration, login, and authentication. The process of various phase are described in detail as the followings.

3.1 Registration Phase

The users have to apply for the smart card at the counter beforehand when they need to access the information on WMSN. First, the users make the registration from the server, so that the server can generate the smart card and user's access control lists toward the users. The phase consists of the following steps:

- The user U_i selects an ID_i , and choose a PW_i and a secret number RN_i that only known by the user.
- The user U_i submits the selected information $\langle ID_i, PW_i, RN_i \rangle$ to the server S .
- After server S receives the registration information submitted by the user U_i , it will be calculated in accordance with PW_i and RN_i :

$$RID_i = H(PW_i, RN_i)$$

- Then, the server S will select he appropriate group GID_i to generate the corresponding ACL_i based on the identity of the user U_i .
- The server S generates the smart cards of user U_i , there are $ID_i, H(\bullet), ACL_i, RID_i$, elliptic curve, reference point G and the public key P_S S of server S have been stored in it.
- The server S saves the related parameters $ID_i, GID_i, RID_i, ACL_i$ for the user U_i .
- To send the smart card to the user U_i .

3.2 Login Phase

After registration, the user can use smart card to login the server by entering the user's ID_i, PW_i and secret number RN_i through a handheld device or a terminal

device when they need to access the information on WMSN. The steps consists of the following steps:

- At the first, the user U_i inserts smart card into the reader and the terminal device or a handheld device, after that, enters ID_i and the password PW_i which is set by their own, and then, enter the secret number RN_i which only known by the user.
- Next, the terminal device or a handheld device will confirm if ID_i entered by a user U_i is equal to ID_i stored on smart cards. If there were not, the procedure of user's login may be interrupted.
- If they are equal, it will be calculated according to PW_i and RN_i entered by user U_i :

$$RID'_i = H(PW_i, RN_i)$$

- According to the calculated RID'_i , making the comparison with the RID_i stored on smart cards to confirm if they are equal. If they were, it means that PW_i and RN_i entered by user U_i are correct. It can continue to do the following procedure of login. If there were not, then, it means the password or secret number entered by the user U_i 's incorrect, the phase of login will be suspended immediately.
- After terminal device or a handheld device generates a random number a , and receives G and PS stored on smart cards, it will be calculated as follows:

$$Z = a \times G$$

$$Z_1 = a \times P_S$$

- The terminal device or a handheld device acquires the existing timestamp T_1 , then, it continues to calculate:

$$M_1 = ID_i \oplus H(Z_1 || T_1)$$

- The calculation through one-way hash function

$$H(RID_i || T_1)$$

- The value Z_1 calculated by the terminal device or a handheld stored device.
- To send $\langle M_1, Z, H(RID_i || T_1), T_1 \rangle$ to the server ACL_i .
- After server S receives the message transmitted from terminal device or a handheld device by user U_i . At the first, it confirms whether $T_2 - T_1 \leq \Delta T$. T_2 is the system time for the server to receive the message. ΔT is the allowable delay time for setting network. If it is true, then, continue with the subsequent procedure of login. If there were not, it means the packet of this message may be delay too long or suffer from replay attack leads by the problems of network. Then, the stage of login will be suspended immediately.
- Then, the server S will calculate with received Z and its private key X_S :

$$Z'_1 = a \times P_S = a \times G \times X_S = Z \times X_S$$

- Server S according to the received M_1 and calculated Z'_1 to receives ID_i by calculations:

$$ID_i = M_1 \oplus H(Z'_1 \| T_1)$$

- The server S finds out its corresponding RID_i by calculating through hash function $H(\bullet)$ with calculated ID_i :

$$H(RID_i \| T_1)'$$

- Next, the server S will compares calculated $H(RID_i \| T_1)'$ with received the value of $H(RID_i \| T_1)$, and it can confirm whether they are the same. If the calculated $H(RID_i \| T_1)'$ and received $H(RID_i \| T_1)$ are not equal, the step of login will be suspended immediately.
- If they are equal, then, the server S will generate the symmetric key B_i by calculating the access control list with available sensor node SN_j , access control lists ACL_i , the server's public key P_S and the shared key MK_{jS} of sensor node SN_j of user U_i :

$$B_i = H(SN_j \| ACL_i \| P_S \| MK_{jS})$$

- Then, the server S produces the symmetric key K_{si} according to ID_i , RID_i and the system time T_2 values for receiving the messages:

$$K_{si} = H(ID_i \| RID_i \| T_2)$$

- The server S uses the encryption of the secret key to encrypt the two parameters SN_j and B_i :

$$E_{K_{si}}(SN_j, B_i)$$

- The server S sends the message $\langle E_{K_{si}}(SN_j, B_i), T_2 \rangle$ to the terminal device or a handheld device of user U_i .
- After user U_i receives the message transmitted by the server S through terminal device or a handheld device, it confirms whether $T_3 - T_2 \leq \Delta T$ at first. T_3 is the time for server S to receive the message. ΔT is the allowable delay time for transmitting the message from network set by the system. If it is true, then, continue with the subsequent procedure of login phase. If there were not, then, the step of login may be suspended immediately due to the packet delay too long or suffer from replay attack causes by the problems of network.

- The terminal device or a handheld device calculates the same symmetric key K_{si} according to the parameters received T_2 and achieved ID_i and RID_i on smart cards.

$$K_{si} = H(ID_i || RID_i || T_2)$$

- It get SN_j and B_i by decrypting the symmetric key K_{si} with calculating $D_{K_{si}}(SN_j, B_i)$.
- The terminal device or a handheld device stores the acquired values of SN_j and B_i , and then, complete the mutual authentication with the server S and user U_i .
- S server obtains current system time T_4 .
- To encrypt SN_j , ID_i , RID_i and ACL_i through shared key MK_{js} and sensor node SN_j by using AES symmetric key algorithm:

$$E_{MK_{js}}(SN_j, ID_i, RID_i, ACL_i)$$

- The server S will transmit $\langle SN_j, E_{MK_{js}}(SN_j, ID_i, RID_i, ACL_i), T_4 \rangle$ to the sensor node SN_j .
- After sensor node SN_j receives a message from server S , it confirms whether $T_5 - T_4 \leq \Delta T$ at first. T_5 is the time for sensor node SN_j to receive a message. ΔT is the allowable delay time for transmitting the message set by the system. If it is true, then, continue with the subsequent procedure of login phase. If there were not, then, the step of login may be suspended immediately because the packet delay too long or suffer from replay attack causes by the problems of network.
- Then, the sensor node SN_j will confirm whether received SN_j is SN_j itself. If they are equal, then, continue with the subsequent procedure of login phase. If there were not, then, the step of login may be suspended immediately.
- To decrypt through shared key MK_{js} of the server S by using a symmetric key algorithm AES:

$$D_{MK_{js}}(SN_j, ID_i, RID_i, ACL_i)$$

- To get SN_j , ID_i , RID_i and ACL_i after decryption, and finish the confirmation of the legal identification of server S .
- If they are equal, the sensor node SN_j will store received ID_i , RID_i and ACL_i in the memory.

3.3 Authentication Phase

When a user enters a user ID , password and the secret number, and completes all steps in login phase, then, it starts to enter the step of authentication phase. The user

can access the WMSN after the confirmation of the legal status. The Z_1 , SN_j and B_i are mainly stored before the step of login phase to be used at the step of authentication phase. The process consists of the following steps:

- The user U_i acquires current time T_6 through the terminal devices or handled devices.
- Use the symmetric key B_i to encrypt ID_i , RID_i and Z_1 :

$$E_{B_i}(ID_i, RID_i, Z_1)$$

- The terminal devices or handled devices store the value of T_6 .
- To transmit $\langle SN_j, E_{B_i}(ID_i, RID_i, Z_1), T_6 \rangle$ to the sensor node SN_j .
- After sensor node SN_j receives a message from U_i through the terminal devices or handled devices, it confirms whether $T_7 - T_6 \leq \Delta T$ at first. T_7 is the time for sensor node SN_j to receive a message. ΔT is the allowable delay time for transmitting the message set by the system. If it is true, then, continue with the subsequent procedure of authentication phase. If there were not, then, the step of authentication may be suspended immediately because the packet delay too long or suffer from replay attack causes by the problems of network.
- The sensor node SN_j will confirm whether received SN_j is equal to the SN_j stored in the memory.
- The sensor node SN_j produces the symmetric key B'_i by calculation according to SN_j stored in memory, the access control list ACL_i , the server's public key P_S and server's shared key MK_{js} :

$$B'_i = H(SN_j || ACL_i || P_S || MK_{js})$$

- To decrypt through symmetric key B'_i of the sensor node SN_j

$$D_{B'_i}(ID_i, RID_i, Z_1)$$

- If the decryption is done successfully, it continues to do the subsequent procedure of authentication phase. The step of authentication will be suspended immediately if it fails to decrypt because of the wrong B'_i .
- To get ID_i , RID_i and Z_1 by decryption. Next, the sensor node SN_j will confirm whether ID_i and RID_i obtained from the decryption is equal to the ID_i and RID_i stored in the sensor node SN_j . If it is true, then, continue to do the producers of authentication phase. If it were not, the step of authentication will be suspended immediately.
- The sensor node SN_j produces the symmetric key K_{ji} according to SN_j , ID_i , RID_i stored in the memory, received Z_1 and T_6 with one-way hash function.

$$K_{ji} = H(SN_j || ID_i || RID_i || Z_1 || T_6)$$

- The sensor node SN_j replies the acknowledgement to the terminal devices and handled devices of user U_i (Acknowledgement).
- The sensor node SN_j deletes ID_i , RID_i and ACL_i stored in the memory.
- After the user U_i receives the message sent from the sensor node SN_j through the terminal devices and handled devices, it will produce the same symmetric key K_{ji} according to the stored parameter SN_j , Z_1 , T_6 , ID_i and RID_i stored in the smart card.

$$K_{ji} = H(SN_j || ID_i || RID_i || Z_1 || T_6)$$

- After generating the symmetric key K_{ji} , the information transition with of user U_i 's terminal devices and sensor node can use this symmetric key K_{ji} to connect the secure channel for making sure the security of transmitting the information through AES algorithm.
- The terminal devices and handled devices of user U_i deletes Z_1 , SN_j , B_i and T_6 stored in the memory.

4 Analysis

It can be seen through Table 1, at the step of registration, our proposed scheme in this study requires only a one-way hash function calculation ($1T_h$). As for the step of login phase, our proposed scheme in this requires two symmetric key algorithm operation, plus eight one-way hash function operation, plus three elliptic curve scalar multiplication ($2T_s + 8T_h + 3T_{mul}$). At the step of authentication, our proposed scheme in this study requires a symmetric key algorithm operation, plus three one-way hash function operation ($1T_s + 3T_h$). To synthesize the required time of computing at various, in total, our scheme mechanism needs three symmetric key algorithm operation, plus 12 one-way hash function calculation, along with three elliptic curve scalar multiplication ($3T_s + 12T_h + 3T_{mul}$).

Table 1 Comparisons of computation costs

Phases	Registration	Login	Authentication	Total cost
Costs	$1T_h$	$2T_s + 8T_h + 3T_{mul}$	$1T_s + 3T_h$	$3T_s + 12T_h + 3T_{mul}$

5 Conclusions

The proposed scheme in this study takes the effectiveness owned by the server itself, the sensor node's computing performance and storage devices into the consideration, so it can store a large number of shared key MK_{js} and key X_j of sensor nodes. The entire calculation of authentication scheme is aimed at the end of user and server because the user device and the server has a high computing effectiveness, and the limitation from hardware of sensor node is less compared to the using of calculation. And, it also takes the function for user to access data on sensor node with the public key or ECC system into consideration. It will take more time and power of computing, so the session key exchange in this study is based on the system of symmetric key. Our proposed scheme provides mutual authentication and the security issues also provides resistance to man-in-the-middle attack, replay attack, password guessing attack, privileged insider attack, user impersonation attack, using the same account attack, and provides a session key and forward/backward security.

References

1. Kumar P, Lee H-J (2011) Security issues in healthcare applications using wireless medical sensor networks: a survey. *Sensors* 12(1):55–91
2. Cao X, Kou W, Dang L, Zhao B (2008) IMBAS: identity-based multi-user broadcast authentication in wireless sensor networks. *Comput Commun* 31(4):659–667
3. Guo P, Wang J, Zhu J, Cheng Y (2013) Authentication mechanism on wireless sensor networks: a survey
4. Milenković A, Otto C, Jovanov E (2006) Wireless sensor networks for personal health monitoring: issues and an implementation. *Comput Commun* 29(13):2521–2533
5. Malan D, Fulford-Jones T, Welsh M, Moulton S (2004) Codeblue: an ad hoc sensor network infrastructure for emergency medical care. In: *International workshop on wearable and implantable body sensor networks*, vol 5

Application of Cloud Computing for Emergency Medical Services: A Study of Spatial Analysis and Data Mining Technology

Jui-Hung Kao, Feipei Lai, Bo-Cheng Lin, Wei-Zen Sun, Kuan-Wu Chang and Ta-Chien Chan

Abstract Out of Hospital Cardiac Arrest (OHCA) is an important medical and public health issue. Emergency first aid service prior to hospital admission is an important indicator for the quality evaluation of the emergency medical service. OHCA frequently occurs without warning, and while there are clear steps in emergency first aid concerning the treatment of OHCA patients, their survivability diminishes if they cannot receive emergency first aid services in time. Using statistical methods such as chi-square test, logistic regression, and decision tree, the influence factors were analyzed and extracted. In addition, combining the strengths of three independent spatial clustering analysis methods, namely, the Global Moran's Index for finding the spatial clustering, as well as the Local Moran's Index and spatial autocorrelation analysis Getis-Ord G_i^* algorithm, a novel summary approach to identify high-risk OHCA areas. The Global Moran's Index of OHCA event locations were 0.025861, with a Z-score of 8.178045, indicating significance spatial clustering phenomenon of OHCA locations, Getis-Ord G_i^* covers more towns (urban areas), but the High-High area reaching statistical standards obtained through the Local Moran's Index also has also appeared in the high clusters Area found through search using the Getis-Ord G_i^* . In addition, the important factors found through the decision tree analysis method have more space distribution coverage. When OHCA occurs, based on findings in this study, the 119-dispatch duty officer may make further inquiries regarding medical history of heart disease or

J.-H. Kao (✉) · F. Lai · W.-Z. Sun
Department of Computer Science and Information Engineering,
Department of Electrical Engineering, Graduate Institute
of Biomedical Electronics and Bioinformatics, National Taiwan University,
Taipei, Taiwan
e-mail: kao.jui.hung@gmail.com

J.-H. Kao · B.-C. Lin · T.-C. Chan (✉)
Center for Geographic Information Science, Research Center for Humanity
and Social Sciences, Academia Sinica, Taipei, Taiwan

K.-W. Chang
Division of Emergency Medical Service, Fire Department,
New Taipei City Government, New Taipei, Taiwan

diabetes, which shall serve as a reference for future dispatch of senior technicians. Based on the OHCA-prone hot zone generated by the Getis-Ord G_i^* and targeting OHCA patients' past medical history of heart disease or diabetes, public health units may adopt information technology or wearable devices as intervention in order to increase the probability of eyewitnesses and prioritize the dispatch of emergency aid resources into the hot zone, thereby enhancing OHCA patient survival rates.

Keywords Out-of-hospital cardiac arrest · Cardiopulmonary resuscitation · Geographic information systems · Spatial statistics · Public health interventions

1 Introduction

Approximately 420,000 out-of-hospital cardiac arrests (OHCA) occur annually in the United States [1]. Research has found that the survival rate of OHCA may be related to the location of the OHCA event [2, 3]. One of the important factors in the survival of OHCA is the presence of witness and timely administration of cardiopulmonary resuscitation (CPR). However, not every OHCA patient receive CPR or has witness at the time of cardiac arrest [4]; such incidences are rare. The guidelines on CPR chain of survival of the American Heart Association stressed the importance of timely administration of CPR. A study by Boller [5] in 2013 stated that approximately 80 % of OHCA and 55 % of intra-hospital cardiac arrest (IHCA) adult patients have not regained ROSC (return of spontaneous circulation) after receiving CPR. The age makeup of the OHCA patients also varied from young to old, with older patients commonly accompanied by various chronic diseases like cardiovascular disease, high blood pressure and diabetes [6].

There are four major types of emergency medical service cases in Taiwan: motor vehicle accidents, acute illnesses, mental illnesses and roadside collapses. The cases of acute illness and roadside collapse include many incidences of acute cardiovascular disease (CVD), which occupies the number 2 position of the 10 major causes of deaths in Taiwan; on average one life is lost every 31 min and 50 s due to CVD [7]. The major concerns of a CVD patient are delayed medical attention and failure to save a life during the critical moments after the onset of a heart attack. Studies have shown that the survival rate of a cardiac arrest patient is significantly improved if the patient receives basic CPR within 4 min, or advanced CPR within 8 minutes [8]. In other words, this has grave implication for the quality of medical service and leaves much room for improvement.

Although there are geographical variations in OHCA survival rates between cities, they also exist variations in populations [9]. The prevalence of bystander CPR also appears to aggregate with in cities [10]. Using geographic information

system (GIS) and spatial cluster analysis, districts identified as “high risk” can be defined as having higher than expected OHCA prevalence, and lower prevalence of the corresponding bystander CPR. The function of identifying these communities is to maximize public health resources through tailoring CPR training and cardiac arrest education programs to communities most in need [11].

There are various spatial analysis methods to test for clusters or districts with statistically significant auto-correlation (e.g., spatial scan statistic [12] and Kernel Density [13]). Each method has its own advantages and limitations. However, there is currently no consensus on how to identify these high-risk districts. The main objective of this study is to propose an integrated method to define districts with high risks of OHCA through statistical approaches and spatial analysis, treatment of OHCA patients, their survivability diminishes if they cannot receive emergency first aid services in time.

However, due to the changes of the social structure in Taiwan, there is great diversity of student characteristics. Especially, there is an educationally disadvantaged generation gradually emergence, such as children of new inhabitants, children of single parent, children from grandparents raising grandchildren and so forth. If we can implant some homogeneous average strategy during the normal class grouping process, not only these students can be placed suitably but also the teachers can evenly share the teaching and counseling loads.

In this paper, we investigate on the normal class grouping operation of elementary and junior high school. Furthermore, we propose a two-dimensional normal class grouping model which can be put into practice.

The rest of this paper is organized as follows. In Sect. 2 we will review the necessary background material on the normal class grouping practiced in Taiwan. In Sect. 3 we will propose a prospective idea about extending the normal class grouping process which is much suitable for the actual educational situation in Taiwan. Based on the prospective idea, we will build a model called “Two-dimensional Normal Class Grouping”, and then we will prove the existence of the solution of the proposed method. Section 4 offers some conclusions and suggestions for future work.

2 Literature Review

DATA

The study analyzed the OHCA data by the Fire Department of the New Taipei City Government from January 1, 2011 to December 31 of the same year. This study was approved by the institutional review board (IRB) of Academia Sinica (IRB#: AS-IRB01-15011). The source of data is from the emergency rescue statistics of the emergency service. The data we used were all stripped of

identifying information and thus informed consent was not needed. There were a total of 2416 OHCA cases in 2011. Eliminating those with unidentifiable addresses, the final samples included 2172 OHCA cases. Data format following the international Utstein-style criteria was collected during a 119 emergency call. This information includes patient demographics (name, age, birth date, gender), event information (event date, time, location, hospital of emergency service and speculated cause of emergency) and first aid factors before admission (bystander witness, time of ambulance arrival, time of onsite treatment, time from site to emergency service, total response time, use of automated external defibrillator (AED), onsite CPR, location type). Also included are first aid factors after admission (AED initial cardiac rhythm, cardiac rhythm before AED shutdown, intubation, drug administration, AED status), prognostic outcomes (return of spontaneous circulation (ROSC), 2-h survival rate after OHCA, cardiac resuscitation before admission) and medical history (past history of cardiovascular diseases, pulmonary diseases, asthma, diabetes, high blood pressure, renal diseases, and cerebrovascular diseases). Based on the aforementioned event descriptions, 28 variables were extracted from the data. A newly added category is first-aid capable hospital, which is classified by the regulations on emergency medical service capability accreditation system of the Ministry of Health and Welfare, Taiwan. The 29 variables were assessed based on 2-h survivability. All OHCA events contained address data as geocoding by the Ministry of Interior Affairs' household number registry (geographical coordinates and graphs) and the Social Economic Database. Geographical encoding of the OHCA events were supplemented with corresponding demographic data (population status, population structure, population characteristics, population changes, basic statistical area) from the 2010 Population Census, social economic data and the geographical borders of the 2010 Population Census.

Study Area

The New Taipei City, Taiwan is situated in the Taipei Basin and is part of the greater Taipei metropolitan area along with Taipei and Keelung City, as well as being one of the largest municipalities in Taiwan. The city has a coastline of about 120 km and contains various terrain features like mountains and hills. The New Taipei City has an estimated population of 3,967,571 and an area of 2052 km². The National Geographic Information System categorized geographical data based on street addresses and geographical coordinates like latitudes and longitudes. The basic geographic mapping was classified to "basic statistical area", "1st level dissemination area", "2nd level dissemination area" and "3rd level dissemination area". The "3rd level dissemination areas" in New Taipei City are made up of 1032 basic statistical areas; a "statistical area" is used as surrogate unit for analysis of communities at the community and neighborhood level as it has been designed to represent about 300–13,000 homogenous social and economic group entities.

Statistical Analysis

High-Risk Neighborhood Identification

The research area is in New Taipei City. Global and local spatial analyses were conducted with spatial autocorrelation analysis (SAA) on statistical area-defined spatial units. To test whether the aggregated statistical areas of OHCA cases in New Taipei City have similar number of values with its neighboring districts, i.e. whether the spatial aggregation of characteristics confer significance to hotspot maps, Global Moran’s I test was used to test for the global spatial autocorrelation. The specific local spatial clusters were analyzed by using Local Moran’s I and Getis-Ord G_i^* indicators, identified where high or low values cluster spatially.

Global Moran’s I [14]

Global Moran’s I is an indicator commonly used to calculate global SAA. Briefly, this indicator has its root in the covariance concept in statistics and is calculated by spatial matrix (W_{ij}). The definition is as follows:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, i \neq j$$

- n number of spatial unit;
- x_i value of variable in each unit;
- \bar{x} mean value of variable in each unit;
- W_{ij} value of weights between the neighboring units;

When the variable is OHCA incidence rate (shown as x), n is the number of OHCA samples; x_i is the total OHCA incidence rates of the spatial unit i ; x_j is the ratio of j ; w_{ij} is an n -order symmetric matrix composed of 0 and 1 and is used to represent the neighboring relationship of all spatial units. If there is neighboring relationship between spatial units i and j , then the corresponding w_{ij} value is 1; if not, then the value is 0. The relation between w_{ii} and w_{jj} is also set as 0. The value of Moran’s I calculated with the above formula will fall between -1 and 1 . When I is larger than 0, it indicates the corresponding district and neighboring districts have similar high values. If I is less than 0, then there is large attribute difference between neighboring districts, and the spatial distribution is spaced between high and low attributes. When I approaches 0, then the correlations among neighboring spatial units are low, meaning there is randomly distribution of high and low OHCA incidences. The Monte Carlo Significance Test converts statistical values into Z-score for testing significance. At 5 % significance level or when Z-score is at 1.96, it means the spatial distribution has a significant positive correlation. When

Z-score is between 1.96 and -1.96, then the spatial distribution is insignificant. When $Z(I) \leq -1.96$, then the spatial distribution has a significant negative correlation.

Local Moran’s I [14]

Each matrix relationship can obtain a Moran’s I, which indicates the level of spatial autocorrelation of the supportive rate under the influence of this matrix relationship. To represent the local spatial distribution of the attributes as either “clustering”, “disseminated” or “randomly distributed” in Moran’s I, or to understand the correlation between each individual district and its neighboring districts, then the Local Moran’s I can be calculated:

$$I_i = \frac{X_i - \bar{x}}{\sum_{i=1}^n (X_i - \bar{x})^2} \sum_{j=1}^n W_{ij}(X_j - \bar{x})$$

- n number of spatial unit;
- X_i value of variable in each unit;
- \bar{x} mean value of variable in each unit;
- W_{ij} value of weights between the neighboring units;

Each I_i is part of the global spatial autocorrelated Moran’s I, and the aggregation of all I_i is the value of Moran’s I. The more influence I_i exerted on I, then the autocorrelation of the district i would be either high or low. The significance test calculates the $Z[I_i]$ value. When $Z[I_i]$ is larger than 1.96, under significance level of $\alpha = 0.05$, the autocorrelation for the spatial unit is positive and significant, meaning the district is surrounding by districts with similar high attributes. This is known as high-high cluster (hot spot); when the observed value of the district and its neighboring areas are low, it is represented as a low-low cluster (cold spot). Additionally, if $Z[I_i]$ is between -1.96 and 1.96, then it would represent either a high-low or low-high value with no spatial autocorrelation”.

Getis-Ord G_i^* [15]

The hot spot analysis of mapping clusters was calculated by the Getis-Ord G_i^* spatial statistic tool. First, the OHCA sites were allocated to 3rd level statistical areas in New Taipei City through the GIS. The incidence of each statistical area was then divided by the population of each statistical area to obtain the incidence rate of the areas. The rate is then substituted into the Getis-Ord G_i^* formula. If the incidence rate of an area and its neighboring districts are similarly elevated, then the Getis-Ord G_i^* value of such area will be high. Getis-Ord G_i^* value represents the degree of spatial clustering with statistical significance; a higher Getis-Ord G_i^* value means the statistical area is a cluster of OHCA events, or an OHCA hot spot.

In G_i^* statistics with a study area with n spatial units, a point in each of the spatial unit i (usually the center of the spatial unit) represents the value x_i of a

random variable X . If a radius of distance d is given and a circle is drawn from the center of each unit i with the radius distance d , then the other spatial units j included in the point of each circle represent the neighboring spatial units of unit i . With this an n amount of subunits can be defined; G_i^* statistic can measure the degree of spatial clustering of the random variable X in one subunit relative to other X values of other subunits. The formula of G_i^* is as follows:

$$G_i^* (d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j}$$

- d distance.
- W_{ij} weighted matrix; when j is within a circle of radius d and center i , then $W_{ij} = 1$; when j is outside the circle of radius d and center i , then $W_{ij} = 0$.
- X_i the value of X at position i and j .
- and X_j

Inferential Statistic
Chi-square analysis

The chi-square analysis is used to test the distribution between two independent categorical variables; however, it does not measure the intensity of the correlation. Chi-square analysis is divided into two parts. The first is the test of statistical significance on variables such as demographics (age and gender), even status (date and time), first-aid factor before admission (bystander witness, EMT-P dispatched for OHCA, time of ambulance arrival, time of onsite treatment, time of arrival at emergency hospital, AED usage, CPR performed on site, types of location), first-aid factors after admission (AED initial cardiac rhythm, cardiac rhythm at AED shut down, intubation, drug administration, AED usage), and prognostic outcomes (ROSC, cardiac rhythm restored before admission, first-aid capable hospital) on 2-h survival rates of OHCA patients was tested. Secondly, the statistical significance of variables on past medical histories (past history of cardiovascular diseases, pulmonary disease, asthma, diabetes, high blood pressure, renal disease and cerebrovascular diseases) on 2-h survival rates of OHCA patient was tested.

Logistic regression analysis

Similar to linear regression analysis, logistic regression is mainly used to analyze binary or multinomial discrete data. A binary data means there are only two possible outcomes from an experiment (e.g. survival or death, 0 or 1). Rather than directly predicting the two possible outcomes, the goal of logistic regression is to establish the simplest and best suitable model; when the dependent variable is a binary data type, it is suitable to use logistic regression. In epidemiological studies, logistic regression is frequently used to determine the relationship between hazardous factors and diseases. The data type of hazardous factors (independent variables) maybe either discrete or continuous.

Data Mining and Decision Tree Analysis

Decision tree is a common analysis structure for supervised data mining. The establishment of a decision tree utilizes the data attribute that is best used to classify the attribute of training data set. The initial tree will be built from model subsets from the training data. These subsets will be used to make the decision tree; the remaining models of the training data set can be used to test the accuracy of the decision tree. If the tree correctly classifies the model cases, then the program will terminate. If there is an error in the model classification, then the model case will be put into the subset of the selected training model cases to start another tree. This process will continue until a tree is able to accurately classify all model cases that have not been selected [16].

3 Results

Study Population

Frequency allocation on the variables in this study showed that most OHCA events occurred during spring and winter seasons; most incidents occurred between day-time hours, with most cases occurring between 7 and 10 am, followed by 15:00–18:00 pm. Male was the predominant gender amongst the cases; subjects were aged between 18 and 65 year. OHCA occurred mostly with internal medical illness; emergency category was predominantly acute illnesses; EMT-P personnel were dispatched for OHCA patients; the level of first-aid capable hospital was normal. The most prevalent past medical history was cardiovascular diseases; events occurred mostly in home residences. The response time from dispatch to the site, from arrival on site to leaving, and from leaving site to hospital admission was mostly completed within 12 min. About 43.4 % of patients received onsite CPR, 15.6 % patients received intubation, 18.4 % patients were given medicine onsite, 2.7 % patients regained ROSC on site, 21.5 % patients used AED onsite, and 10.5 % of OHCA patients had witnesses or bystanders on site. 6.8 % subjects exhibited Ventricular Tachycardia and Ventricular Fibrillation (VT/VF) at initial cardiac rhythm before AED, 1.6 % subjects exhibited VT/VF in cardiac rhythm before AED shutdown and 4.6 % patients regained cardiac rhythm before admission. 95.3 % of OHCA patients had past history of CVD; OHCA occurs mainly in household residences.

Analysis of OHCA incidence in statistic areas

The first phase of the statistical analysis was performed on the following variables: demographic of the OHCA areas, event condition, first-aid factors before admission, first-aid factors after admission, prognostic outcomes and past medical history. Univariate analysis was carried out by Pearson Chi-Square Test to test for their correlation; factors with significant ($p < 0.05$) univariate results were treated as

independent variables. Logistic regression was carried out via stepwise method using SPSS statistical software; odds ratio (OR) and their 95 % confidence interval (CI) were used to represent whether the correlation between independent and dependent variables are statistically significant, after all other factors have been accounted for.

The second phase of the analysis utilizes data mining. The CART (Classification and Regression Trees) algorithm is used for classification. A complex decision tree was established and was pruned into optimal size by cross-validation or results from testing clusters. When building a decision tree, predictors can make different predictions to the data. A predictor is chosen for its ability to lower data disorder. Based on the aforementioned characteristics and strengths, the decision tree algorithm is best suited for the needs of our study, and we have selected the CART decision tree algorithm for testing analysis.

The third phase of the analysis combines spatial statistical analysis. In our study, we used Chi-square analysis, regression analysis and decision tree analysis to statistically analyze the OHCA hot spots, and to interpret or predict how environment variables affect OHCA incidences. However, traditional regression mode does not take spatial heterogeneity and spatial auto-correlation into consideration, which resulted in reduced explanatory power of the various variables being affected by spatial variation. Therefore, for the spatial analysis in our study, we used different spatial models to determine the optimal pattern and validate that spatial variation exists between OHCA incidence rate and spatial characteristics of different areas.

Chi-Square Test

The study variables have been re-encoded into categorical variables in order to determine variables with significance. Study by March [17] et al in 2004 mentioned that Chi-square test is useful to test variables with significance, and the accuracy is not lowered, thus we used the Chi-square analysis to select significant variables from within the total variables of the OHCA dataset. The α value was set at 0.05; a 2-h survival period was used to segregate the survivors and ratio of each variable, as well as mortalities and their ratio. The chi-square value and P-value for the variable was calculated; a P -value less or equal than 0.05 denotes statistical significance; a higher chi-square value means the variable has the characteristic of classification discriminatory power. In our study, the important factors related to 2-h survival of OHCA patients included OHCA type, EMT-P dispatched, intubation, drug administration, onsite ROSC, AED usage, bystander witness, AED initial cardiac rhythm, cardiac rhythm recovered before admission, past histories of diabetes and renal diseases. These 11 factors all shown significance in chi-square test, and the results are shown in Table 1.

Table 1 2-h survival period of chi-square significant factor

Variables		2-h survival period		P-value
		NO N = 1652 (%)	YES N = 520 (%)	
OHCA type	Medicine	1523 (92.2%)	493 (94.8%)	0.025
	Surgery	129 (7.8%)	27 (5.2%)	
EMT-P dispatch	YES	934 (56.5%)	342 (65.8%)	0.000
	No	718 (43.5%)	178 (34.2%)	
Intubation	YES	234 (14.2%)	106 (20.4%)	0.001
	No	1418 (85.8%)	414 (79.6%)	
Drug administration	YES	288 (17.4%)	112 (21.5%)	0.022
	No	1364 (82.6%)	408 (78.5%)	
Onsite ROSC	YES	6 (0.4%)	53 (10.2%)	0.000
	No	1646 (99.6%)	467 (89.8%)	
AED usage	YES	213 (12.9%)	96 (18.5%)	0.001
	No	1439 (87.1%)	424 (81.5%)	
Bystander witness	YES	124 (7.5%)	104 (20.0%)	0.000
	No	1528 (92.5%)	416 (80.0%)	
AED initial cardiac rhythm	Unable to determine	1556 (94.2%)	467 (89.8%)	0.001
	VT/VF	96 (5.8%)	53 (15.2%)	
Cardiac rhythm recovered before admission	YES	21 (1.3%)	79 (84.8%)	0.000
	No	1631 (98.7%)	441 (21.3%)	
Past histories of diabetes	YES	354 (21.4%)	136 (26.2%)	0.015
	No	1298 (78.6%)	384 (73.8%)	
Past histories of renal diseases	YES	130 (7.9%)	69 (13.3%)	0.000
	No	1522 (92.1%)	451 (86.7%)	

Logistic Regression analysis

An important task before data mining is feature selection. The purpose of which is to remove unrelated attributes and to determine key attributes, increasing the accuracy of classification. Common feature selection method is through information

Table 2 2-h survival period of logistic regression factor (*P*-value ≤ 0.2)

Variables	<i>P</i> -value	OR	95% CI	
OHCA type	0.200	0.745	0.475	1.170
EMT-P dispatch	0.038	1.284	1.014	1.626
Onsite ROSC	0.000	10.296	4.045	26.207
Bystander witness	0.000	2.593	1.915	3.510
Cardiac rhythm recovered before admission	0.000	6.072	3.478	10.601
Past histories of diabetes	0.102	1.231	0.960	1.579
Past histories of renal diseases	0.072	1.377	0.972	1.953

gain, mutual information (MI) and logistic regression. We selected a multiple regression analysis with P value less than 0.2 to determine factors related to OHCA incidence. The estimated 95 % CI of the adjusted odds ratio (AOR) was excellent [18, 19].

Using logistic regression can also calculate the statistical test of all variables, with result similar to the previously mentioned chi-square test. The advantage is that statistical analysis can be performed whether the variable is categorical or continuous. With logistic regression, we obtained a P value less than 0.2 for the following variables: OHCA type, EMT-P dispatched, onsite ROSC, bystander witness, cardiac rhythm recovered before admission, past histories of diabetes and renal diseases), indicating that these variables are related with 2-h survival of OHCA patients. The results are shown in Table 2.

Decision Tree analysis

We used the Decision Tree C5.0 algorithm in our study. Chi-square analysis was used to determine factors related to OHCA and to use them as research variables; logistic regression was then performed to find the important variables within the related factors. The C5.0 algorithm was then used to build a pattern amongst important factors and to look for pattern in the data. A complex decision tree was established and pruned into optimal size via cross-validation or results from testing of clusters. When building a decision tree, predictors can make different predictions to the data. A predictor is chosen for its ability to lower data disorder. Based on the aforementioned characteristics and strengths, the decision tree algorithm is best suited for the needs of our study, and we have selected the CART decision tree algorithm for testing analysis. The results are shown in Fig. 1a, b.

Spatial Statistical Analysis

Global Moran's Index

In this study, we investigated whether there are spatial clusters of OHCA occurrences within New Taipei City. Through analysis of Global Moran's Index, we calculated the Moran's Index to be 0.025861, indicating a positive spatial correlation. A P -value of <0.001 indicated statistical significance for spatial auto-correlation of these data. A Z-score of 8.178045 indicated high spatial clustering of attributes within the areas (as shown in Fig. 2), denoting high spatial clustering phenomenon in incidence sites within New Taipei City.

Local Moran's Index

After obtaining the result with spatial clustering forms in New Taipei City from the Global Moran's Index, we conducted analysis of Local Moran's Index to analyze SAA and calculate hot spots with Local Moran's Index. Significant results are shown in four types. HH indicated higher incidence in the area and neighboring areas (black areas). LL indicated lower incidence in the area and surrounding areas (blue areas). HL indicated high incidence in the area and low incidences in the

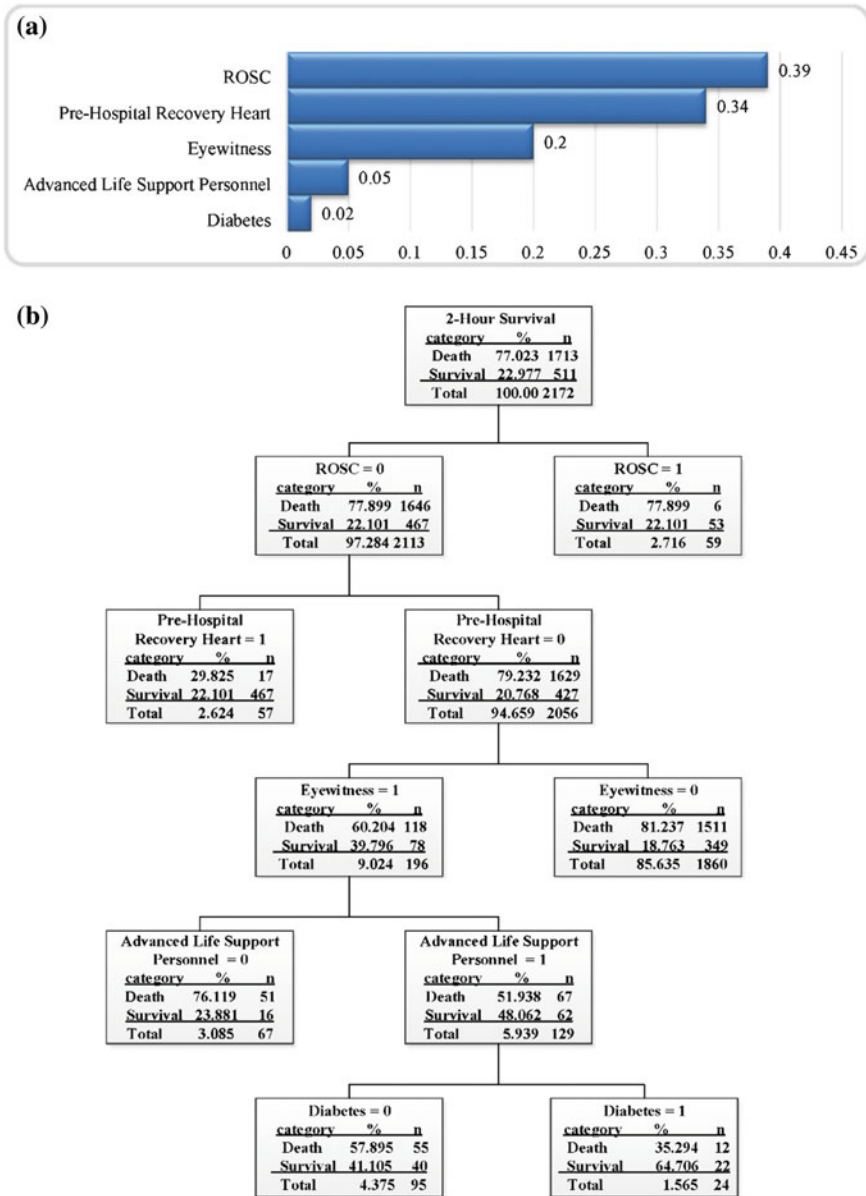


Fig. 1 Correlation analysis of decision tree. **a** OHCA of important factors decision trees. **b** OHCA of decision trees dendrogram

surrounding areas (yellow areas) and LH indicated low incidence in the area and high incidences in the surrounding areas (white areas). Results showed that New Taipei City has the highest incidence areas (shown in Fig. 3a).

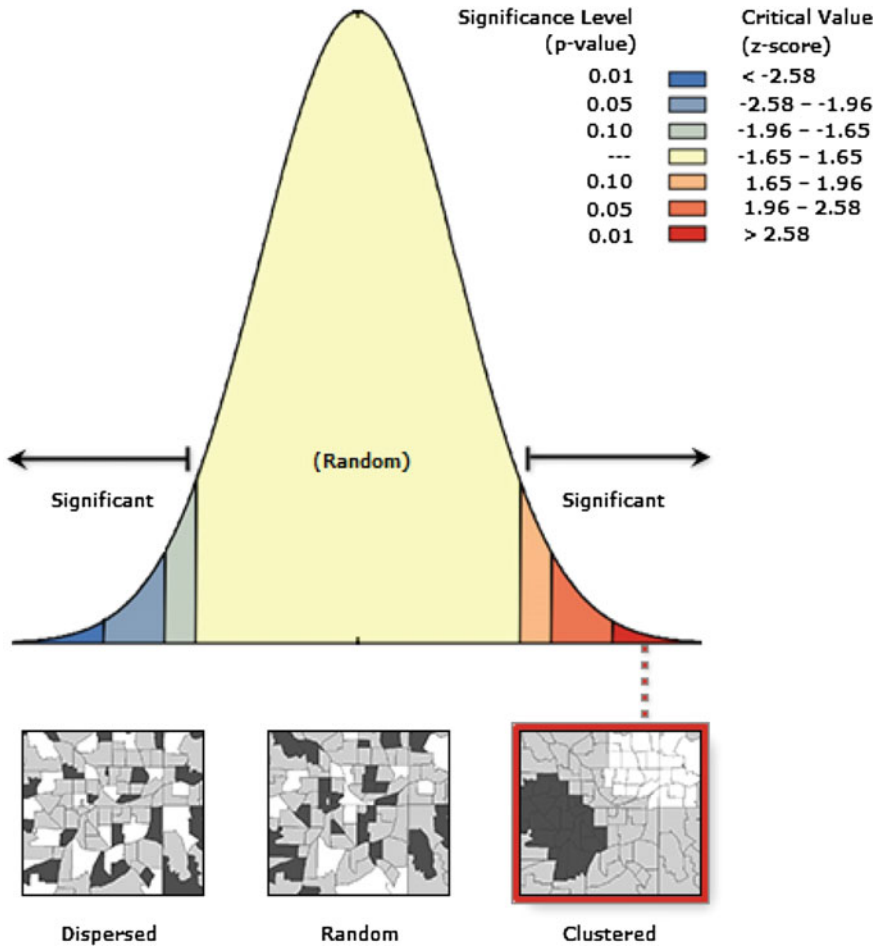


Fig. 2 Global Moran's index coefficient

Getis-Ord Gi*

The study used Getis-Ord Gi* as the indicator for hotspot analysis. The Gi* value was divided into 7 parts to represent various degree of significance. A hot spot denotes significant clustering of high values; a cold spot denotes significant clustering of low values. In this study, the OHCA frequency was treated as attribute value; the results from the hotspot analysis are shown in Fig. 3b. In our use of two different spatial statistical methods on spatial feature analysis, regardless of the greater difference between of the analysis results, there were more populated areas (cities and counties) covered by Getis-Ord Gi*. However, HH areas with statistical significance as identified by Local Moran's Index existed also in High Clusters

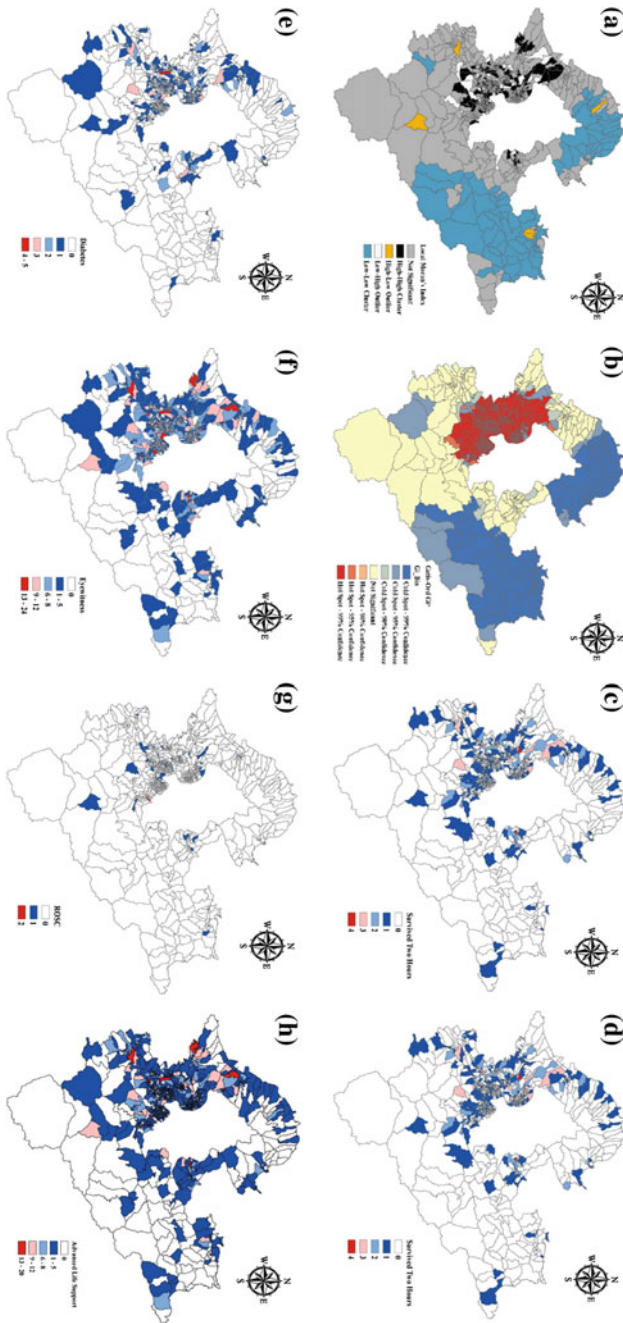


Fig. 3 Cardiac arrest incidence using Getis-Ord, local Moran’s I and spatial number distribution. **a** Local Moran’s index coefficient; **b** Getis-Ord G_i^* coefficient; **c** Survived two hours of spatial number distribution; **d** Diabetes of spatial number distribution; **e** Eyewitness of spatial number distribution; **f** ROSC of spatial number distribution; **g** Pre-hospital recovery heart of spatial number distribution; **h** Advanced life support of spatial number distribution

areas determined by Getis-Ord G_i^* (as shown in Fig. 3a, b). The coverage of important factors in the spatial distribution charts (Fig. 3c–h) was also higher than decision tree method.

4 Conclusions

According to our results, the OHCA first-aid factors tested by the chi-square analysis—OHCA type, EMT-P dispatched, intubation, drug administration, onsite ROSC, AED usage, bystander witness, AED initial cardiac rhythm, cardiac rhythm recovered before admission, past histories of diabetes and renal diseases, even though they could not be changed, they could still be preventable. For example, when a patient's disease severity reaches a certain degree, the medical staff could be alerted immediately to transfer the patient to the ICU, which not only allows first-aid CPR procedure but also reduces time of emergency treatment. We discovered that intubation and cardiac agents (inotropes) are related to the outcomes of emergency service; studies have shown that use of cardiac inotropes could improve the outcomes of CPR, making it a viable option for emergency treatment and drug selection to reduce risk of CPR failure [20–23].

This is a first study that experiments on a new summary method to identify areas with high OHCA risks. The method combined the advantages of three independent spatial clustering analyses to determine OHCA hot spots. Using New Taipei City as the scope of the study, we used the Global Moran's Index to determine spatial clustering phenomenon, then analyzed the spatial distribution of OHCA hot spots with Local Moran's Index and the Getis-Ord G_i^* SAA algorithm. For statistical analysis of OHCA incidence areas, we established a multi-variate linear regression model with emergency statistical data from fire and emergency services as variables. We determined the location factors of OHCA and investigated the special characteristics of OHCA hot spots. We compared between traditional regression model and spatial mode to find the most suitable spatial pattern, and validated that spatial variation existed between OHCA incidence rate and spatial characteristics of districts [10, 24].

Our study used Chi-square analysis, logistic regression and decision tree analysis to determine whether the following factors are important to the survival of OHCA: onsite ROSC, EMT-P dispatched, presence of bystander witness and past history of diabetes. Together with geospatial analysis, we determined that within a hotspot, 98 and 22 % of OHCA patients with onsite ROSC ($N = 53$) and cardiac rhythm regained before admission ($N = 79$) had past histories of cardiac diseases and diabetes, respectively. We found that survival of OHCA patients may increase if the 119 emergency dispatcher had made detailed inquiries to the patient on past histories of cardiac diseases or diabetes, or when EMT-P were dispatched. From the OHCA hot spots generated in Fig. 3b, public health institutions may improve the survival of OHCA patients by introducing intervention measures such as using information technology (e.g. wearable technology) for patients with past histories of cardiovascular or diabetes, or to increase the presence of bystander witness [25, 26].

Acknowledgement This research was supported by grant entitled “Multidisciplinary Health Cloud Research Program: Technology Development and Application of Big Health Data” from the Academia Sinica. We would also like to express our sincere gratitude to Mr. Kent M. Suárez for his English editing.

Competing Interests

The authors declare that we do not have any competing interests related to this study.

References

1. Go AS, Mozaffarian D, Roger VL et al (2014) Heart disease and stroke statistics–2014 update: a report from the American Heart Association. *Circulation* 129(3):e28
2. Nichol G, Thomas E, Callaway CW et al (2008) Regional variation in out-of-hospital cardiac arrest incidence and outcome. *JAMA* 300(12):1423–1431
3. Sasson C, Rogers MA, Dahl J et al (2010) Predictors of survival from out-of-hospital cardiac arrest: a systematic review and meta-analysis. *Circ Cardiovasc Qual Outcomes* 3(1):63–81
4. McNally B, Valderrama AL (2011) Out-of-hospital cardiac arrest surveillance: Cardiac Arrest Registry to Enhance Survival (CARES), United States, Oct 1, 2005–Dec 31, 2010
5. Manuel B (2013) Will models of naturally occurring disease in animals reduce the bench-to-bedside gap in biomedical research? *Zhonghua wei zhong bing ji jiu yi xue* 25(1):5–7
6. Lee C-K (2010) The Analysis of Ilan’s Out-of-Hospital Cardiac Arrest (OHCA) Patients
7. China MoHaWRO. The cause of death statistics. Secondary The cause of death statistics. http://www.mohw.gov.tw/cht/DOS/Statistic.aspx?f_list_no=312. Access date, 2015/05/18
8. Valenzuela TD, Roe DJ, Cretin S et al (1997) Estimating effectiveness of cardiac arrest interventions: a logistic regression survival model. *Circulation* 96(10):3308–3313
9. Sasson C, Keirns CC, Smith D et al (2010) Small area variations in out-of-hospital cardiac arrest: does the neighborhood matter? *Ann Internal Med* 153(1):19–22
10. Root ED, Gonzales L, Persse DE et al (2013) A tale of two cities: the role of neighborhood socioeconomic status in spatial clustering of bystander CPR in Austin and Houston. *Resuscitation* 84(6):752–759
11. Sasson C, Meischke H, Abella BS et al (2013) Increasing cardiopulmonary resuscitation provision in communities with low bystander cardiopulmonary resuscitation rates: a science advisory from the American Heart Association for healthcare providers, policymakers, public health departments, and community leaders. *Circulation* 127(12):1342–1350
12. Kulldorff M (1997) A spatial scan statistic. *Commun Stat Theory Methods* 26(6):1481–1496
13. Waller LA, Gotway CA (2004) *Applied spatial statistics for public health data*. Wiley, New York
14. Anselin L (1995) Local indicators of spatial association-LISA. *Geograph Anal* 27(2):93–115
15. Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geograph Anal* 24(3):189–206
16. Geatz MW, Roiger R (2011) *Data mining: a tutorial based primer*. Pearson Education, London
17. Automated icd9-cm coding employing bayesian machine learning: a preliminary exploration. *Simpósio de Informática y Salud*; 2004
18. Hosmer DW Jr, Lemeshow S, Sturdivant RX (2000) *Model-building strategies and methods for logistic regression*. Third Edition, Applied Logistic Regression, pp 89–151
19. Chan T-C, Fu Y-c, Wang D-W, et al (2014) Determinants of receiving the pandemic (H1N1) 2009 vaccine and intention to receive the seasonal influenza vaccine in Taiwan

20. Gardner LS, Nguyen-Pham S, Greenslade JH et al (2014) Admission glycaemia and its association with acute coronary syndrome in Emergency Department patients with chest pain. *Emerg Med J* emermed-2014-204046
21. Tandon N, McCarthy M, Forehand B et al (2013) Comparison of intubation modalities in a simulated cardiac arrest with uninterrupted chest compressions. *Emerg Med J* emermed-2013-202783
22. Chang AM, Edwards M, Matsuura AC et al (2013) Relationship between renal dysfunction and outcomes in emergency department patients with potential acute coronary syndromes. *Emerg Med J* 30(2):101–105
23. Henry K, Murphy A, Willis D et al (2012) Out-of-hospital cardiac arrest in Cork, Ireland. *Emerg Med J* emermed-2011-200888
24. Ong MEH, Wah W, Hsu LY et al (2014) Geographic factors are associated with increased risk for out-of hospital cardiac arrests and provision of bystander cardio-pulmonary resuscitation in Singapore. *Resuscitation* 85(9):1153–1160
25. Lam SSW, Zhang J, Zhang ZC et al (2015) Dynamic ambulance reallocation for the reduction of ambulance response times using system status management. *Am J Emerg Med* 33(2):159–166
26. Nhavoto JA, Grönlund Å (2014) Mobile technologies and geographic information systems to improve health care systems: a literature review. *JMIR mHealth and uHealth* 2(2)

Social Event Detection and Analysis Using Social Event Radar

Jin-Gu Pan and Ping-I Chen

Abstract This research describes the social event collection and analysis system developed by Institute for Information Industry. The system can collect more than 30,000 web data items per day for the government to understand public opinion on policy and for companies needing business insights or seeking to provide exposure for their brands on the Internet.

1 Introduction

According to Facebook usage statistics, there are more than 15 million active users per month in Taiwan, and 11 million users log in every day. The interaction between users and Facebook Pages can yield important information for government and companies seeking to discover what users want, what they are thinking, and their opinions on public issues and company brands. Besides Facebook, other venues such as bulletin board systems, forums, blogs, and news websites also provide a great opportunity for researchers to mine and analyze important clues for risk management and marketing use [9].

The most basic functions of social media analysis include counting the number of likes, shares, and comments on each topic and determining whether discussions are trending higher. Users can employ single or multiple keywords to represent topics of interest, and a typical analytics platform will generate statistics or sentiment charts for advanced analysis. These methods can be deployed on the cloud platform, using an API or a web crawler to gather and store more than one hundred million data items. However, as the amount of data increases, the traditional SQL query method becomes inefficient and error-prone. Thus, building a social media

J.-G. Pan (✉) · P.-I. Chen
Innovative DigiTech-Enabled Application and Service Institute,
Institute for Information Industry, Taipei, Taiwan
e-mail: jgpan@iii.org.tw

P.-I. Chen
e-mail: be@iii.org.tw

analysis platform requires a lot of financial support, a variety of techniques, and cooperation of domain experts working together.

Institute for Information Industry is a government-subsidized nonprofit organization that focuses on new technology development and research. About five years ago, we began designing a platform capable of collecting both social and non-social media unstructured information to provide information services for government and companies. Now, the amount of data has grown substantially, and we not only have acquired experience in NLP, big data storage, and other techniques but also provide analysis of this data as a service for risk management and social marketing for companies [4, 7].

In this study, we introduce our system, especially the sentiment analysis mechanism, and we provide a complete picture of this social media analysis platform. In our research, we worked to develop our techniques based on customers' real world requirements. Many customers' requirements are based on their experience in using commercial products. Therefore, our challenge is finding solutions to their needs and providing accurate analysis for their data. In some cases, we also can provide source code and knowledge to software companies to enhance their systems. Perhaps one day, we may publish the complete system, allowing researchers to use the data and share their algorithms in the platform so that it will become a major data ecosystem.

2 Related Works

2.1 *Social Media Data Collection*

A crawler is a type of software application that can systematically browse a target web page and extract specific data from the page. It is the basis of social network analysis (SNA) research. Most of the work can be done by sending an http request to a website and using an HTML parser to obtain the information. However, one limitation of the web crawler is its crawling speed. Moreover, most websites regard such crawlers as a threat and ban IP addresses that make too many requests or make requests too often. Various workarounds for dealing with this problem are possible, such as changing the crawler's IP address or issuing requests at random time intervals.

Apache Nutch is an open source web crawler that can be used to crawl a large number of target websites and directly index the data by Solr or Lucene search engine [5]. We can use it to monitor updated information and crawl that new information automatically. However, we still need to consider the request time and speed. Once all of the information has been gathered from the social network, we can start to evaluate the friend relations and provide related insights to clients.

The Facebook Graph API provides an easy way for data scientists and companies to obtain public interaction information [3]. We do not need to develop a

web crawler and resort to various workarounds to collect data from the website anymore. Data gathered using the API is also clearer and allows our research to concentrate more on social network analysis.

2.2 Chinese Word Segmentation

English articles can be tokenized by using punctuation; however, there is no punctuation in a Chinese sentence. The most famous Chinese word segmentation tool is the CKIP, which was built by Academia Sinica [2]. We can send articles to tokenize using the tool's API. The only problem is that the response time of the API is not immediate. Mmseg4j is another open source Chinese segmentation tool. It can not only use local word dictionaries to segment sentences but also easily combine them into the Solr or Lucene search engine.

Because our database is too large, we are not able to use SQL queries to find the desired information. Thus, we use Solr and Elasticsearch to index all of the data, which solves the problem. When the data comes in, mmseg4j segments sentences into words and indexes or updates the frequency of the words in the search engine. In our experience, each search engine's core can handle about two hundred million data items. In addition, the indexing time and method require more attention to deal with the huge amount of daily data collected.

2.3 Opinion Mining

Opinion mining refers to use of the text analysis technique to identify positive and negative emotions of an article's author. Generally, we can use a sentiment word dictionary that identifies each keyword and its degree of emotion. Turney proposed a method that considers a phrase as having a positive semantic orientation when it has good associations and a negative semantic orientation when it has bad associations. Then, the PMI-IR algorithm is used to estimate the semantic orientation of the extracted phrases. The PMI-IR algorithm was originally designed to measure the similarity of pairs of words. Turney [8] calculated the semantic orientation as follows, and the score of SO will be positive when a phrase is more strongly associated with "excellent."

$$SO(Phrase) = PMI(Phrase, "excellent") - PMI(Phrase, "poor") \quad (1)$$

Pang et al. [6] proposed a method that tried to infer an author's implied numerical rating. Their meta-algorithm can provide significant improvements over the traditional SVM classifier. Basari et al. proposed an algorithm that uses particle

swarm optimization (PSO) to enhance the accuracy of the SVM. PSO can control the selection of possible subsets using the maximum suitable particle to give rise to the next creation of n-candidate particle to achieve optimal forecast accuracy, thus making the SVM results more accurate than before [1].

In this research, we assume it will not be easy to extract the author's opinion by using the supervised or unsupervised learning algorithm directly. One main reason is that there are no sufficient or obviously positive or negative words in Chinese. Many words used in different situations will have different meanings. Likewise, an author of social media content sometimes will define new words to represent the meaning of the discussion. Therefore, we tried to find a new method that uses social tagging to train for a period of time, analyzing topics and using the resulting information to classify the emotion of the article.

3 System Design

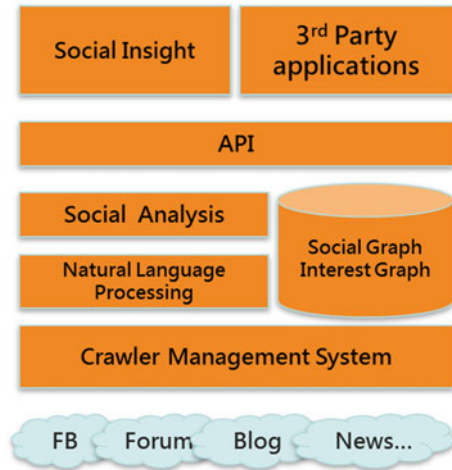
In this section, we introduce our data collection system and the topic tracking system. In addition, based on the properties of public issues in social media data, we use a simple mechanism for topic sentiment analysis. We introduce the reason and findings in this section.

3.1 *SER Data Collection*

SER (Social Event Radar) is a solution for social media analysis that allows collection of social event information 24/7. The database has been collecting this data since the year 2012; hence, information can be extracted and re-examined by using the SER platform. The data sources include all major news websites, important Facebook Pages, PTT (Taiwan's best-known bulletin board system), blogs, and Flickr. All of the data has been processed in a manner that identifies the author, the data source, and how many people like or share each item. The resulting data is then stored in a well-formed structure.

During the past few years, we have focused on the web crawling technique and tried several ways to obtain and store the required information efficiently. Thus, we not only provide the data to the customer, but we also provide the system's architecture and performance tuning experience to several companies. Our database contains approximately 10 terabytes of data, and each month an additional 300 gigabytes of data are added. Our server farm consists of 42 machines for crawling data, 13 machines for data storage, and 6 machines for computing (Fig. 1).

Fig. 1 SER architecture



3.2 Social Event Tracking System

Some companies require weekly or monthly reports to understand product or brand comments collected from the Internet. Our data scientists initially contact the customer management team and determine what type of chart the client wants and what information is needed. After that, on the basis of discussion and our prior experience, the data scientist can associate the tracking topic with its related keyword set. First, we use a testing tool to search using the keyword set. This tool will export an Excel file, allowing the data scientist to examine the accuracy of the outputted data. If there are too many inconsistencies or too much “noise” in the data, we will add or delete some keywords and attempt to devise a new keyword set. This process is repeated until the keyword set theoretically represents the target topic. After that, we configure the topic and its keyword set in the auto tracking system, which, in turn, will provide information about data update status and some basic statistical graphing that will enable users to understand the discussion trend of the topic (Fig. 2).

Our system allows users to assign the time range and data sources to draw the topic trend map. As shown in Fig. 3, we selected the time range from April 15 to April 29, and then determined that we want all kinds of data in our database. The trend map shows there are only a few posts and comments about this topic. Most social media users just skim the topic and click the “Like” button. This means that the topic attracts users’ attention but without too much impact. Our system also provides a multiple topic trend comparison map, allowing data scientists to compare several related topics on the same page.

This system has been in use since 2014, and we employ it to monitor nearly all of the important news topics in Taiwan to help government agencies and companies make decision and minimize risk impacts.

Keyword 趨勢 關鍵字設定

[設定](#) [繪圖](#) [列表](#)

新增關鍵字

新增

更新時間

- 每日07:00更新前一天資料

組名	關鍵字	建立日期	最後更新時間	
0420-1_台股一年新軌9到10起	台股+新軌((軌道)新軌)+新	2015-04-22	2015-06-17	更新
0419-3_特教賦得評講缺失	特教+採購+審計	2015-04-22	2015-06-17	更新
0421-4_軍公教會館違規	軍公教+會館	2015-04-22	2015-06-17	更新
0417-6_霜害重創春茶	霜害+春茶	2015-04-22	2015-06-17	更新
0423-8_高雄將三階限水	限水+(三階)限+(高雄)高市	2015-04-23	2015-06-17	更新
0423-6_RCA員工立法院抗議勞動部	(RCA R C A)+(職災)工僑工類工作傷害)+勞動部	2015-04-23	2015-06-17	更新
0423-4_缺水農產並損農業	(乾旱)缺水 旱災 旱災)+(瓜)果農產 農作 農民 農捐(農業)	2015-04-23	2015-06-17	更新

Fig. 2 Keyword setting for topic tracking system



Fig. 3 Trend analysis and resource selection

3.3 *Sentiment Analysis Module*

In the past few years, much research has already been conducted on sentiment analysis techniques. We tried to use some of these methods in our system to quickly determine whether each user's opinion is positive or negative. However, our system's data sources and topic domains are extensive, making it difficult to construct the sentiment dictionary for each domain. Additionally, in our experience, the life cycle of most keywords related to each topic is no more than two weeks. Consequently, users will construct many new keywords to describe or represent the situation when they engage in discussion.

Many commercial products claim that they can deal with sentiment analysis and provide useful results. However, many of our customers have tried these products before coming to us, and they complain about the results achieved. From the start, we knew that sentiment analysis in social media analytics is not easy, and we made a special effort to ensure that the results are sufficiently accurate for commercial use. Thus, our team uses human intelligence processes to examine each comment and feel what users are thinking to provide sentiment analysis results. This is a very time-consuming job, however, and we finally decided to address this problem.

Our system contains nearly all of the popular social media in our database. We quickly came up with a new idea, utilizing each topic's user tagging information to train and fine-tune the process. We used the results to classify similar topics' comments on other social media and determine whether the sentiment is positive or negative.

The PTT bulletin board system has more than 1.5 million registered users, with over 150,000 users online during peak hours. It has over 20,000 boards covering a multitude of topics, and more than 20,000 articles and 500,000 comments are posted every day. One interesting feature is that users can "Like" or "Dislike" a particular topic when they post a comment.

We assume that when news is breaking, it will be shared to different social media. Therefore, we can use sentiment information from PTT to reliably predict the response on news sites and Facebook. We extract each topic's comments and the corresponding sentiment tags in Fig. 4, where "推" means "Like" and "嘘" means "Dislike" in Chinese. The topic in this example is devoted to discussion of the financial issue of Buddhist Compassion Relief Tzu Chi Foundation, the largest NGO in Taiwan. Using these comments as training data, we can employ the keyword segment and compile the positive comments' keywords into a positive keyword dictionary and vice versa.

We use naive Bayes classifiers to train positive and negative words so that if a similar topic on Facebook and its accompanying comments come in, we can use this simple classifier to decide whether users' comments on this topic are positive or negative. Figure 5 shows an example of the comments' classifying results on Facebook. In our experience, the accuracy of the sentiment analysis results is approximately 60 percent. We believe this is adequate because most of the comments are very short and do not contain too many sentiment words, so it will not be

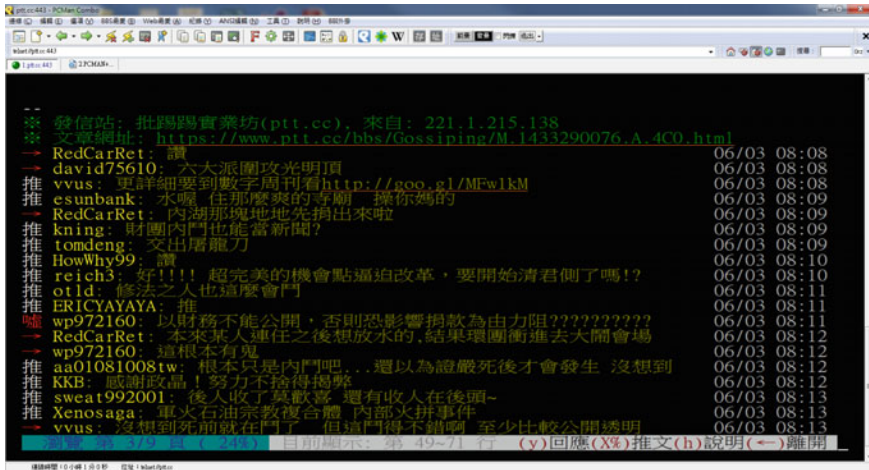


Fig. 4 PTT bulletin board system



Fig. 5 Sentiment analysis results

easy to use the traditional dictionary based method to identify their sentiment. To improve the accuracy of our sentiment analysis system and ensure that the results are reliable, we provide a mechanism for the data scientists to examine the classifying results of the comments and modify the sentiments if something appears to be incorrect as shown in Fig. 6. We plan to hire several domain experts to examine and double-check the results of our system. All of the actions will be

內容	情緒	情緒-正	情緒-負
有陣期的詐騙益善機構	positive	-44.3124697304	-44.3125411599
還不就是欺騙的邪教?	negative	-51.2036829974	-50.2742676784
詐騙集團!	positive	-24.9645050333	-25.9825713032
會不會馬上有人要被吞了?	positive	-51.2191383758	-52.5080905257
佛門之地如此六根不清淨,這就是絕對的魔力。	negative	-120.0468995072	-118.240641428
它們現在是狗咬狗了麼有沒有見骨了	negative	-81.8456099688	-80.8076149474
內鬥了!內鬥了!	positive	-36.8945021402	-37.8537209308
出家人不乾不淨的利益為重 噁心味	negative	-74.8322579152	-73.0005045324
幹的好,必需將那些邪惡貪婪者趕出經濟	negative	-106.070940334	-105.29389464
經濟之權利與慾望第二季	positive	-60.3749905147	-62.9290746566
財務透明了嗎?	positive	-37.4916325646	-38.1156516059
壞蛋大法師人	positive	-37.8120596885	-38.741883777
恐怖啊,恐怖啊這是佛門地方	positive	-63.0413287166	-65.0950056206
公佈財務,不要讓你們黑手。	negative	-75.6742647098	-75.6134395137
通常無賴的善都是小人物,因為他們心裡單一	positive	-103.689403501	-106.217470111
阿彌陀佛慈悲為懷,是凡間人作孽,佛記受罰很重啊!作惡多端的人,下場很悲慘的。	positive	-196.675219653	-196.82550035
缺.....複雜的經濟!	positive	-48.5488901534	-48.9705487176
	positive	-0.308156925641	-1.32727102067
還真的看不懂,霧裡看花,無解,回不去創始的精神了!	positive	-118.646298121	-119.005049897
改革有用嗎,還不是一條要人勇健 錢撈了也不知道是不是真的幫助需要幫助的人 根據了解弱特團體要申請救助幫忙是複雜	positive	-252.9271102416	-255.800825071
啊!對全國宗教財務都要比擬辦理	positive	-70.3880254748	-71.6867634116
解散 財產歸民權黨管理 國民黨監督 這種比較不會有爭議	positive	-100.435695596	-102.702823353

Fig. 6 Comments and sentiment classifier's detail information

recorded and feedback to our classifier will be immediate so that the next related topic examined will yield more accurate results.

4 Conclusions

In this study, we introduce our social event collection and tracking system, which we have used to provide risk management for the government and companies. We also describe our new finding and propose a new mechanism for sentiment analysis. Although there is still much to be done to improve the accuracy of sentiment analysis, our method is not only based on users' sentiment tagging results in social media but also relies on a domain expert to recheck and give feedback to the training system. The only weakness is that if there are not enough comments or related topics, the classifier cannot be implemented correctly.

At present, we identify a problem wherein each topic consists of several discussions. If we want to measure a topic's sentiment results, should we deal with the emotion of the author individually and identify a representative emotion or just merge those discussions to give a complete picture of the topic's sentiment? If we want to find the sentiment results for a discussion, how should we decide the most representative discussion in our dataset? By using our SER data collection system,

we can easily try various ways to research and pinpoint the best solution to the problem. We also want to make it an open platform so that research organizations and companies can participate and use our SER data to continue inventing new techniques on the platform, like the Amazon Mechanical Turk. Companies can send their data and requests to the platform; the data scientist can analyze it and then use the data and algorithms to solve the problem.

Acknowledgments This study is conducted under the “Big Data Technologies and Applications (1/4)” of the Institute for Information Industry, which is subsidized by the Ministry of Economic Affairs of the Republic of China.

References

1. Basaria ASH, Hussina B, Anantaa IGP, Zeniarjab J (2012) Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. In: Malaysian technical universities conference on engineering and technology, Elsevier, Amsterdam
2. CKIP, <http://ckipsvr.iis.sinica.edu.tw/>
3. Facebook Graph API, <https://developers.facebook.com/docs/graph-api>
4. Hoffman DL, Fordor M (2010) Can you measure the ROI of your social media marketing? MIT Sloan Manag Rev 52(1):40–50
5. nutch, <http://nutch.apache.org/>
6. Pang B, Lee LL (2005) Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the association for computational linguistics (ACL), pp 115–124
7. Sohn D (2009) Disentangling the effects of social network density on electronic word-of-mouth (eWOM) intention. J Comput Med Commun 14(2):267–352
8. Turney PD (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the association for computational linguistics, pp 417–424
9. Wilson K, Fornasier S, White KM (2010) Psychological predictors of young adults’ use of social networking sites. Cyberpsychol Behav Soc Netw 13(2):173–177

Social Network and Consumer Behavior Analysis: A Case Study in the Retail Store

Pin-Liang Chen, Ping-Che Yang and Tsun Ku

Abstract The goal of this study was to analyze the characteristics and purchase probability of different customers groups in the retail store POYA. We abstracted CheckMe app user records about the retail store POYA. The dates of those user records were between January 2015 and April 2015. Furthermore, we collected the Facebook information of the users. All statistical procedures were performed with our Persona Analysis Platform. The purchase probability of female subjects was nearly twice than male subjects in POYA. Although, the main subjects were 18–39 years old, the purchase probability of 40–44 years old subjects was higher than others. Most of the customers were from the metropolises. However, the purchase probability of the subjects in Hsinchu and Pingtung was higher than that in other cities. After we provided analysis results to POYA, it improved its promotions and got an up to 10 % conversion rate improvement.

1 Introduction

Traditional advertising and marketing used in the past are focused on the non-specific customers. However, most of them have diverse interests and the effects are not significant. Besides, the increasing popularity of social networking services and the mobile devices have changed people's life. Social network analysis has become more and more important. To overcome the new business challenges, service providers have to improve their marketing strategy, analyze the existing and potential customers, and develop the precision marketing methods. The goal is to offer the right products to the right customers at right time and right location.

P.-L. Chen (✉) · P.-C. Yang · T. Ku
Institute for Information Industry, Taipei, Taiwan, ROC
e-mail: mileschen@iii.org.tw

P.-C. Yang
e-mail: maciaclark@iii.org.tw

T. Ku
e-mail: cujing@iii.org.tw

In the past, there were several studies focused on the analysis of customers [1–6]. Liao et al. [5] found the differences between heavy and non-heavy users in top video apps based on Chunghwa Telecom network connection records. He et al. [3] analyzed the trend of tweets numbers for the big three pizza chains by text mining. Such analysis results can be used to design appropriate marketing plans to improve sales [7]. However, those studies used only social network data or private customer data.

In this study, we combined two types of user records, the social network user information and the customer shopping records. We got the customer shopping records from the online to offline (O2O) corporation, OPEN-LIFE, we cooperated with. It combined the social media marketing and the service of earning reward points. Customers could use its app, CheckMe, and get reward points from completing its assignments, for example, spending over NT\$100 in the designated store. Its clients included POYA, Carrefour, SK-II, etc. In this paper, we used the customer shopping records from POYA, a retail store sells cosmetics and groceries.

The primary goal of this study was to analyze the characteristics and purchase probability of different customers groups in the retail store POYA. All statistical procedures were performed with the Persona Analysis Platform we developed before. The secondary goal was to provide the analysis results for our customers and help them to improve their promotions.

2 Subjects/Materials and Methods

2.1 Data Source

In this section, we describe the data source we used in the paper. We collected the information of the users who installed the CheckMe app. The CheckMe app would give the users two types of assignments. One was to check into the store or scan the barcode of some products; another was to purchase some products in the store. When a user finished one assignment, he/she would be rewarded some reward points and the record would be saved in the database. Besides, the CheckMe app needed the authorization to access user's information on Facebook. Therefore, we had two types of information, the user records on CheckMe app and their information on Facebook.

In the data collection process, we collected the user records of CheckMe app from October 14, 2014 to April 8, 2015. The user records contained user Facebook ID, user email, invite code, shop name, assignment name, execution time, synchronization time, beacon ID, reward points, app platform, invoice information, invoice amount, and status. There were two types of records on CheckMe app according to the assignment name. One was the check-in assignment; another was the shopping assignment.

Furthermore, we collected the Facebook information of CheckMe app users by their Facebook ID. The user information on Facebook contained user Facebook ID,

user email, gender, birthday, city he/she lived in, number of followers, the fan pages' IDs he/she liked, the fan pages' names he/she liked, and the fan pages' categories.

2.2 Study Samples

In the study, we abstracted user records about the retail store POYA. The dates of those user records were between January 29, 2015 and April 8, 2015. Those records covered all the POYA stores in Taiwan. There were twenty-seven kinds of check-in assignments, included checking into the POYA store, scanning the barcode of Mini Elisa, scanning the barcode of KIRIN Bar beers, etc. There were six kinds of shopping assignments, included spending over NT\$100, spending over NT\$150, spending over NT\$300, spending over NT\$400, spending over NT\$800, and spending over NT\$1000.

To avoid the influence of the error records, we deleted the duplicate records or the records with incomplete status.

2.3 Case and Control

In the study, the case definition was different in each analysis. In the analysis of male group, the case group contained all male users and the control group contained the other users, including female users and the users with no gender information. Similarly, in the analysis of Kaohsiung group, the case group contained all users lived in Kaohsiung and the control group contained the other users, including users lived in other cities and the users with no city information. Thus, we could compare the group we wanted to focus on with the other general users and find out if there was any special in the case group.

2.4 Events

The goal of the CheckMe app was to encourage the customers to purchase some products in the store. Therefore, the event was whether the user buy or not in the store. If one user finally bought something in the store, we put him/her into the event group. Otherwise, we put him/her into the non-event group if he/she only checked into the store or scanned barcode.

2.5 *The Analysis Tool*

The Persona Analysis Platform was a social data analysis tool we developed in Institute for Information Industry. It was written in Java, PHP and the statistical procedures were built with the Apache commons mathematics library. The Persona Analysis Platform contained several web crawlers to obtain different heterogeneous social and customer shopping data, including Facebook, Taobao, Tmall, PTT, blogs, web forums, etc. It also included the customer shopping data and beacon data of the corporations we cooperated with.

The Persona Analysis Platform merged different heterogeneous data and provided the analysis results to the corporations we cooperated with, helping them to improve their sales performance. It provided several analysis methods: (1) clustering the users and finding the differences between those user groups. (2) finding the hidden association rules in the shopping behavior. (3) finding the customer shopping periods. (4) finding the different shopping behavior in different shopping mall. Several analysis methods are developing and will be included in the Persona Analysis Platform in the future.

2.6 *Statistical Analysis*

All statistical procedures were performed with our Persona Analysis Platform. We estimated the odds ratio (OR) and 95 % confidence interval (CI) of each case and control group. The *P* value was calculated using the chi-square test. All tests were 2-tailed, and a *P* value of less than 0.05 was considered statistically significant.

3 **Results**

A total of 4208 CheckMe app users who completed any assignments in POYA were included in the case study cohort. Table 1 shows the demographics of CheckMe app users who completed any assignments in POYA. Most of the subjects were female (81.4 %). The main subjects were 18–39 years old (84.6 %). The 22.9 % subjects had one or more followers on Facebook. Most of the subjects (70.2 %) purchased some things in POYA.

Table 2 shows the associations of purchase probability with subjects' gender. 438 of the 752 (58.2 %) male subjects and 2496 of the 3426 (72.9 %) female subjects purchased some things in POYA. The purchase probability of female subjects was nearly twice than male subjects in POYA (OR, 1.89; 95 % CI, 1.61–2.22; *P* < 0.001).

Table 3 shows the associations of purchase probability with subjects' age. The purchase probability of the subjects under 18 years old was the lowest (OR, 0.62;

Table 1 Demographics of CheckMe app users who completed any assignments in POYA*

Characteristics	Subjects (n = 4208)	
Age, median (IQR), year	30	(24–33)
<18	57	(1.4)
18–24	1188	(28.2)
25–29	1034	(24.6)
30–34	832	(19.8)
35–39	505	(12.0)
40–44	217	(5.2)
≥45	139	(3.3)
No data	236	(5.5)
<i>Gender</i>		
Male	752	(17.9)
Female	3426	(81.4)
No data	30	(0.7)
<i>Follower</i>		
No followers	3246	(77.1)
One or more	962	(22.9)
<i>Purchase</i>		
Yes	2955	(70.2)
No	1253	(29.8)

Abbreviation: *IQR* interquartile range

*Data are number (percentage) except where indicated

Table 2 Odds ratios of purchase in CheckMe app users who completed any assignments in POYA stratified by gender

Gender	Case, No. (%)		Control, No. (%)		OR	(95% CI)	<i>P</i> value*
Male/others	438	(58.2)	2517	(72.8)	0.52	(0.44–0.61)	<0.001
Female/others	2496	(72.9)	459	(58.7)	1.89	(1.61–2.22)	<0.001
No data/others	21	(70.0)	2934	(70.2)	0.99	(0.45–2.17)	1.000

Abbreviations: *CI* confidence interval; *OR* odds ratio

*Group comparisons by the chi-square test

95 % CI, 0.37–1.06; *P* = 0.378). The result was not significant because of the fewer people in the group. Although the main subjects were 18–39 years old, the purchase probability of 40–44 years old subjects was higher than others (OR, 1.37; 95 % CI, 0.99–1.88; *P* = 0.296).

Table 4 shows the associations of purchase probability with subjects’ number of followers on Facebook. We found that the purchase probability of the subjects who had one or more followers on Facebook was significantly higher than those with no followers (OR, 1.19; 95 % CI, 1.02–1.38; *P* = 0.029).

Table 3 Odds ratios of purchase in CheckMe app users who completed any assignments in POYA stratified by age

Age	Case, No. (%)		Control, No. (%)		OR	(95% CI)	P value*
	No.	%	No.	%			
<18/others	34	(59.6)	2921	(70.4)	0.62	(0.37–1.06)	0.378
18–24/others	831	(69.9)	2124	(70.3)	0.98	(0.85–1.14)	0.996
25–29/others	718	(69.4)	2237	(70.5)	0.95	(0.82–1.11)	0.940
30–34/others	598	(71.9)	2357	(69.8)	1.11	(0.93–1.31)	0.717
35–39/others	351	(69.5)	2604	(70.3)	0.96	(0.79–1.18)	0.986
40–44/others	165	(76.0)	2790	(69.9)	1.37	(0.99–1.88)	0.296
≥45/others	101	(72.7)	2854	(70.1)	1.13	(0.78–1.65)	0.938
No data/others	157	(66.5)	2798	(70.4)	0.83	(0.63–1.10)	0.651

Abbreviations: *CI* confidence interval; *OR* odds ratio

*Group comparisons by the chi-square test

Table 4 Odds ratios of purchase in CheckMe app users who completed any assignments in POYA stratified by number of followers on Facebook

Followers	Case, No. (%)		Control, No. (%)		OR	(95% CI)	P value*
	No.	%	No.	%			
No followers/others	2250	(69.3)	705	(73.3)	0.82	(0.70–0.97)	0.133
One or more/others	705	(73.6)	5205	(69.7)	1.19	(1.02–1.38)	0.029

Abbreviations: *CI* confidence interval; *OR* odds ratio

*Group comparisons by the chi-square test

Table 5 shows the associations of purchase probability with subjects' current city. Most of the customers were from the metropolises (Taipei, Taichung, Tainan and Kaohsiung). However, the purchase probability of the subjects in Hsinchu (OR, 1.37; 95 % CI, 0.90–2.07; $P = 0.530$) and Pingtung (OR, 1.51; 95 % CI, 0.90–2.52; $P = 0.479$) was higher than that in other cities. The purchase probability of the subjects in Chiayi is the lowest (OR, 0.54; 95 % CI, 0.33–0.89; $P = 0.108$). The results were not significant because of the fewer people in the groups.

Table 6 shows the conversion rate of CheckMe app users who completed any assignments in POYA during different periods. 1657 of 2811 (58.9%) subjects purchased some things in POYA during January 29, 2015 to March 4, 2015. According to the analysis results we provided, POYA improved its promotions. 2955 of 4208 (70.2 %) subjects purchased some things in POYA during January 29, 2015 to April 8, 2015. The conversion rate got an up to 10 % improvement during one month.

Table 5 Odds ratios of purchase in CheckMe app users who completed any assignments in POYA stratified by current city

Current city	Case, No. (%)		Control, No. (%)		OR	(95% CI)	P value*
Taipei	359	(69.2)	2596	(70.4)	0.95	(0.77–1.15)	0.958
Taoyuan	90	(72.6)	2865	(70.2)	1.13	(0.76–1.68)	0.952
Hsinchu	96	(76.2)	2859	(70.0)	1.37	(0.90–2.07)	0.530
Taichung	386	(69.9)	2569	(70.3)	0.98	(0.81–1.20)	0.999
Changhua	85	(73.3)	2870	(70.1)	1.17	(0.77–1.77)	0.912
Yunlin	64	(75.3)	2891	(70.1)	1.30	(0.79–2.14)	0.785
Chiayi	36	(56.3)	2919	(70.4)	0.54	(0.33–0.89)	0.108
Tainan	304	(70.7)	2651	(70.2)	1.03	(0.82–1.28)	0.997
Kaohsiung	312	(70.3)	2643	(70.2)	1.00	(0.81–1.24)	1.000
Pingtung	67	(77.9)	2888	(70.1)	1.51	(0.90–2.52)	0.479
No data or others+	1156	(69.6)	1799	(70.7)	0.95	(0.83–1.09)	0.899

Abbreviations: *CI* confidence interval; *OR* odds ratio

*Group comparisons by the chi-square test

+The group included users with no city information or users lived in the city with less than 20 users

Table 6 Conversion rate of CheckMe app users who completed any assignments in POYA during different periods

Period	Customers	All subjects	Conversion rate (%)
2015/01/29–2015/03/04	1657	2811	58.9
2015/01/29–2015/04/08	2955	4208	70.2

4 Discussion

In this study, we analyzed the characteristics and purchase probability of the subjects in POYA in different groups and got some interesting findings. For example, the main subjects were female between 18 and 39 years old. However, we found that the purchase probability of 40–44 years old subjects was higher than others. Furthermore, POYA improved its promotions according to our analysis results and got an up to 10 % improvement in conversion rate.

We found that the purchase probability of the subjects who had one or more followers on Facebook was significantly higher than those with no followers. In our other studies, the similar results were also found in other stores, especially in the restaurants. Most of the subjects had no followers on Facebook (77.1%). A person had followers probably because he/she was an opinion leader and usually shared experiences of using some products or eating in restaurants. That may be the reason why the purchase probability of the subjects who had one or more followers was higher.

We also found that most of the customers were from the metropolises (Taipei, Taichung, Tainan and Kaohsiung). However, the purchase probability of the subjects in Hsinchu and Pingtung was higher than that in other cities. The purchase probability of the subjects in Chiayi is the lowest. Although the population was higher in metropolises than in other cities, there were more shopping streets too. Much people went out for entertainment instead of only shopping. That may be the reason why the purchase probability of the subjects from the metropolises was lower than Hsinchu and Pingtung.

The strengths of this study are that we cooperate with the information service providers. We combine the heterogeneous data from the social network and the consumer behavior data. We provide the analysis results for our customers and improve their conversion rates. There are also some limitations to this study. We collect the consumer behavior data from the information service providers' app. If someone is POYA's customer but not the information service providers' app user, we don't have his/her information.

In conclusion, we analyzed the characteristics and purchase probability of the subjects in POYA in different groups. The main subjects were female between 18 and 39 years old. However, the purchase probability of 40–44 years old subjects was higher than others. The purchase probability of the subjects who had one or more followers on Facebook was significantly higher than those with no followers. Most of the customers were from the metropolises. However, the purchase probability of the subjects in Hsinchu and Pingtung was higher than that in other cities. Besides, we provided analysis results for POYA. POYA improved its promotions and the conversion rate increased from 58.9 to 70.2 % during one month, an up to 10 % improvement.

Acknowledgment This study is conducted under the “Online and Offline integrated Smart Commerce Platform (1/4)” of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

References

1. Bonchi F, Castillo C, Gionis A, Jaimes A (2011) Social network analysis and mining for business applications. *ACM Trans Intell Syst Technol* 2
2. Dhaliwal S, Jahangirli H, Aljomai R, Sarhan A, Almansoori W, Alhadj R (2013) Integrating social network analysis and data mining techniques into effective e-market framework, presented at the 6th international conference on information technology
3. Hea W, Zha S, Li L (2013) Social media competitive analysis and text mining: a case study in the pizza industry. *Int J Inf Manag* 33:464–472
4. Wu H-Y, Liu K-L, Trappey C (2014) Understanding customers using Facebook Pages: data mining users feedback using text analysis. In *Proceedings of the 2014 IEEE 18th international conference on computer supported cooperative work in design*, Hsinchu, pp 346–350
5. Liao C-H, Lei Y-H, Liou K-Y, Lin J-S, Yeh H-F (2015) Using big data for profiling heavy users in top video apps. In *IEEE bigdata congress*

6. Prassas G, Pramataris KC, Papaemmanouil O, Doukidis GJ (2001) A recommender system for online shopping based on past customer behaviour. In: Proceedings of the 5th international conference on intelligent user interfaces
7. Bambini R, Cremonesi P, Turrin R (2010) A recommender system for an IPTV service provider: a real large-scale production environment. *Recommender systems handbook*, pp 299–331

Novel Scheme for the Distribution of Flyers Using a Real Movement Model for DTNs

Tzu-Chieh Tsai and Ho-Hsiang Chan

Abstract In delay tolerant networks (DTNs), simulations used to verify the performance of a routing algorithm usually employ a mobility model, either trace or synthetic. Trace models record the actual movement of individuals in the real world; however, obtaining data can be difficult. Synthetic models use mathematical modelling, which eliminates the need to obtain data from participants; however, this means that the results are not necessarily representative of the actual movement of individuals. This study collected information related to the movement of university students using a specially designed APP featuring location aware and behavior aware functionality. This APP tracks the movement of students in a campus environment and then exports the data for simulation. We also developed a method for the distribution of flyers only to individuals who express an interest in the content of that particular message who can then forward the flyers to others. Simulation results demonstrate that the proposed method is able to enhance the successful delivery ratio while reducing delivery overhead and thereby improve the dissemination of data on campus.

1 Introduction

Delay tolerant networks are sparse mobile wireless networks without a fixed infrastructure. As a result, network connections are unstable, nodes move in unpredictable directions, and establishing an end-to-end route for connections can be difficult because communication between nodes tends to be intermittent.

Mobility models are generally classified into two types: trace and synthetic. Trace models, such as Infocom 06 trace data [1], Cambridge traces data [2] and MIT Reality trace [3], record the movement of individuals in the real world.

T.-C. Tsai (✉) · H.-H. Chan
Department of Computer Science, National Chengchi University, Taipei, Taiwan
e-mail: ttsai@cs.nccu.edu.tw

H.-H. Chan
e-mail: 100753503@nccu.edu.tw

Unfortunately, the collection of the data used in trace models is not easy. Synthetic models, such as the random walk mobility model and random waypoint mobility model [4], are based on mathematical models.

Considerable research has also been conducted on the dissemination of messages using DTNs. Most these studies rely on the contact utility of nodes to determine whether messages should be forwarded to another node close to the destination. Unfortunately, this kind of distribution is considered a form of spam, is irritating and greatly increases delivery overhead. This study proposes a real movement model to overcome the problem of message flooding and reduce distribution costs.

We conducted an experiment based on MIT Reality [3] for the collection of data related to the movement of university students using a specially designed Android App, called NCCU Trace Data. We sought to obtain trace data that includes a diversity of users with the aim of obtaining results that are more generalizable than that achieved using previous schemes [1–3]. The mobility trace data can then be exported to the ONE Simulator to verify the effectiveness of the routing algorithms.

We also developed a scheme for the distribution of flyers suitable for an actual campus environment. In the proposed scheme, referred to as “Direct Contact Distribute”, students send flyers only to target students that may be interested in receiving the flyer. These same students can then forward the flyers to others using a scheme referred to as InDirect Contact Distribute. Finally, we used the proposed NCCU Trace Data as a mobility model with which to evaluate the effectiveness of the flyer distribution scheme. Simulation results demonstrate that compared with other routing algorithms, the proposed method is able to enhance the successful delivery ratio while reducing delivery overhead.

2 Related Work

This study focused on two topics: (1) mobility models and (2) efficient routing schemes for DTNs.

(A) Mobility models

Mobility models can be classified into two types: (1) trace and (2) synthetic.

- (1) Trace: Previous studies on the collection of trace data include MIT reality trace data [3], Cambridge trace data [2] and Infocom 06 trace data [1]. MIT Reality trace data was obtained by deploying 100 NOKIA cellular phones to users working in a single building at MIT over a period of nine months. Infocom 06 trace data was obtained from 78 users attending a conference over a period of four days. Cambridge trace data was obtained from fifty-four users living in the city of Cambridge.
- (2) Synthetic: Each node moves independently from the others [4]. The random walk mobility model is based on random directions and speeds. The random waypoint mobility model is similar; however, the duration in which the node remains in a given location is longer.

Previous studies [1–3] are limited to activity within a single building, wherein participants frequently encounter one another. In this study, we investigated mobility models with a focus on individuals spread over a university campus, thereby more closely approximating a real world environment. Student volunteers for our experiment were recruited at random from all departments at Chengchi University to ensure that students would appear in every corner of the campus, rather than only one building. This helps to ensure that the results are more generalizable.

(B) Efficient routing scheme for DTNs

In Epidemic routing [5], nodes replicate messages and indiscriminately forward them to a single destination. In cases where a node already has a copy of the message, the message is not forwarded.

Spray and wait routing [6] proceeds through two phases: (1) spray and (2) wait. The source node initially sets L copies of messages. In the spray phase, the source node is given a strict upper bound regarding the number of copies of each message that it is allowed to introduce into the network. Half of the messages are forwarded to another node and when only one message is remaining, it is not forwarded and enters the wait phase. In the wait phase, the last message copy is forwarded to the final destination.

PROPHET routing [7] uses previous encounters with other nodes to determine which node has a higher probability of contacting the destination node and then uses this node for the delivery of that particular information.

The PeopleRank scheme [8] uses information related to a social network to identify the most popular node, which is selected to actually transmit the message.

Most routing algorithms base their message forwarding decisions on the probability of encounters between nodes. They do not take into account whether the receiver is actually interested in receiving that particular information. In this study, we employ a message forwarding strategy in which messages are delivered only to individual nodes that express an interest in the content of the message. We then used NCCU Trace Data to enhance the efficiency of the system.

3 Proposed Approach

Consider a campus environment, in which activities are advertised through the distribution of flyers via a DTN. Our research was divided into two parts: (A) the collection of data related to the mobility of students within a campus; (B) a scheme for the dissemination of messages based on the stated interest of the recipients in order to overcome the difficulties of message flooding problem and excessive traffic.

(A) Data collection

Based on previous works [1–3], we designed an Android App with location-aware and behavior-aware functionality referred to as “NCCU Trace Data” for the

Table 1 Data obtained in movement tracing experiment

Experiment dates	2014/12/17–2014/12/31
Number of volunteers	115
Number of departments	29
School days	From monday to friday
Campus size	3764 m × 3420 m

collection of daily trace data from university students. We recorded the movements of 115 university students while living on campus over a period of two weeks. The resulting data are presented in Table 1.

NCCU Trace Data was used to collect data related to the location of the user, Bluetooth devices, the usage of applications, and Wi-Fi access points, as shown in Fig. 1. Data was collected only while participants were present on campus.

NCCU Trace Data records the data as follows:

1. Location

The positioning of users is determined using the global positioning system (GPS) as well as Wi-Fi and 3G GPS when users are in buildings.

2. Usage of Android Applications on mobile phones

The frequency and duration spent using applications, such as Google Maps and Facebook, are recorded for the characterization of usage behavior.

3. Bluetooth devices

We also recorded how frequently students connect with other devices via Bluetooth as a proxy for the relationships among students.

4. Wi-Fi access points

NCCU Trace Data records how many Wi-Fi access points are proximal to a given destination by using information, such as SSID, MAC addresses, and RSSI.

Fig. 1 Screenshot showing NCCU trace data



To reduce battery consumption, the data collection APP is triggered only every ten minutes. The App runs in the background to prevent interference with the use of other applications.

We used questionnaires for the collection of personal information from users, including age, sex, grade, major, Facebook ID, personal interests, and Facebook usage behavior. Personal interests are classified into five categories: academic, athletic, artistic, community, and social activities. Personal interests were rated using a Lickert-type scale from 1 (strongly dislike) to 5 (strongly like). Summed scores are converted into standardized scores ranging from 0 to 1, with higher scores representing stronger interest. We also investigate the usage characteristics of Facebook including frequency, duration, and specific tasks, such as posting articles, sharing video links, and sending private messages.

These experiments involved the collection of personal information; therefore, prior written consent was obtained from all users and all personal information was encrypted prior to use.

(B) Targeted flyer distribution scheme

Every day, numerous flyers are published and distributed to students on campus. Unfortunately, many students are annoyed by flyers that do not interest them. Thus, we developed a novel scheme for the distribution of flyers based on the personal interests of students.

In the proposed scheme, there are two situations in which students forward flyers to other students:

1. Direct Contact Distribution

When student S_A , who has a flyer about a sporting event, meets student S_B , they first exchange hobby-related information. If student S_B is interested in sports, then S_A delivers the flyer to S_B .

2. InDirect Contact Distribution

In some cases, student S_A tries to send a flyer to student S_B ; however, S_B has no interest in sports. Nonetheless, S_B may be linked to student S_C who is interested in sports. In such cases, student S_A is able to deliver the flyer to Student S_C via Student S_B .

To determine whether a student is interested in a particular type of flyer, we employed m-dimensional cosine similarity for the calculation of attribute similarity between students and flyers. When student come into contact with other students, they exchange interest-related data and store it within a matrix. In this study, we used questionnaires to collect information related to the interests of 115 volunteers, which was then imported into a simulation program for the calculation of similarity between personal interests and flyers.

Interests are divided into *community* (I_1), *academics* (I_2), *athletics* (I_3), *arts* (I_4), and *social activities* (I_5) as well as into five preference levels: strongly like, like, neutral, dislike, strongly dislike. For example, student S_A likes community, strongly likes academics, strongly dislikes athletics, dislikes arts, and likes social activities.

These are represented using the following values: 0.75, 1, 0, 0.25, and 1, and are then shared with contacts.

When, student S_A , who has k number of flyers to distribute, meets student S_B , m -dimensional cosine similarity is used to calculate the similarity of attributes between the personal interests expressed by student S_B and the flyers that student is holding F_K , as follows:

$$\text{Cos}(S_B(I), F_k(I)) = \frac{S_B(I) \cdot F_k(I)}{\|S_B(I)\| \cdot \|F_k(I)\|} = \frac{\sum S_B(I) \times \sum_{k=1}^n F_k(I)}{\sqrt{\sum (S_B(I))^2 \times \sum_{k=1}^n (F_k(I))^2}} \quad (1)$$

The resulting similarity ranges from 0 to 1. A result of 0 indicates that the individual is not interested in flyer F_K . A result of 1 indicates that the individual is very interested in the flyer. Intermediate values indicate moderate levels of interest. The calculation of similarity enables the distribution of the flyer F_K only to students who are genuinely interested in it.

This algorithm is outlined in the following:

Algorithm Used in Flyer Distribute Scheme	
Input: Simulation_Finished_Time SFT , x-th Students' Interest attribute $S_x(I)$, k-th Flyer attribute $F_k(I)$	
Output: Distribute Flyer	
Meta_Data: {list of Student S_x 's contacts $S_x_meetlist$, Flyers' index F_k }	
Interest threshold = 0.75	
1:	While(SFT !=System_current_time){
2:	If (Connect_Student S_B)
3:	Exchange_Meta_Data();
4:	Compute Cos($S_B(I)$, $F_k(I)$);
5:	
6:	If (Cos($S_B(I)$, $F_k(I)$) > Interest threshold)
7:	Delivery one copy to S_B
8:	// Direct Contact Spread
9:	EndIf
10:	
11:	Else
12:	For(int x =0 ; x < $x_meetlist.size()$; x++)
13:	Compute Cos($S_x(I)$, $F_k(I)$);
14:	
15:	If (Cos($S_x(I)$, $F_k(I)$) > Interest threshold)
16:	Forwarding one copy to Student S_B
17:	// InDirect Contact Spread
18:	EndIf
19:	EndFor
20:	Increase(current_time);
21:	EndIf
22:	EndWhile

4 Simulation Results

Patterns of movement affect the performance of routing protocols in DTNs. Thus, we compared the performance of the proposed scheme with that of other DTN routing protocols, including Spray and Wait [6], Epidemic routing [5] and PROPHET [7]. In The ONE simulator [9] we used NCCU Trace Data as a mobility model to represent the actual movement of students on the campus throughout the week. We used the following performance metrics to measure success in the delivery of flyers:

- (1) Deliver Success Ratio: the rate of each flyer being successfully delivered to multiple destinations
- (2) Delivery Delay: the average delay in the delivery of flyers
- (3) Delivery Overhead: the average number of relays required for the delivery of flyers

(A) Simulation methods

In the following, we compare the performance of the proposed scheme with existing routing schemes:

- (1) Epidemic [5]: Students replicate flyers and forward them to all students with whom they are in contact (including all new contacts) except those who already have a copy.
- (2) Spray and Wait [6]: The source node initially sets L copies of messages. In the spray phase, the source node is given a strict upper bound regarding the number of copies of each message that it is allowed to introduce into the network. Half of the messages are forwarded to another node and when only one message is remaining, it is not forwarded and enters the wait phase. In the wait phase, the last message copy is forwarded to the final destination.
- (3) PROPHET [7]: This approach uses previous encounters with other nodes to determine which node has a higher probability of contacting the destination node and then uses this node for the delivery of that particular information. This method is well suited to situations involving multiple destinations.

(B) Simulation

The proposed scheme was implemented using The ONE Simulator (Opportunistic Network Environment Simulator) [9]. Mobility models can be imported into the ONE Simulator by converting them into movement trajectories in order to provide actual trace data for simulation. We ran simulations of the 115 students for a period of 86,400 s, which represents a day in the life of a student. The coverage of the map is 3764 m \times 3420 m, encompassing the main area of the campus. The novelty of this mobility model is its ability to model the movement of students as they proceed to classes or extracurricular activities. We considered the fact that flyers are not generated randomly, but rather, they are based on the message sending habits of the students. Our previous questionnaire asked students how often they send dynamic or

Table 2 Simulation parameters

Simulation times	86,400 s
Area	3764 m × 3420 m
Radio range	10 m
Flyer size	500 kB–1 MB
Interval of message creation	The student behavior
Data rate	2 Mbps
Buffer size	500 MB
Time to live	18,000 s

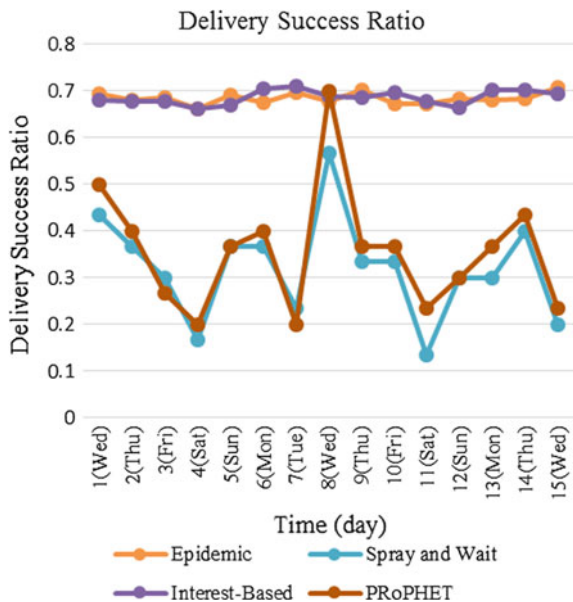
private messages on Facebook. We used the results of the questionnaire to give us an indication of the time intervals involved in the generation of messages. The simulation parameters are listed as Table 2.

(C) Simulation result

The original version of the Epidemic, Spray and Wait and PRoPHET routing schemes include only a single destination. Thus, to ensure a fair comparison, we modified the flyer to enable delivery to multiple destinations in a DTN.

As shown in Fig. 2, the performance of the proposed scheme is similar to that of the Epidemic routing scheme and clearly outperformed Spray and Wait and PRoPHET with regard to delivery success ratio. Our objective was to eliminate instances of scanning by limiting delivery only to students with an expressed interest in the content of the messages. Thus, unlike Epidemic routing, which floods every student that it meets with flyers, the proposed scheme enables the selection of students to whom specific flyers should be directed, thereby decreasing network

Fig. 2 Delivery success ratio



overhead. In addition, the performance of the Spray and Wait and PRoPHET routing schemes dropped noticeably on Saturdays, because the stored movement patterns do not match the actual movement of students on weekends. On weekends, few students encounter one another on campus, which makes it difficult to deliver or forward flyers. The Spray and Wait routing scheme limits the number of flyer copies that can be forwarded and fails to take into account the specific attributes of the flyers. Thus, in the spray phase, the flyers are forwarded indiscriminately and in the wait phase, flyers cannot be delivered. Eventually, the lifespan of the flyer is exceeded, and which point it is dropped. The PRoPHET routing scheme uses previous contact with students to identify individuals with the highest probability of reaching the contact destination. As with the Spray and Wait approach, when there are few students on campus, many of the flyers are dropped before they can be delivered.

As shown in Fig. 3, the Spray and Wait routing scheme resulted in the lowest overhead ratio because this method limits the number of copies that can be made of each flyer, such that copies are not always forwarded to all of the students who are contacted. Compare this to the Epidemic routing scheme in which an unlimited number of copies of each flyer can be made. In Epidemic routing scheme, students replicate flyers and forward them to all students with whom they are in contact (including all new contacts), except those who already have a copy. This approach increases the number of relays and subsequently in delivery overhead.

The PRoPHET routing scheme forwards flyers in accordance with the movement of students, which results in heavy delivery overhead on school days, as shown in Fig. 3. The proposed scheme distributes flyers only to students with an expressed

Fig. 3 Overhead

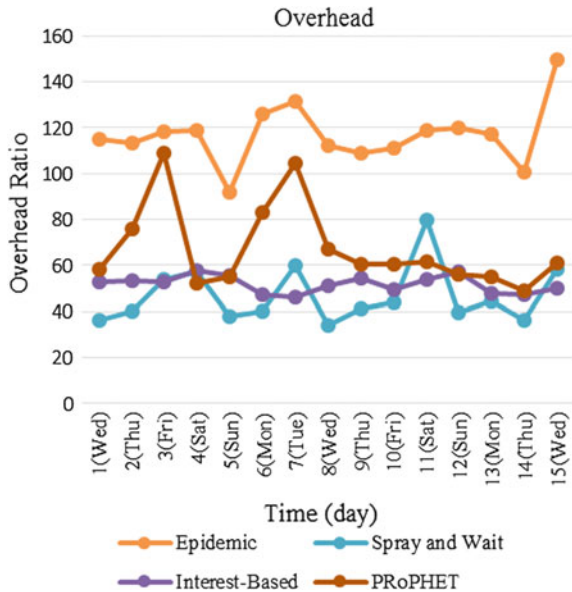
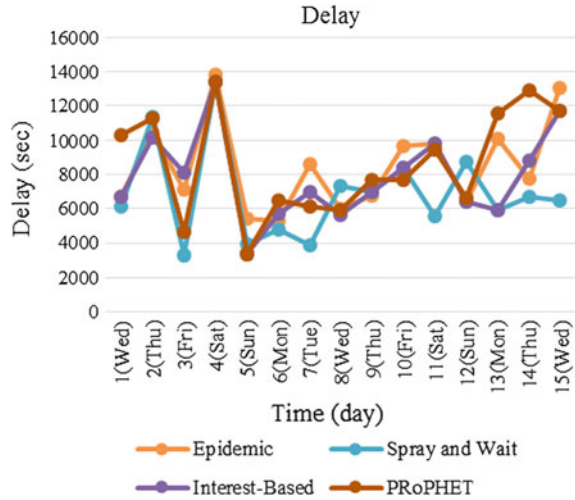


Fig. 4 Delivery delay



interest in the topic of the flyer, thereby reducing the number of unnecessary relays and decreasing delivery overhead. The proposed method reduces delivery overhead to a level far below that of Epidemic and PROPHET routing schemes

As shown in Fig. 4, all of the routing schemes have similar performance with regard to delivery delay. Particularly on Saturdays, all of the routing schemes are prone to serious delivery delays of approximately 13,000 s, due to a lack of students on campus.

The above simulation results demonstrate that the proposed NCCU Trace data can improve the performance of routing schemes on DTNs by taking into account the movement of potential targets. The efficacy of this approach was verified through a comparison with four existing routing schemes, which have particularly poor performance on weekends when students are not on campus. Our simulation results clearly demonstrate the superior performance of the proposed scheme with regard to delivery success ratio and reducing delivery overhead.

5 Conclusions and Future Work

In this paper, we developed a scheme for the distribution of flyers based on the topic of the flyer and the stated preferences of potential targets. We then developed an Android App with location-aware and behavior-aware functionality, referred to as NCCU Trace Data, to record the movement of university students on campus for use as a mobility model in simulations. Simulation results demonstrate the efficacy of the mobility model in enhancing the successful delivery ratio beyond that achieved using the PROPHET and Spray and Wait routing schemes with a delivery overhead ratio far exceeding that of Epidemic routing.

References

1. Infocom'06. Available: <http://crawdad.org/>
2. Hui P, Chaintreau A, Scott J, Gass R, Crowcroft J, Diot C (2005) Pocket switched networks and the consequences of human mobility in conference environments. In: WDTN '05: proceedings of the 2005 ACM SIGCOMM workshop on delay-tolerant networking
3. Eagle N, Pentland A (2006) Reality mining: sensing complex social systems. *Pers Ubiquitous Comput* 10(4):255–268
4. Bettstetter C, Hartenstein H, Perez-Costa X (2004) Stochastic properties of the random-waypoint mobility model. *Wirel Netw* 10(5):555–567
5. Vahdat A, Becker D (2000) Epidemic routing for partially-connected ad hoc networks. Technical Report. CS-2000-06, Duke University, July 2000
6. Spyropoulos T, Psounis K, Raghavendra CS (2005) Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: Proceedings of WDTN '05, ACM Press, pp 252–259
7. Lindgren A, Doria A, Schel'en O Probabilistic routing in intermittently connected networks, LUT, Sweden, In: Proceedings of SIGMOBILE, vol 7–3, July 2003
8. Mtibaa A, May M, Ammar M, Diot C (2010) PeopleRank: combining social and contact information for opportunistic forwarding, INFOCOM
9. Keränen A, Ott J, Kärkkäinen T (2009) The ONE simulator for DTN protocol evaluation. In: Proceedings of SimuTools, March 2009

A Study of Two-Dimensional Normal Class Grouping

Ruey-Gang Lai and Cheng-Hsien Yu

Abstract This paper aims to investigate the heterogeneous class grouping operation practiced in elementary and junior high schools and propose a two dimensional heterogeneous class grouping model which facilitates the equality of class size and students' backgrounds.

Keywords Heterogeneous class grouping · Normal class grouping

1 Introduction

According to the normal class grouping guidelines issued by the Ministry of Education of Taiwan [1], the procedure of student assignment into different classes begins with dividing all students into two groups by sex and then the heterogeneous class grouping process is executed.

However, due to the changes of the social structure in Taiwan, there is great diversity of student characteristics. Especially, there is an educationally disadvantaged generation gradually emergence, such as children of new inhabitants' children of single parent' children from grandparents raising grandchildren and so forth. If we can implant some homogeneous average strategy during the normal class grouping process, not only these students can be placed suitably but also the teachers can evenly share the teaching and counseling loads.

In this paper, we investigate on the normal class grouping operation of elementary and junior high school. Furthermore, we propose a two-dimensional normal class grouping model which can be put into practice.

The rest of this paper is organized as follows. In Sect. 2 we will review the necessary background material on the normal class grouping practiced in Taiwan.

R.-G. Lai (✉) · C.-H. Yu
China University of Technology, Taipei 106, Taiwan
e-mail: larrylai@cute.edu.tw

C.-H. Yu
e-mail: chyu@cute.edu.tw

In Sect. 3 we will propose a prospective idea about extending the normal class grouping process which is much suitable for the actual educational situation in Taiwan. Based on the prospective idea, we will build a model called “Two-dimensional Normal Class Grouping”, and then we will prove the existence of the solution of the proposed method. Section 4 offers some conclusions and suggestions for future work.

2 Literature Review

Normal class grouping (or so called heterogeneous grouping) is a type of distribution of students among various classrooms of a certain grade within a school. In this method, children of the same grade are placed into different classrooms in order to create a relatively even distribution of students of different abilities as well as different educational and emotional needs [2, 3].

Since this paper was to investigate the scope of technical aspects of the normal class grouping operation of elementary and junior high school in Taiwan, we do not pay much attention to discuss the pros and cons of the education reform policy or its related educational philosophy.

Based on the guidelines issued by the Ministry of Education of Taiwan, the normal class grouping procedure just separates students into two groups, which are boys and girls, and then executes the normal grouping process. Since there is only one parameter—gender for the grouping procedure, we call it one dimensional normal class grouping.

In recent years the changes in the social structure of Taiwan result in the changes in enrollment structure of the elementary schools and junior high schools. According to the statistics [4–6] issued by the Government, there are about a quarter of students can be summarized as educationally disadvantaged generation, such as children of new inhabitants’ children of single parent, children from grandparents raising grandchildren and so forth. Tables 1, 2 and 3 illustrates the statistics.

Faced with the complexity, if some strategies can be implemented during the normal class grouping process, students can be appropriately placed; in addition, teaching and counseling loads can be equally shared by teachers.

Table 1 Children of new inhabitants

Year	Elementary schools		Junior high schools	
2010	149,164	9.8 %	27,863	3.0 %
2011	159,181	10.9 %	33,881	3.9 %
2012	162,021	11.8 %	41,690	4.9 %
2013	157,647	12.2 %	52,631	6.3 %
2014	146,877	11.7 %	65,568	8.0 %

Table 2 Parent’s native nationality (new inhabitants, 2010–2014)

Vietnam	38.32 %
China	36.70 %
Indonesia	13.47 %
Philippines	2.68 %
Cambodia	2.39 %
Thailand	2.11 %
Myanmar	1.16 %
Malaysia	0.80 %
Others	2.37 %

Table 3 Family background of students

Year		2008 (%)	2010 (%)	2012 (%)	2013 (%)
Low income	Elementary	1.88	2.39	3.18	3.36
	Junior high	2.34	2.98	3.86	4.10
Single parent	Elementary	9.54	10.00	10.14	10.66
	Junior high	11.40	11.72	12.03	11.02
GRG	Elementary	2.32	2.36	2.43	2.38
	Junior high	1.84	1.96	2.44	2.19
Foster family	Elementary	0.72	0.66	0.56	0.50
	Junior high	0.81	0.75	1.06	0.70

GRG Grandparents Raising Grandchildren

3 Two-Dimensional Grouping Model

In this section, we propose a model called “Two-dimensional Normal Class Grouping”, namely, in addition to using one parameter “gender” in the original normal class grouping model, we introduce one more parameter and separate the students into four groups.

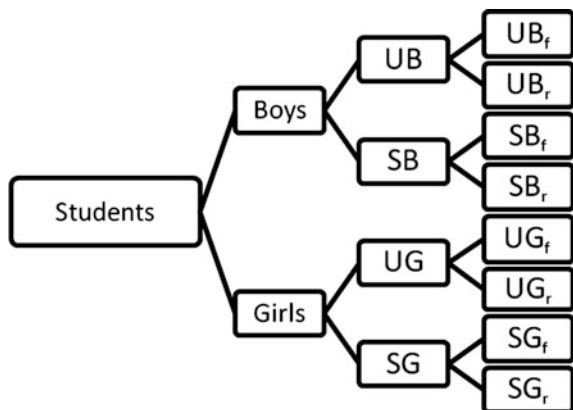
Sufficiently but not necessarily the second parameter is to distinguish if the student is educationally disadvantaged or not. For example, there are schools using bi-entrances system to recruit new students and our model is also suitable for these schools. Table 4 gives a brief description of the notation we use in this paper.

According to the definitions above, one can clearly notice that the include relationship between these different student groups as Fig. 1.

Table 4 Symbol table

N	Number of classes
UB	General group of boys
SB	Special group of boys
UG	General group of girls
SG	Special group of girls
UB _f	After first stage distribution operation, those boys in UB who can be distributed into some class
UB _r	After first stage distribution operation, those boys in UB who can not be distributed into any class
SB _f	After first stage distribution operation, those boys in SB who can be distributed into some class
SB _r	After first stage distribution operation, those boys in SB who can not be distributed into any class
UG _f	After first stage distribution operation, those girls in UG who can be distributed into some class
UG _r	After first stage distribution operation, those girls in UG who can not be distributed into any class
SG _f	After first stage distribution operation, those girls in SG who can be distributed into some class
SG _r	After first stage distribution operation, those girls in SG who can not be distributed into any class
B	Individual in UB
G	Individual in UG
B	Individual in SB
G	Individual in SG
•	Number in group •

Fig. 1 Include relationship between different student groups involved



3.1 The Objective of Two-Dimensional Normal Class Grouping

In order to perform our model, “the objective of two-dimensional normal class grouping” is defined as the following.

Definition 1 The assembly of the following nine conditions is called “The objective of two-dimensional normal class grouping ($\equiv OJ$)”:

The objective of two-dimensional normal class grouping (OJ)
 Between any two classes:

1. The difference number of boys distributed from SB is 0 or at most 1.
2. The difference number of girls distributed from SB is 0 or at most 1.
3. The difference number of boys distributed from UB is 0 or at most 1.
4. The difference number of girls distributed from UB is 0 or at most 1.
5. The difference number of boys is 0 or at most 1.
6. The difference number of girls is 0 or at most 1.
7. The difference number of students of special group is 0 or at most 1.
8. The difference number of students of general group is 0 or at most 1.
9. The difference number of students is 0 or at most 1.

In order to perform our distribution method and to check if it satisfies the condition OJ, we first require that there is an order within each UB, SB, UG or SG group (the order may be decided by birthday order, intelligence quotient (IQ) or some other quotient.) Secondly, we introduce the S-shape distribution (see Fig. 2) and the inverse S-shape distribution (see Fig. 3) which are two different ways to distribute students into each class.

After the preparation, now we can carry out the steps as the flow chart shows as Fig. 4.

Fig. 2 S-shape distribution

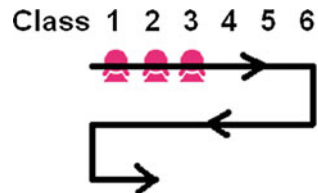
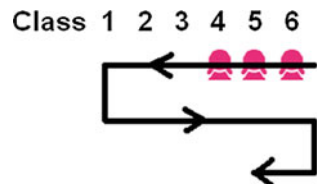


Fig. 3 Inverse S-shape distribution



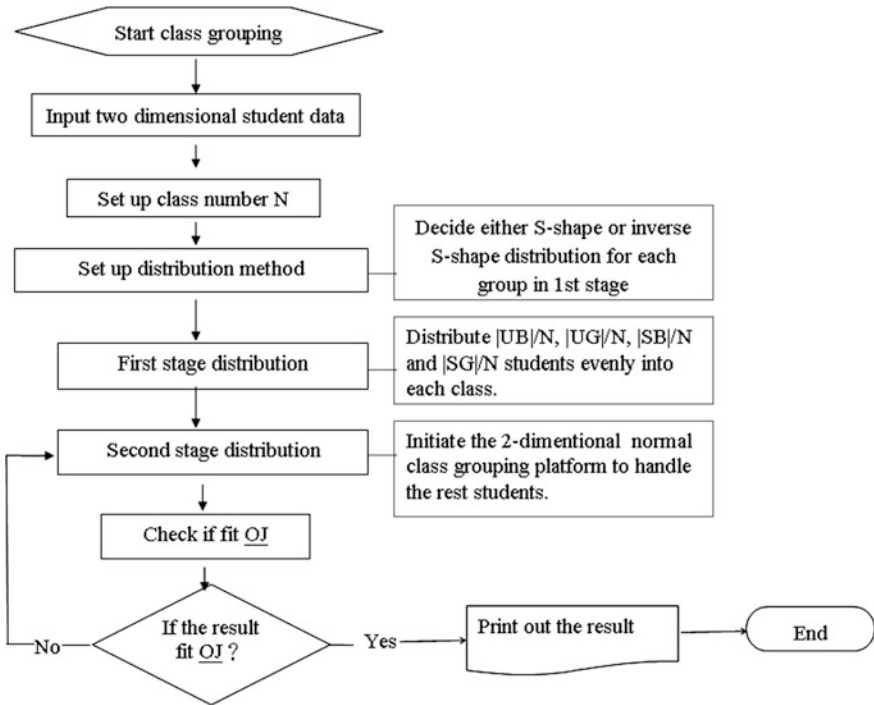


Fig. 4 Flow chart of the proposed method

3.2 First Stage Distribution

At this stage, one needs to deal with students in these four groups UB, UG, SB and SG which can be distributed into each class evenly. It is not difficult to see that for each group, it can distribute (the number of this group)/N students into each class evenly. Without loss of generality, we take Group UB as an example:

Example 1 $N = 6$, $|UB| = 14$, if we chose the S-shape distribution for group UB in first stage, the distribution process, the groups UB_f and UB_r show in Fig. 5.

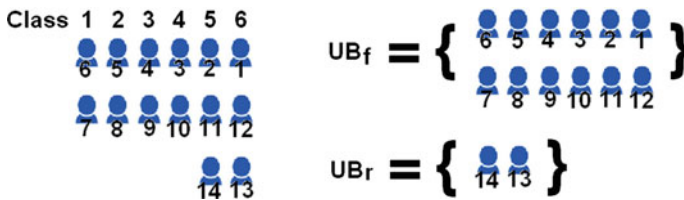


Fig. 5 First stage distribution and the groups UB_f, UB_r

We notice that $|UB_r| < N$, $|UG_r| < N$, $|SB_r| < N$ and $|SG_r| < N$. For students in groups UB_r , UG_r , SB_r and SG_r that can not be distributed into some class in the first stage, we will need to introduce “two-dimensional normal class grouping platform” to deal with this issue.

3.3 Two-Dimensional Normal Class Grouping Platform and Second Stage Distribution

Before we define “two-dimensional normal class grouping platform”, first we introduce the concepts of “homogeneous student pair” and “heterogeneous student pair” as follows.

As shown in Fig. 6, all adjacent pairs, such as $\{B, G\}$, $\{G, g\}$, $\{g, b\}$ and $\{b, B\}$, are called “homogeneous student pairs”, and all opposite pairs, such as $\{B, g\}$ and $\{G, b\}$ are called “heterogeneous student pairs”. These two concepts are particularly useful when we deal with the grouping problem of those students from groups UB_r , UG_r , SB_r and SG_r . That is, in order to keep the balance among classes and fit condition OJ, during the second stage distributing process, if one particular class has been distributed with a “B”, next student been distributed into this class better comes from B’s heterogeneous student pair, which is a “g”.

Now we can build up a two-dimensional normal class grouping platform as Fig. 7 shows.

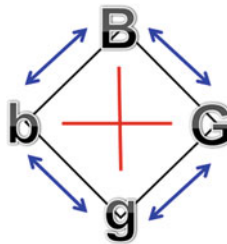


Fig. 6 Schematic diagram for “homogeneous student pair” and “heterogeneous student pair”

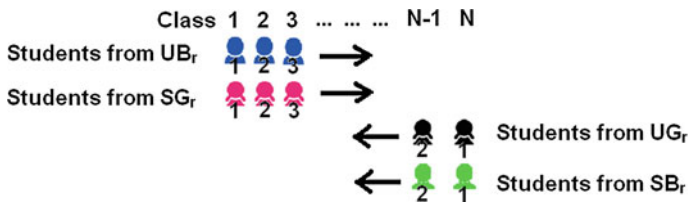


Fig. 7 Two-dimensional normal class grouping platform

As shown in Fig. 7, students from UBr and SGr (heterogeneous student pair) are distributed into classes in order (i.e. the direction from 1 to N), and students from UGr and SBr (heterogeneous student pair) are distributed into classes in reverse order (i.e. the direction from N to 1).

Here we need to point out that our proposed method takes advantage of the two-dimensional normal class grouping platform to distribute students of groups UBr, UGr, SBr and SGr will automatically fulfill condition 1 to condition 8 of OJ. Hence one only need to check if the preliminary distributing result via two-dimensional normal class grouping platform fits condition 9 of OJ, namely, between any two classes, the difference number of students is 0 or at most 1.

Actually, we can prove the following theorem, in other words, even the preliminary distributing result via two-dimensional normal class grouping platform does not fit condition 9 of OJ, as long as one makes some adjustment, there exists a distributing way which fulfills the condition OJ.

Theorem 1 (Existence of Solution) *For a fixed class number N and any finite number of students which can be divided into four groups UB, UG, SB and SG, there exists a distributing way to distribute all these students into N classes and fulfill condition OJ.*

To prove this theorem, we need to build up two lemmas.

Lemma 1 *After the preliminary distributing result via two-dimensional normal class grouping platform, if between any two classes, the difference number of students is 0 or at most 1, then this distributing way fits condition OJ.*

Lemma 2 *After the preliminary distributing result via two-dimensional normal class grouping platform, it is impossible to have two classes (say Class X and Class Y) with $||X| - |Y|| \geq 3$.*

Brief proof of the Theorem With the help of Lemmas 1 and 2, one can see that if the distributing result does not fit condition 9 of OJ, there exist at least 2 classes, say Class X and Class Y with $|X| = |Y| + 2$. The adjustment is to move a student from class X to class Y and this particular student belongs to one of the Groups UBr, UGr, SBr or SGr such that class Y does not have any student in that group. With finite iterations of adjustment, one can prove the existence of the solution and the theorem.

Although we can prove the existence of the solution, namely, there exists a distributing way to distribute all students into N classes and fulfill condition OJ. The solution is not unique.

4 Conclusions

In this paper, we propose a prospective idea about extending the normal class grouping process which is called “two dimensional normal class grouping” and we make a fine definition of its objective. Finally, we prove the existence of its

solution. One clear future direction is to extend this two dimensional problem to multi-dimension problem.

We, in the end, hope to arouse the educational authorities' awareness. We expect they can cogitate on the imbalanced classrooms and the issue we try to address.

References

1. The Ministry of Education of Taiwan (2009) The normal class grouping guidelines for elementary schools and junior high schools
2. HESS, Natalie (2005) Teaching large multilevel classes. Cambridge University Press, Cambridge
3. Prodromou, Luke (1992) Mixed ability classes. Macmillan Publisher, London
4. of Household Registration, M.O.I. (2015) The number of foreign spouses by nationality and mainland China (including Hong Kong and Macao)
5. The Department of Statistics of the Ministry of Education (2015) The statistics of children of new inhabitants attending elementary schools and junior high schools
6. The Department of Statistics of the Ministry of Education (2015) The statics of the family background of the students of elementary schools and high schools

Visualized Comparison as a Correctness Indicator for Music Sight-Singing Learning Interface Evaluation—A Pitch Recognition Technology Study

Yu Ting Huang and Chi Nung Chu

Abstract This paper demonstrated the efficiency of visualized interface for the music sight-singing learning on the Internet. The self-generated visualization on music sight-singing learning system incorporates pitch recognition engine and visualized pitch distinguishing waveforms with descriptions for each corresponding stave notation on the web page to bridge the gap between singing of pitch and music notation. There are three anticipated effects from the design of web-based learning system of self-generated visualization on pitch recognition for the music education in sight-singing: visualizing sight-singing music notes, tuning errors in sight-singing visually with quantified scale, and transforming hearing into vision resulting in individualized sight-singing learning. This paper shows the conducted research results that this web-based sight-singing learning system could scaffold cognition about aural skills effectively for the learner through the Internet.

Keywords Pitch recognition · Self-generated visualization · Sight-singing · Music education

1 Introduction

Sight-singing skills tone up a music learner's ability to be more accurate when performing unread music [1, 2]. Sight-reading is the essential skill in music learning to improve further music performance, such as better understanding of music repertoire, ability to play more complex scores, music composition, greater

Y.T. Huang (✉)

Department of Music, Shih Chien University,
No.70 Ta-Chih Street, Chung-Shan, Taipei, Taiwan, ROC
e-mail: yuting11@mail.usc.edu.tw

C.N. Chu

Department of Management of Information System, China University of Technology,
No. 56, Sec. 3, Shinglung Rd, Wenshan Chiu 116, Taipei, Taiwan, ROC
e-mail: nung@cute.edu.tw

retention of music notations, and even be a musician [3–6]. Many constituent elements are involved in the process of sight-singing which includes an individual’s perceptive competence, knowledge, and experiences. Multiple cognitive procedures are involved concurrently when learners read music by sight [7–9]. In traditional schooling, where students learn music from notations, sight-singing demonstrates a student’s music literacy and music understanding. However the process of sight-singing involves the conversion of musical information from sight to sound [10, 11]. It is hard for the learners to distinguish their sound of soundness by themselves. As the web environment is becoming an effective educational media [12], the goal of this study with pitch recognition design is to construct a facilitating sight-singing learning interface which adopts visualization strategy to transform the singing sound to a wave curve. By comparing with the standard wave curve of music notations, the learners would identify the correctness of their music sight-singing from the self-generated wave curve. Thus learners can build their own foundation of sight-singing skills by themselves from the Internet.

2 Self-generated Visualization on Music Sight-Singing Design

The design of web-based learning system of self-generated visualization on pitch recognition (Fig. 1) is based on pitch recognition engine running on Windows platform. It integrates a voice recorder as a music sight-singing input producer. It also connects with the standard MIDI pitch producer from the music software Finale. The standard MIDI pitch producer could act like threshold for learners to distinguish their sight-singing status. Each pitch of music note sang by a user could be explicitly recognized by web-based learning system of self-generated visualization on pitch recognition and responded visually with the compared results of accuracy (Fig. 2). The integrated web-based learning system of self-generated visualization on pitch recognition software is installed at client side. This design not only reduces the overload of server computation, but also avoids redesigning of the existing web sites for the user’s special hearing needs.

The design of web-based learning system of self-generated visualization on pitch recognition provides a useful visual mechanism of accessible learner sight-singing. The goal of this design is to construct automatic visual interfaces for sight-singing learning in web-based environment. For each the music notes would be sampled via the voice recorder once sung by a learner, the pitch recognition engine could then

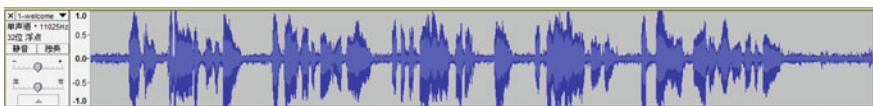


Fig. 1 Web-based learning system of self-generated visualization on pitch recognition

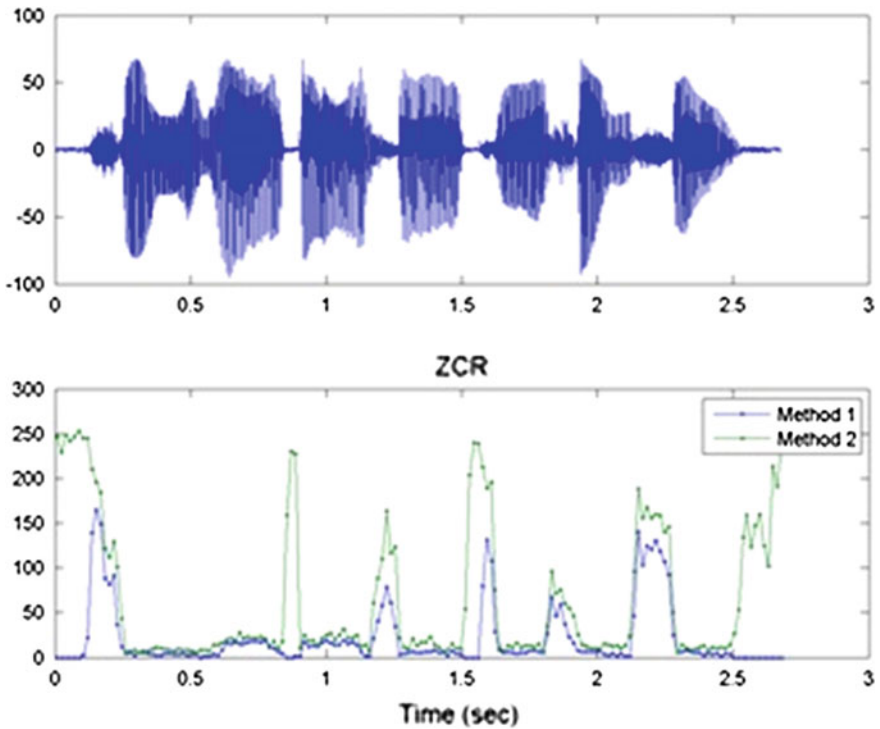


Fig. 2 Visual responding accuracy of sight-singing note

perform the following functions as requested to facilitate the access of sight-singing of music notes.

```
private static Double[] bytesToDoubles(byte[] bytes) {
    Double[] double = new Double[bytes.length / 2];
    for(int i=0; i < bytes.length; i+=2) {
        double[i/2] = bytes[i] | (bytes[i+1] << 8);
    }
    return double;
}
```

1. Transforming each music into corresponding wave forms, which is used to compare with the standard MIDI pitch producer.
2. Producing compared result of each music note with the wave forms corresponding to the music staff.
3. Highlighting the error of sight-singing note.
4. Replaying the sight-singing of user with feedback of pitch accuracy comparing.

Thus, user can access the sight-singing of music notations visually on the Internet through this web-based learning system of self-generated visualization on pitch recognition.

3 Benefits Evaluation

The performances of web-based learning system of self-generated visualization on pitch recognition were evaluated for each subject with the tasks of sight-singing. There were 32 participants conducted in this research, 16 students majoring music in senior high school and 16 students majoring music in university. Subjects participated in four 90-min sessions. Each session consisted of 30 items of trials.

3.1 Efficiency of Self-generated Visualization on Pitch Recognition

The analyses of learning time showed the beneficial effects of implementing self-generated visualization on pitch recognition versus traditional sight-singing environment. There was a significant main effect of learning environment ($F(1,26) = 58.66, p < .001$). There were also two significant, two-way interactions: learning environment and age ($F(1,26) = 5.78, p < .05$); and learning environment and music intervals ($F(3,78) = 36.53, p < .001$). Furthermore, there was a significant three-way interaction of learning environment, music intervals and age ($F(3,78) = 4.16, p < .01$). The university students made better sight-singing learning curve than did senior high school students, and the sight-singing learning curve was better with the interface of self-generated visualization on pitch recognition than with traditional sight-singing environment for both age level students. However, the senior high school students benefited more from the interface of self-generated visualization on pitch recognition when singing more complex music intervals than did the university students.

3.2 Discrimination of Self-generated Visualization on Pitch Recognition

The interaction between self-generated visualization on pitch recognition and complex music intervals were significant ($F(2,48) = 34.93, p < .001$). The follow-up tests showed that the self-generated visualization on pitch recognition managed to get a better sight-singing learning curve. More importantly, there was no difference between self-generated visualization on pitch recognition versus

traditional sight-singing environment when the visualized comparison was showed up. This exhibited that both the senior high school students and the university students were able to use the self-generated visualization on pitch recognition with the visualized comparison to distinguish the complex music intervals in sight-singing.

3.3 Visualized Comparison as a Correctness Indicator

The analyses yielded significant effects ($p < .10$) with the quantified scales which worked as the correcting indicator of visualized comparison in sight-singing. There was, however, a significant interaction of self-generated visualization on pitch recognition and music intervals ($F(3,78) = 18.94, p < .001$). The follow-up tests showed that there was a beneficial effect of visualized comparison as a correctness indicator when music intervals were complex as compared with the simple ones. The senior high school students seemed to benefit more from visualized comparison as a correctness indicator than the university students.

4 Conclusion

There are three anticipated effects from the design of web-based learning system of self-generated visualization on pitch recognition for the music education in sight-singing as the follows:

1. Visualizing sight-singing music notes
As implementing the web-based learning system of self-generated visualization on pitch recognition, each music note sung by the sight-singing learner could be recognized and transformed into waveforms which are displayed under each corresponding music notes.
2. Tuning errors in sight-singing visually with quantified scale
Erroneously sung music notes in sight-singing can be identified and depicted by the web-based learning system of self-generated visualization on pitch recognition with quantified scale in waveforms with each corresponding music notes. The learner could visually distinguish his/her own waveforms from the standard waveforms for the erroneous music notes sung. And thus the learner could specifically adjust his/her errors during sight-singing.
3. Transforming hearing into vision results in individualized sight-singing learning
As the sound of sight-singing can be visualized, the web-based learning system of self-generated visualization on pitch recognition would facilitate sight-singing learners in the off class practice alone without the help of hearing correctness identification from other people in traditional sight-singing learning environment.

The visualized learning interface of self-generated visualization on music sight-singing could move sight-singing learners beyond basic drill exercises to a competence that is tailored to the content of individual needs in the sight-singing training. And the recording and replay functionalities created from the web environment facilitate learners to master the skills. The remediation would be visualized with quantified scales by the exercise itself. Sight-singing skills development occurs when learners interact with the visualized learning interface of self-generated visualization on music sight-singing in a continuous drill. Many singing sound faults could then be tuned up through the combination of practice and immediate visual feedback.

References

1. Lehmann AC, Ericsson KA (1996) Performance without preparation: structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology* 15(1-2):1-29
2. Sloboda JA (1984) Experimental studies of music reading: a review. *Music Percept* 2(2):222-236
3. Draai-Zerbib V, Baccino T (2013) The effect of expertise in music reading: cross-modal competence. *J Eye Mov Res* 6(5):1
4. Gudmundsdottir HR (2010) Advances in music-reading research. *Music Educ Res* 12(4):331-338
5. Kopiez R, Weihs C, Ligges U, Lee JI (2006) Classification of high and low achievers in a music sight-reading task. *Psychol Music* 34(1):5-26
6. Zhukov K (2014) Evaluating new approaches to teaching of sight-reading skills to advanced pianists. *Music Educ Res* 16(1):70-87
7. Grutzmacher PA (1987) The effect of tonal pattern training on the aural perception, reading recognition, and melodic sight-reading achievement of first-year instrumental music students. *J Res Music Educ* 35(4):171-181
8. Mishra J (2013) Improving sightreading accuracy: a meta-analysis. *Psychol Music*. doi:[10.1177/0305735612463770](https://doi.org/10.1177/0305735612463770)
9. Wolf TE (1976) A cognitive model of musical sight-reading. *J Psycholinguist Res* 5(2):143-171
10. Hagen SL, Cremaschi A, Himonides CS (2013) Effects of extended practice with computerized eye guides for sight-reading in collegiate-level class piano. *J Music Technol Educ* 5(3):229-239
11. Lehmann A, McArthur V (2002) Sight-reading. The science and psychology of music performance: creative strategies for teaching and learning. Oxford University Press, New York
12. Web-based Education Commission (WBEC) (2000) The power of the Internet for learning: moving from promise to practice (Report). U.S. Department of Education, Washington, DC, pp 17-49

A Fuzzy Genetic Approach for Optimization of Online Auction Fraud Detection

Cheng-Hsine Yu

Abstract According to the Internet Crime Complaint Center (IC3) reports from 2006 to 2014, we can find the online fraud cases are increasing rapidly year by year. Although the online auction web site is the biggest platform for online transaction, it brings the huge chances to do the online auction frauds. To prevent the online auction frauds, this research will propose a fuzzy genetic approach to learn the detection rules for detect the fraudster accounts. The goal of this research is to help the users to identify which seller is more dangerous. The seller behavior features will transform into fuzzy rules which can represent the detection rules. Then optimize the fuzzy rules by genetic algorithms to build the auction fraud detection model. For implementation, we collect the real auction data from “Ruten Auction” which is the most popular auction site in Taiwan. Then we use the proposed detection model to analyze the fraudster accounts and find out the optimal detection rules of them. We hope the result of this research can help the website administrators to detect the possible seller fraudsters easier in online auction.

Keywords Fuzzy control system • Genetic algorithms • Fraudster detection

1 Introduction

Because of huge amount of transactions in online auction, there are a lot of chances to do auction fraud. According to the IC3 reports from 2006 to 2014, it shows that there are more and more online fraud cases occur year by year. The auction related frauds (e.g. auction fraud and no-delivery fraud) are the top popular fraud in both U. S. and Taiwan. The internet crime complaint center (IC3) report of 2014 shows that there are 269,422 internet complaints submitted. From 2006 to 2014, auction related frauds were continuing to stand the top position of Internet complaint

C.-H. Yu (✉)
China University of Technology, Taipei 106, Taiwan
e-mail: chyu@cute.edu.tw

comparing [1–9]. According to the 2014 report of the fraud prevention hotline “165” in Taiwan, the auction related frauds are also hold the top popularity [10].

To prevent the auction frauds, every online auction site builds the reputation system to help users to evaluate seller behaviors for check out the fraud problems. Despite this, the current reputation system is too simple and provides too less information to help users to prevent frauds. To solve the problem of current reputation system of online auction sites, we propose a fuzzy genetic approach for optimization of online auction fraud detection. This approach can monitor the suspicious collusive accounts, analyze the cheating behavior features and detect the fraudster accounts from real transaction data of online auction sites. We hope this approach can remedy current reputation system and help the police to detect the fraudster accounts of online auction sites.

2 Literature Review

In this section, we review the fuzzy control system and genetic algorithm which are used to construct the detection model. The related detection approaches are also surveyed for fraudster features and detection rules design. The literature review is discussing as below.

2.1 *Fuzzy Control System and Genetic Algorithms*

Fuzzy Control system which is an intelligent control approach has been applied in many fields in the past few years [11]. Fuzzy control system includes four parts: membership function, proportional factor, control rules and quantization factor. Control rules play a very important role in the system performance, and represent the dynamic behaviors of the controller. Generally, fuzzy control rules were extracted from experts’ domain know-how. In the rule design process, it is difficult to generate the parameters of the membership function. If system has too many input variables, the decision of the fuzzy rules will bring a huge selection space and increase the complexity of computation. It is equivalent to a parameter optimization problem and is difficult to obtain an optimum solution. There are some optimization methods which are used to optimize fuzzy control rules, such as genetic algorithm, neural network, combination of genetic algorithm with neural network and others [12]. GA (genetic algorithm) has been used to generate the parameters of membership function and fuzzy control rules. The evolutionary learning of GA can help to find the optimal fuzzy control rules.

2.2 *Fraud Detection in Online Auctions*

According to the literature, fraud detection approaches can be classified into three kinds of methods: statistical methods, data mining techniques and formal methods [13]. With regard to statistical methods, Rubin et al. [14] propose a new reputation system for auction sites to help users protect their interests by warning them of the risk of auction fraud. The model uses three variables (average number of bids, average minimum starting bid, and bidders' profiles) to identify whether an account shows activity typical of shill biddings [14]. Other researchers also use statistical methods to detect behaviors of fraudsters [15–17]. With regard to data mining techniques, Pandit et al. [18] propose the NetProbe approach, which models auctioneers with their transactions as a Markov Random Field to detect the fraudsters' suspicious patterns. It also constructs a belief propagation mechanism to detect similar fraudsters [18]. Formal notation and logic are used in mathematically rigorous techniques for the specification, development, and verification of software and hardware designs. Xu et al. [19] and Clarke et al. [20] use a formal approach to detect shilling behaviors. Gavish and Tucci [21] report the results of an investigation that checked for the amount of fraudulent activities on auction sites. Livingston [22] studies the functional forms of relationships between price and the seller's reputation.

Recently, Wang et al. [23] used social network analysis (SNA) indicators based on k-core and centered-weights algorithms to detect auction fraud groups because of the normal transactions do not have a complicated relationship. However, if collusive auctioneers are to manipulate their reputation, they must have transactions among themselves. Thus, collusive auctioneers' transactions have a complicated relationship and have the high-core characteristic in the transactional structure. Taking a structural perspective by focusing on the relationships between traders rather than their attribute values, Wang and Chiu used k-core and centered-weights algorithms to create a collaborative-based recommendation system that could suggest risks of collusion associated with an account [23, 24].

3 **Detection Model**

In our previous research, we build a web crawler system [25] to collect real auction data. The web crawler has agents to explore auction users and capture the users' profile data with reputation records. The web crawler also has a central server with a database used to control the parallel crawling. The web crawler can start crawling from any one user account and expand the crawling area level by level until the stop condition is done.

The process of data collection describes as below.

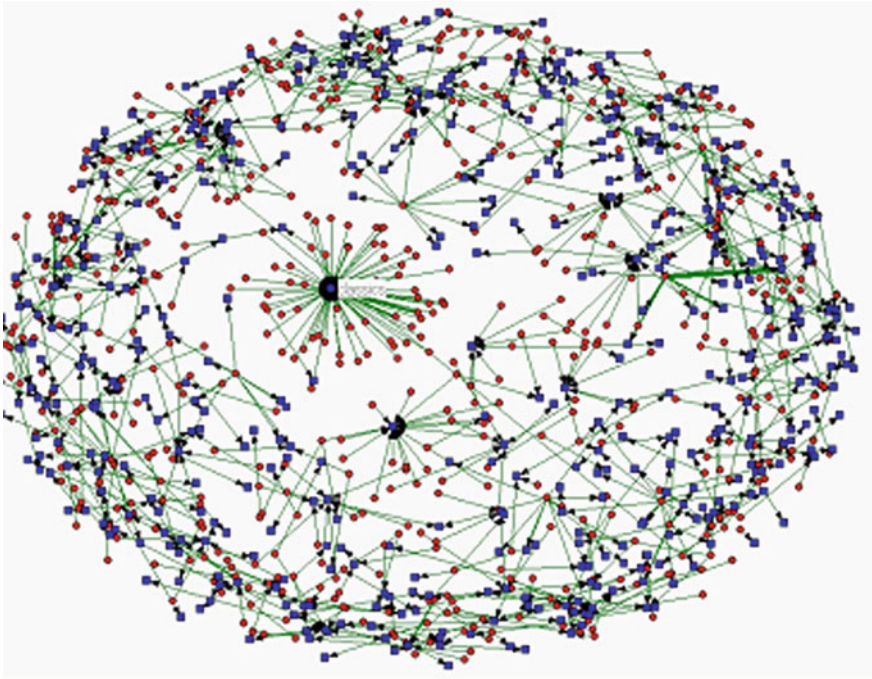


Fig. 1 Transaction network of online auction users

3.1 Web Crawler System for Online Auction Data Collection

Because of the limitation of research cost, we select three accounts randomly from the fraudster list of our collected auction data and start the data crawling from a social network perspective on the real auction site environment (<http://www.ruten.com.tw>). A parallel crawling system which is developed by our previous research is conducted to perform the web crawling tasks. The web crawling will search the direct participants account first, and then expand the crawling area to the next level of the transaction network. Three levels of the transaction network will be crawled and related accounts and bidding history will be collected too (as Fig. 1). There are three auction data sets collected and applied to the experiments.

3.2 Fuzzy Genetic Approach

This research provides a fuzzy genetic approach to learning the optimal detection rules for online auction frauds detection. The fuzzy rule base is used to encode the genes of the genetic algorithms. Membership function, fuzzification function, inference mechanism and defuzzification function of fuzzy control system are used

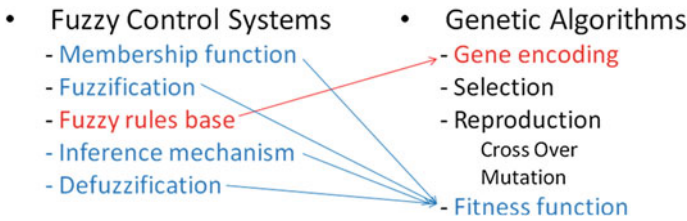


Fig. 2 The fuzzy genetic approach

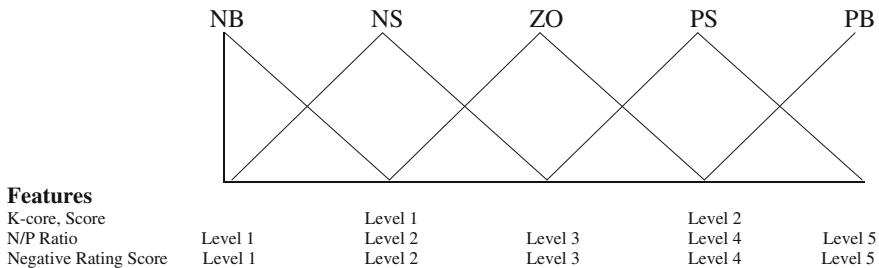


Fig. 3 Membership function

to evaluate genes in fitness function of the genetic algorithms (as Fig. 2). Fuzzy control system can generate fuzzy scores to evaluate every one gene. One gene represents one detection rule and can be evaluated by its fuzzy score. A population which includes a set of genes is evaluated to learning the optimal detection rules by selection, reputation and fitness function.

3.2.1 Membership Function

The membership function is designed as Triangular Membership Function (as Fig. 3). The parameters include Seller-Density features (K-core or Score), Crime-Economics features (Negative-Positive ratio) and Reputation-System features (Negative rating score) which are defined in our previous researches. Seller-Density features are divided into only two levels include Negative-Small (NS) and Positive-Small (PS). Negative-Positive Ratio and Negative Rating Score are divided into five levels include Negative-Big (NB), Negative-Small (NS), Zero (ZO), Positive-Small (PS), and Positive-Big (PB).

3.2.2 Gene Encoding

The gene is encoded as a fuzzy control rule table (see G0–G49 in Table 1). Every cell represents an output of a control rule. In Table 1, G0 represents the output when total negative rating score is NB, negative-positive ratio is NB and density (K-core

Table 1 Gene encoding

N/P ratio	Negative rating score				
	NB	NS	ZO	PS	PB
<i>Density (K-core or score) = NS</i>					
NB	G ₀	G ₁	G ₂	G ₃	G ₄
NS	G ₅	G ₆	G ₇	G ₈	G ₉
ZO	G ₁₀	G ₁₁	G ₁₂	G ₁₃	G ₁₄
PS	G ₁₅	G ₁₆	G ₁₇	G ₁₈	G ₁₉
PB	G ₂₀	G ₂₁	G ₂₂	G ₂₃	G ₂₄
<i>Density (K-core or score) = PS</i>					
NB	G ₂₅	G ₂₆	G ₂₇	G ₂₈	G ₂₉
NS	G ₃₀	G ₃₁	G ₃₂	G ₃₃	G ₃₄
ZO	G ₃₅	G ₃₆	G ₃₇	G ₃₈	G ₃₉
PS	G ₄₀	G ₄₁	G ₄₂	G ₄₃	G ₄₄
PB	G ₄₅	G ₄₆	G ₄₇	G ₄₈	G ₄₉

or Score) is NS. G25 represents the output when total negative rating score is NB, negative-positive ratio is NB and density (K-core or Score) is PS, and so on.

3.2.3 Fitness Function

The fitness function is designed for recognize which genes can detect fraudster accounts effectively. The optimal gene can also use to compare the difference between detection features. The important feature will get a high threshold. Otherwise, lower threshold score represent lower importance. The formula of fitness function lists as below.

$$E_i = \text{Fuzzy_Score} \tag{1}$$

$$\text{Threshold} = \frac{\text{Normal_Max_Fuzzy_Score} + \text{Fraud_Min_Fuzzy_Score}}{2} \tag{2}$$

$$\text{Recall} = \frac{\text{Count}(\text{Fraud_E}_i > \text{Threshold})}{\text{Count}(\text{Fraud_Account})} \tag{3}$$

$$\text{Precision} = \frac{\text{Count}(\text{Fraud_E}_i > \text{Threshold})}{\text{Count}(\text{Fraud_E}_i > \text{Threshold}) + \text{Count}(\text{Normal_E}_i > \text{Threshold})} \tag{4}$$

$$\begin{aligned} \text{Goal} &= \text{Max}(\text{F - measure}) \\ \text{F-measure} &= \frac{2 * \textit{precision} * \textit{recall}}{\textit{precision} + \textit{recall}} \end{aligned} \tag{5}$$

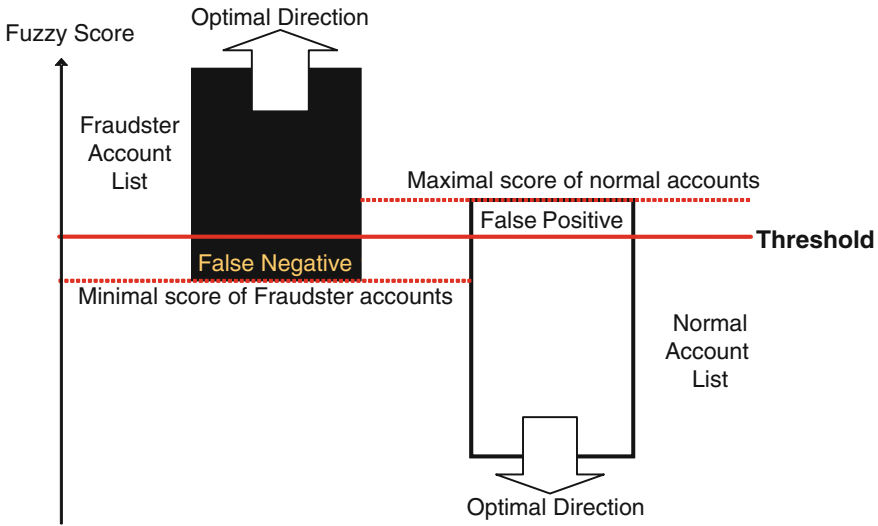


Fig. 4 Threshold design

In the formula (1), E_i is an fuzzy output value of an instance in the population. In the formula (3), Recall means the recall rate of the detection rule. $\text{Count}(\text{Fraud_}E_i > \text{Threshold})$ means the count of accounts in the fraudster list when $E_i > \text{Threshold}$. The threshold is designed as formula (2) and show as Fig. 4. Threshold is set as average of minimal fuzzy score of fraud account list and maximal fuzzy score of normal account list. $\text{Count}(\text{Fraud_Account})$ means the count of total accounts in the fraud account list. In formula (4), Precision means the precision rate of the detection rule. $\text{Count}(\text{Fraud_}E_i > \text{Threshold})$ is defined as above, and $\text{Count}(\text{Normal_}E_i > \text{Threshold})$ means the count of accounts in the normal account list when $E_i > \text{Threshold}$. In the formula (5), Goal means the goal of the genetic algorithms, and the maximal F-measure is the optimal solution. F-measure is defined as the general approach as formula (5) and provide the balance measurement between recall rate and precision rate. Through the evaluation by this fitness function, the optimal detection rules can detect all the fraudster accounts and get the maximal value of $\text{Count}(\text{Fraud_}E_i > \text{Threshold})$ and the optimal value will equal to $\text{Count}(\text{Fraud_Account})$. So the best value of Recall will be close to 1. In addition, the optimal detection rules will prevent the normal account be detected as an fraudster account, so it will get the minimal value of $\text{Count}(\text{Normal_}E_i > \text{Threshold})$ and the optimal value will be equal to zero. The best value of Precision will be close to 1 and the best value of F-measure will be close to 1 too.

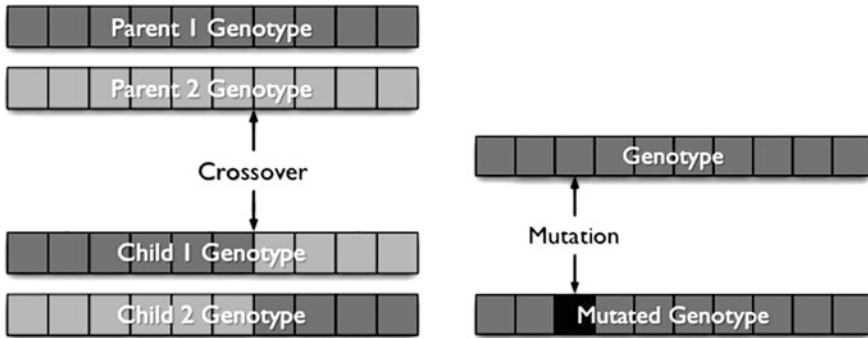


Fig. 5 Reproduction methods

3.2.4 Selection and Reproduction

The roulette wheel selection is performed in this research. In the selection operation, the higher fitness gene will have bigger roulette area to represent the higher probability of selection. Otherwise, the lower fitness gene will have smaller roulette area to represent the lower probability of selection.

There are three general types of crossover of genetic algorithms. In this research, we choose the One-Point crossover to reduce the computation of crossover and stable the evolution of the population (as Fig. 5).

The genetic algorithms have a possible problem that is the local optimal solution problem. The mutation can change the bit of gene randomly and drive the evolution to the other ways to search the possible path to get the optimal solution (as Fig. 5).

4 Experimental Results

In this research, we conduct three experiments with three data sets which collect from a real online auction website. According to the results of three experiments, we can find that the experiment #2 can find the optimal detection rule in data set #1 and #2 which can detect all fraudster accounts correctly and perform the 100% recall rate and precision rate. In the other hand, it has the general results in data set #3. The experiment #3 performs a similar result with experiment #2. Especially in data set #3, experiment #3 has the better Precision than experiment #2. It means the feature Score can has the better Precision than feature K-core. Though, the experiment #2 has the better Recall than experiment #3. It means the feature K-core can has the better Recall than feature Score. On the other hand, the data size of data set #3 is significant larger than other two data sets and the ratio of fraudster amount is significant less than other two data sets too. These two characteristics make the difficulty to recognize the fraudsters. In addition, we implement C4.5 Decision Tree

Table 2 Performance comparisons

	C4.5 decision tree		Our approach	
Data set #1	Precision	0.33	Precision	1
	Recall	1	Recall	1
	F-measure	0.49	F-measure	1
Data set #2	Precision	0.28	Precision	1
	Recall	1	Recall	1
	F-measure	0.44	F-measure	1
Data set #3	Precision	0.33	Precision	1
	Recall	0.67	Recall	0.8
	F-measure	0.44	F-measure	0.89

on our data sets and compare with our results, as show in Table 2. The parameter complexity-penalty is setting for value 0.5, minimum-support is setting for value 10 and the entropy score method is used in the decision tree implementation. Compare our approach with the C4.5 decision tree, the precision and F-measure of our approach are both better than C4.5 Decision Tree’s results. Although the decision tree provides good recall outcome, but it provides very bad precision outcome too. It means that the non-strict rules generated by the decision tree and there are many normal seller accounts will be detected with fraudster seller accounts. According to the experimental results we can find that our approach can improve the performance of recall rate and keep good precision rate of the auction fraud detection system.

5 Conclusions

In summary, this research provides a fuzzy genetic approach for online auction fraudster detection. It combined the social network analysis, economics of crime, and reputation system of online auction to build detection rules for online auction fraudster detection. For detection rule optimization, the genetic algorithm is conducting for evolve the optimal fuzzy rules used to recognize who are possible fraudsters. Through the real world data collection on Ruten auction site (<http://www.ruten.com.tw>) in Taiwan, the crawler system collected three big data sets and used in evaluation of genetic algorithms. Three experiments were conducting with three data sets. Regarding the results of these three experiments, we can find the features extracted from only original reputation system are not enough to cover the all conditions of fraudster detection. Both of the features K-core and Score can increase the detection rates especially when the data sets size were increasing. After performance comparisons with C4.5 decision tree, our approach provides better detection rates and produce readable detection rules for future applications. Finally, this research produced detection rules which can detect fraudster accounts effectively. We hope these results can provide helps for the website administrators to detect possible online auction fraudsters.

References

1. Internet Crime Complaint Center, 2006 Internet crime report (2007)
2. Internet Crime Complaint Center, 2007 Internet crime report (2008)
3. Internet Crime Complaint Center, 2008 Internet crime report (2009)
4. Internet Crime Complaint Center, 2009 Internet crime report (2010)
5. Internet Crime Complaint Center, 2010 Internet crime report (2011)
6. Internet Crime Complaint Center, 2011 Internet crime report (2012)
7. Internet Crime Complaint Center, 2012 Internet crime report (2013)
8. Internet Crime Complaint Center, 2013 Internet crime report (2014)
9. Internet Crime Complaint Center, 2014 Internet crime report (2015)
10. National Police Agency (2015) <http://www.165.gov.tw/>, Taiwan
11. Lee CC (1990) Fuzzy logic in control system: fuzzy logic controller. *IEEE Trans Syst Man Cybern* 20(2):404–435
12. Pan W, Zhang B, Zhu L (2009) Genetic algorithm combined with immune mechanism and its application in skill fuzzy control. In: Proceedings of the IEEE Chinese control and decision conference. pp 4870–4874
13. Dong F, Shatz SM, Xu H (2009) Combating online in-auction fraud: clues, techniques and challenges. *Comput Sci Rev* 3(4):245–258
14. Rubin S, Christodorescu M, Ganapathy V, Giffin JT, Kruger L, Wang H et al (2005) An auctioning reputation system based on anomaly. In: Proceedings of the 12th ACM conference on computer and communications security. p 279
15. Dong F, Shatz SM, Xu H (2009) Inference of online auction skills using dempster-shafer theory. In: Proceedings of the 6th international conference on information technology: new generations 2009, pp 908–914 (2009b)
16. Trevathan J, Read W (2007) Detecting collusive shill bidding. In: Fourth international conference on information technology 2007, ITNG'07. pp 799–808
17. Trevathan J, Read W (2009) Detecting shill bidding in online english auctions. *Handbook of Research on Social and Organizational Liabilities in Information Security*
18. Pandit S, Chau DH, Wang S, Faloutsos C (2007) Netprobe: a fast and scalable system for fraud detection in online auction networks. In: Proceedings of the 16th international conference on world wide web. p 210
19. Xu H, Cheng YT (2007) Model checking bidding behaviors in internet concurrent auctions. *Int J Comput Syst Sci Eng* 22(4):179–191
20. Clarke EM, Wing JM (1996) Formal methods: State of the art and future directions. *ACM Comput Surv (CSUR)* 28(4):643
21. Gavish B, Tucci CL (2006) Fraudulent auctions on the internet. *Electron Commer Res* 6 (2):127–140
22. Livingston JA (2010) Functional forms in studies of reputation in online auctions. *Electron Commer Res* 10(2):167–190
23. Wang JC, Chiu CC (2008) Recommending trusted online auction sellers using social network analysis. *Expert Syst Appl* 34(3):1666–1679
24. Wang JC, Chiu C (2005) Detecting online auction inflated-reputation behaviors using social network analysis. In: Annual Conference of the North American association for computational social and organizational science
25. Yu CH, Lin SJ (2008) Parallel crawling and capturing for on-line auction. In: Yang et al (eds) ISI 2008 workshops. LNCS 5075, pp 455–466

A Study on the Use Intention of After School Teachers Using Interactive e-Learning Systems in Teaching

Chih-Ching Ho and Horng-Twu Liaw

Abstract With the development of information technology, many industries imported information system to improve the efficiency and output of work. In recent years, interactive digital learning has become the focus of future development projects domestic educational institutions teaching. The after-class school entrepreneurs to import the interactive e-Learning technology system to assist teachers in teaching. Because the overlap between functionality of the interactive e-Learning technology system and after-class schools teachers, those teachers are using the system would therefore be excluded assisted instruction. When the after-class school teachers use this system for assisted teaching, the factors which affecting their willingness is similar with the factors which affecting to other users suffered new information systems imported in other area or not, is also an important issue in this thesis is wish to explore and to discuss. In this paper, through literature analysis and discussion, the use of technology acceptance model 3 as the theoretical model of this thesis. This thesis collection and research framework and measure dimensions scales converted into suitable investigation cram school teacher for the acceptance of the interactive e-Learning technology system questionnaire. In this thesis, teachers who use the system to support teaching cram as research object, paper questionnaires distributed and collected manner survey and sampling data were described in the statistical analysis of samples, mean test, reliability and validity analysis, regression analysis, to verify the impact of the various facets of each other. The paper hope through this research can help to complement modern e-Learning education industry development. The social event collection and analysis system developed by Institute for Information Industry. The system can collect more than 30,000 web data items per day for the government to understand public opinion on policy and for companies needing business insights or seeking to provide exposure for their brands on the Internet.

Keywords After-class school teacher · Interactive e-Learning technology system · Technology acceptance model 3

C.-C. Ho (✉) · H.-T. Liaw

Department of Information Management, Shih Hsin University, Taipei, Taiwan
e-mail: Giniho914@gmail.com

1 Introduction

In recent years, several hardware environment matures, the Ministry of Education to promote the pace accelerated-learning program, under this spacetime have been promoting innovation and change software on promoting cooperation between industry and academia, the private sector has also led to fill teaching industry, the rise of a wave of construction of the boom-learning environment, where “interactive” learning is popular in recent years the focus of attention, but also continue to shape future learning. Therefore, interactive e-learning has become the development spindle future domestic educational institutions teaching.

Under this background the present study, after study supplement industry introducing interactive teaching-learning system, investigate whether teachers make teaching industry would thus exclude the use of the secondary education system, to further explore the factors that influence the will of teachers and other fields whether imported new differences in the user’s information systems.

Based on the above motivation, purpose of this study set as follows: First, the collected and aggregated up to teach business interactive digital learning technology systems development status (Fig. 1).

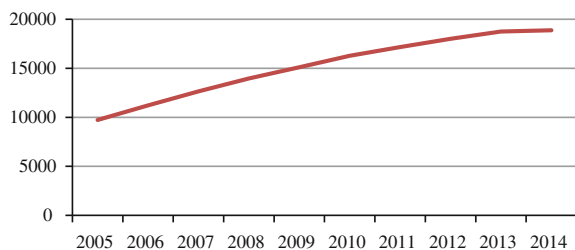
1.1 Purpose and Scope of the Study

According to the research purposes, to establish the scope of this study are as follows: First, fill teach teachers to receive the impact and complement industry sector teachers teach cognitive and behavioral interactive digital learning science and technology system [1].

Based on the above motivation, purpose of this study set as follows:

First, the collected and aggregated up to teach business interactive digital learning technology systems development status.

Fig. 1 National cram the last decade the growth statistics. Source municipalities, counties and various short-term remedial information management system (2014)



2 Review

This chapter discuss today's digital technology system, to understand the user and cognitive differences on their operations, explore the background and knowledge, but research over the years, scholars have completed the reference mechanism, procedures and expect to understand the current e-learning and technology acceptance Model studies [2, 3] and other related issues to do literature review, and sorted out the theory of architecture as a conclusion of this chapter.

3 Research Methods

This paper aims to introduce hypotheses interactive e-learning technology systems to make teaching business students learning the satisfaction of the effectiveness of the impact of the adoption of the Institute, the research process and research methods, and instructions before measurement results of this study in the last section of this chapter.

3.1 Design

In this study, using a complement of interactive e-learning technology systems to teach business teacher, the investigation by questionnaire, please fill teach business teacher in accordance with current practical experience and feelings of interactive e-learning technology systems, and through questionnaires filled reflect its subjective perception, according to the sampling and analysis of the present study. The research by pooling and bit of collating information, will gather to sample survey data analysis and collation of data, and according to the aggregated conclusions of this study and provide recommendations related to future research.

The study hypothesis based TAM 3 and interactive e-learning technology systems characteristics, use the facets have for subjective norms, image, job relevance, output quality, the results show sex, computer self-efficacy, perceived external control, computer anxiety, computer joy perception entertainment, perceived ease of use, perceived usefulness, behavioral intentions and actual use of a total of 14 facets combined propose a hypothesis to assume 16 described as follows:

Technology Acceptance Model 3:

- H1 User perceived subjective norms on behavioral intentions interactive e-learning technology systems are significantly correlated.
- H2 User perceived subjective norms on the image of interactive e-learning technology systems are significantly correlated.
- H3 User perceived subjective norms perceived usefulness of interactive e-learning technology systems are significantly correlated.

- H4 User perception of the image of the interactive digital learning perceived usefulness of the technology systems are significantly correlated.
- H5 User awareness of the work associated with perceived usefulness of interactive e-learning technology systems are significantly correlated.
- H6 News output quality user perception of perceived usefulness of interactive e-learning technology systems are significantly correlated.
- H7 The results demonstrate the possibility of interactive user perception perceived usefulness of e-learning technology systems are significantly correlated.

The study hypothesized that eight to twelve assumptions discussed according to TAM 3 scholars Venkatesh and Davis [4, 5] of the mentioned:

- H8 The user's computer self-efficacy of interactive e-learning technology systems of perceived ease of use has a significant correlation.
- H9 Perception external users control over the perceived ease of use of interactive e-learning technology systems are significantly correlated.
- H10 User's computer anxiety interactive digital learning perceived ease of use are significantly related to science and technology system.
- H11 User's computer joy of perceived ease of use of interactive e-learning technology systems are significantly correlated.
- H12 The user's perception of the fun of interactive e-learning perceived ease of use are significantly related to science and technology system."

According to TAM, TAM2 and TAM3 both retain the relevance of perceived usefulness and perceived ease of use and behavioral intention between, this study presents the following hypothesis thirteen to sixteen hypothesis:

- H13 User's perceived ease of use of interactive e-learning perceived usefulness has significantly related technology systems.
- H14 User behavioral intentions perceived usefulness of interactive e-learning technology systems are significantly correlated.
- H15 The perceived ease of use of user behavior intentions interactive e-learning technology systems are significantly correlated.
- H16 User behavioral intentions for practical use interactive e-learning technology systems are significantly correlated.

The study was analyzed after TAM3 relevant theoretical literature found unsuitable for the purpose of usability measuring questionnaire, so the variables discarded.

3.2 Research Facets and Operational Definition

This section describes the operational definition of each facet of the study in interactive digital learning technology systems design of the questionnaire and

measure the adoption of variables. After the pre-test questionnaires were collected, using IBM SPSS 20 software analyzes each facet reliability. In this study, Cronbach’s α be assessed in various facets of the front part of this study Cronbach’s α were measured falls between 0.8 and 0.9, indicating the degree of good faith before the test. So you can learn all the questions of this questionnaire have a high level of confidence can be measured accordingly. Data collection and analysis.

The study mainly in interactive digital learning technology systems user questionnaire distributed to the parent. For the sake of accurately select the appropriate respondents to exclude invalid sample, the use of sampling methods purposive sampling method to entities questionnaire survey.

Questionnaire data of this study, taking into consideration variables to measure the relevance of scale and analysis tools, the basic analysis using IBM SPSS 20 software questionnaire reliability and validity analysis, descriptive statistics analysis, factor analysis, variance analysis; analysis of the overall pattern of using regression analysis.

4 Data Analysis

This chapter is divided into six sections, the first description of the official release and recovery research questionnaire analysis, Sect. 2 sample basic data analysis, rational analysis of the third quarter, the fourth quarter of factor analysis, single factor V mutation and analyze the relationship between the number of respondents to those basic t-test analysis with various facets of variables, Sect. 4 of regression analysis.

4.1 Questionnaires Were Collected and Analyzed

In this study, the teachers make teaching industry, in April 1, 2014–April 31, 2014, a total of 450 paper questionnaires, 436 questionnaires, the recovery was 96.8 %, 21 parts of the questionnaire after deducting invalid questionnaires 415, the effective rate was 95.1 %. According to the sample’s gender, age, education, job title, years of teaching experience, weekly basic information system time according to the number of samples and after the share consolidation ratio as shown: (Tables 1, 2, 3, 4 and 5).

Table 1 Sample sex ratio

Sex	Times	Gender percentage of times
Male	87	21.0
Female	328	79.0

Table 2 Sample-age

Age	Times	The percentage of the number of age
22 years of age	39	9.3
22–31	222	53.5
32–41	116	27.8
42–51	19	4.7
over 51 years	19	4.7

Table 3 Proportional sample education

Educational attainment	Times	The percentage of the number of educational attainment
Specialist	48	11.6
University	319	76.8
Above institute	48	11.6

Table 4 Sample the proportion of time each week to use the system

The teaching years	Times	The percentage of the number of funded teaching
Less than 1 year	58	14.0
1- less than 3 year	145	34.8
3- less than 5 year	116	27.9
5- less than 7 year	9	2.3
7- less than 9 year	29	7.0
Under 9 years	58	14.0

Table 5 Sample KMO and Bartlett

Research framework	KMO	Bartlett's	Distinctiveness
Subjective norms	0.850	121.010	0.000
Image	0.612	64.707	0.000
Related work	0.628	84.689	0.000
Output quality	0.628	60.766	0.000
The results demonstrate the possibility of	0.802	154.149	0.000
Computer selfefficacy	0.686	68.594	0.000
Perceived external control	0.836	118.155	0.000
User anxiety	0.778	95.196	0.000
User joy	0.802	73.953	0.000
User fun	0.763	126.420	0.000
Perceived usefulness	0.856	133.546	0.000
Perceived ease of use	0.788	96.234	0.000
Behavioral intentions	0.764	94.441	0.000

4.2 Reliability and Validity

The questionnaire for the questions of the reliability of each measurement performed reliability analysis, evaluation index using the internal consistency coefficient (Cronbach's Alpha) to measure the variables of each facet, Each Facets of Cronbach's α values ranged from 0.832 to 0.952, greater than 0.8, a high reliability of the range, so each facet are a considerable degree of internal consistency.

4.3 Mean Assays

In this study, t test between gender and whether each facet will produce significant sexual difference, thereby to understand gender using the t test and ANOVA test t test no significant effect ANOVA test Age and output quality, the results demonstrate the possibility of computer joy, behavioral intentions, there are significant, multiple comparisons Scheffe method showed no significant differences in educational attainment with dimensions no significant effect between groups years of teaching experience and perceived ease of use only facet, there are significant, multiple comparisons Scheffe method showed no significant difference between the groups was no significant difference in funding between the groups.

4.4 Factors

In this study, KMO and Bartlett's test data are spherical test each fall acceptable range, a representative sample of the correlation matrix has a common factor, it is suitable for factor analysis. Factor analysis: All facets are extracted only one factor. The survey questionnaire questions of construct validity of the various facets quite good.

4.5 Regression Analysis

In this study, the overall pattern of development is TAM 3 main study architecture, and based on previous academic studies on various facets make the appropriate adjustments. Working copies of the study reference scholar Yang [6] and academics Wu and Tu [7] regression (Multiple Regression) analysis of cognitive processes associated auxiliary facets, output quality dimensions, according to the results certified TAM 3 and interactive e-learning System features the use of a total of 14 facets facet, after 16 hypothesis regression test, all the assumptions hold, and there is no collinearity.

In order to enhance the complementary teaching sector teachers use results, and to allow use to fill interactive e-learning technology systems to teach the industry understand the current interactive digital learning technology systems actual use situation, the present study architecture through data analysis to verify the sample data, The results confirm the hypothesis of this study are true.

5 Conclusions

This paper on the results of data analysis presented in this study shows that the first quarter will show the relevant conclusions of this study; as described in Sect. 2 of this study and recommend future research directions.

This study is to understand the complement of teachers to teach industry acceptance for interactive digital learning technology systems, the use of technology acceptance model TAM3 investigate subjective norms, image, associated with the work, output quality, the results demonstrate the possibility of computer self-efficacy, perceived external control, Correlation between computer anxiety, computer joy, pleasure perception, perceived usefulness, perceived ease of use, behavioral intentions and actual use and other facets.

The study was collecting and sampling and actually achieved up to teach business teacher interactive e-learning technology systems assist the evaluation and cognitive experience of its teaching, verify the Technology Acceptance Model TAM3 use of interactive e-learning technology systems in up to teach business teacher The teaching also apply on behalf of up to teach business teacher or not, there is no significant difference in the presence of other users in the field of information systems projects for the consideration of the use of information systems, information systems may mean for other areas by the user user acceptance, Model measure and assess the satisfaction and loyalty, there is a great probability applicable complement of teachers to teach industry on cognitive and behavioral information systems, this paper provides a follow-up studies scholars reference.

References

1. Wu M, Tu J (2011) Digital teaching children's language usability of the system-live interactive english center as an example, Shih Hsin University Department of Information and Communication master;s Thesis
2. Fishbein M, Ajzen I (1975) Belief, attitude, intention and behavior: an introduction to theory and research. Addison-Wesley, Reading
3. Davis FD (1986) A technology acceptance model for empirically testing new end-user information systems: theory and results. Doctoral Dissertation, MIT Sloan School of Management
4. Venkatesh V, Davis FD (1996) A model of the antecedents of perceived ease of use: development and test. *Decis Sci* 27(3):451–481

5. Venkatesh V, Davis FD (2000) Theoretical extension of the technology acceptance model: four longitudinal field studies. *Manage Sci* 46(2):186–204
6. Yang S (2012) *SPSS Statistical analysis*. (Acer Feng INFORMATION INC)
7. Wu M, Tu J (2006) *SPSS and statistical applications*, 2nd edn. Wu Nan Book Publishing Co., Taipei

Bibliometric Analysis of Emerging Trends in High Frequency Trading Research

Jerome Chih-Lung Chou, Mike Y.J. Lee and Chia-Liang Hung

Abstract This study reviews and demonstrates the diverse issues and findings in the research field of high frequency trading. This diversity may root from the emerging nature of computing technology and its wide appeal as well as unique researcher and practitioner viewpoints. The authors propose Bibliometric Analysis might be used to identify some fruitful research opportunities.

1 Introduction

As the stock market has become nearly exclusively electronic, advances in computer technology and automated algorithm trading have speeding the transmission and execution of security transaction orders, and establishing High Frequency Trading (HFT). History of HFT can be traced back at least since 1998, after the U.S. Securities and Exchange Commission (SEC) adopted Regulation ATS (Alternative Trading Systems), including electronic exchanges. After that, SEC's Regulation NMS (National Market System), which was adopted in 2005, further provided strong incentives for trading venues to automate, especially the NYSE, which was the last

J.C.-L. Chou
Department of Management Information Systems, Hwa Hsia University
of Technology, New Taipei City, Taiwan
e-mail: jerome@cc.hwh.edu.tw

M.Y.J. Lee (✉)
Department of Management Information Systems, National Chengchi University,
Taipei, Taiwan
e-mail: yjlee@cute.edu.tw

M.Y.J. Lee
Department of Business Administration, China University of Technology,
Taipei, Taiwan

C.-L. Hung
Department of Information Management, National Chi Nan University,
Nantao, Taiwan
e-mail: clhung@ncnu.edu.tw

major floor-based exchange in the U.S. Until 2010, SEC issued a Concept release [1] seeking public comments on issues such as HFT. SEC admitted, “The term (HFT) is relatively new and is not yet clearly defined”.

HFT was not a well-known topic outside the financial sector, until an article [2] published by the New York Times in July 2009 which was one of the first to bring the subject to the public’s attention. HFT has radically changed the stock markets. Some view the Flash Crash of May 6, 2010 as evidence of the potential harmful effects of HFT. Michael Lewis’s recent book in 2014, *FLASH BOYS*, even raised the controversy concerning about HFT by pointing out that electronic trading has rigged the market against ordinary investors, particularly in America, but Lewis paid little attention to the market benefits of HFT. The development of HFT has ignited a heated debate among participants, researchers and regulators about the benefits and concerns related to HFT.

2 Diverse Issues of HFT Research

Since history of HFT is not so long, the researches in this area are highly diverse regarding issues for this new research agenda. According to a recent review article by editor Goldstein [3], unlike established topics in finance, such as dividend policy, capital structure, or asset pricing, HFT is a new, emerging, and rapidly evolving area for the markets, regulators, and the public.

In more recent years, many attempts have been made to research HFT by a number of scholars. Goldstein et al. [4] reviewed the works in this area from the empirical and theoretical papers and assigned them into following six categories on the basis of different topics: market performance [5–15], strategies and practices [16–18], evolution [11, 17, 19, 20], speed [14, 18, 21–23], fairness [1, 24, 25], regulatory implications [26, 27].

Table 1 summarizes findings by topics about each category which include the current state, topics of debate, and empirical and theoretical researches in HFT.

3 Bibliometric Analysis

While there are plenty of findings by topics shown by Table 1 which include the topics of debate, controversial issues, and the current arguments against HFT, however, those findings about HFT are largely qualitative in nature. Based on previous research, merely depending on review articles cannot completely reveal the developmental trends or future orientation of a new research field. Despite the high growth rate of publications, there have been few attempts to gather systematic data on the global scientific production of research on HFT.

A quantitative research tool to fill this gap is the bibliometric method, which has already been widely applied in many disciplines of science and engineering

Table 1 Research categories and findings in HFT

Category	Authors	Findings by topic
• Market performance	• Budish et al. [5] and Menkveld [6] and Schwartz and Wu [7]	<ul style="list-style-type: none"> • Adverse selection <ul style="list-style-type: none"> – HFT’s “socially wasteful arms race” could disadvantage other ordinary investors and then reduce market quality, as measured by liquidity and price informativeness
	• Jarnecic and Snape [8]	<ul style="list-style-type: none"> • Liquidity <ul style="list-style-type: none"> – If HFT activity can improve the liquidity of markets?
	• Brogaard et al. [9] and Hendershott and Riordan [10]	<ul style="list-style-type: none"> • Market structure <ul style="list-style-type: none"> – Market structure changes due to the incremental effect of algorithmic trading and HFT
	• Popper [11]	<ul style="list-style-type: none"> • Transaction costs <ul style="list-style-type: none"> – As the recent volume of HFT has decreased, the benefits of HFT in reducing trading costs for ordinary investors have stalled
	• Baron et al. [12]	<ul style="list-style-type: none"> • Profitability (of HFT vs. non-HFT) <ul style="list-style-type: none"> – HFT profits are earned at the expense of other traders – HFT markets are effectively a “zero sum game”
	• Jones [13]	<ul style="list-style-type: none"> • Volatility <ul style="list-style-type: none"> – After surveying 30 theoretical and empirical papers on the topic of HFT, Jones [13] concludes that HFTs are making markets better
	• Hasbrouck and Saar [14]	<ul style="list-style-type: none"> – The impact of HFT on market quality and volatility
	• Credit Suisse [15]	<ul style="list-style-type: none"> – By the findings that long-term volatility in recent years remained within historical norms, while short-term volatility declined, Credit Suisse concludes that markets are “not worse” for the presence of HFT
• Strategies and practices	• Aldridge [16]	<ul style="list-style-type: none"> • Market efficiency
		<ul style="list-style-type: none"> • Algorithmic strategies <ul style="list-style-type: none"> – Statistical arbitrage strategies – Directional trading around events – Automated market making – Modeling information in order flow – Latency arbitrage – Spread scalping – Rebate capture – Quote matching – Layering • Market manipulation <ul style="list-style-type: none"> – Pinging/sniping/sniffing/phishing – Quote stuffing

(continued)

Table 1 (continued)

Category	Authors	Findings by topic
		<ul style="list-style-type: none"> – Spoofing – Pump-and-dump – Ignition
	• Kirilenko and Lo [17]	<ul style="list-style-type: none"> • Manipulative trading activities <ul style="list-style-type: none"> – ‘Order anticipation’ trading strategy (e.g. a “pinging” tactic to discover the price other traders are willing to pay or to discover undisplayed liquidity)
	• Laughlin et al. [18]	<ul style="list-style-type: none"> • Low-latency strategies <ul style="list-style-type: none"> – Many HFT firms are concerned about transmission speed across geographic distances and utilize strategies that capitalize on their geographic location • Co-location <ul style="list-style-type: none"> – The ability to access direct data feeds from exchanges which includes sophisticated order execution algorithms services
• Evolution	• Rubenstein [19]	<ul style="list-style-type: none"> • Trading volume <ul style="list-style-type: none"> – HFT now accounts for almost 50 % of daily stock trades
	• Goldstein et al. [20] and Kirilenko and Lo [17]	<ul style="list-style-type: none"> • Trading activity <ul style="list-style-type: none"> – HFT accounted for between 40 % and 60 % of trading activity across all U.S. financial markets for stocks, options and currencies
	• Popper [11]	<ul style="list-style-type: none"> • Volumes and profits downtrend in US <ul style="list-style-type: none"> – HFT volume down from 61 % in 2009 to 51 % in 2012 – HFT profits were estimated at most \$1.25B in 2012, down 35 % from 2011 and 74 % lower than the peak of about \$4.9B in 2009
• Speed	• Angel [21]	<ul style="list-style-type: none"> • Data transmission <ul style="list-style-type: none"> – Physical limitations on current trading due to Einstein’s theories and related quantum physics to finance
	• Laughlin et al. [18] and Wissner-Gross and Freer [22]	<ul style="list-style-type: none"> • Data transmission <ul style="list-style-type: none"> – Techniques to minimize transmission delays and execution latencies and affected price discovery when HFT firms trade securities in different locations around the world
	• Brogaard et al. [23] and Hasbrouck and Saar [14]	<ul style="list-style-type: none"> • Technology upgrades <ul style="list-style-type: none"> – Exchanges upgrading for lower latencies
• Fairness	• SEC [1]	<ul style="list-style-type: none"> • Market structure <ul style="list-style-type: none"> – The SEC [1] concept release directly questions the fairness of the current market structure, HFT, and the use of a variety of HFT tools and strategies

(continued)

Table 1 (continued)

Category	Authors	Findings by topic
		<ul style="list-style-type: none"> • Unfair access concerns <ul style="list-style-type: none"> – The SEC [1] concept release directly questions if co-location provides HFTs an unfair advantage because of greater resources and sophistication to take advantage of co-location services than other market participants, including long-term investors?
	• Narang [24]	<ul style="list-style-type: none"> • Rebate structure <ul style="list-style-type: none"> – If the current rebate structure based on volume unfairly benefits HFT firms over non-HFT firms?
	• Patterson et al. [25]	<ul style="list-style-type: none"> • Insider advantages <ul style="list-style-type: none"> – HFTs are using a hidden facet of the Chicago mercantile exchange's computer system to trade on the direction of the futures market before other investors get the same information
• Regulatory implications	• Piwowar [26]	<ul style="list-style-type: none"> • The role of speed <ul style="list-style-type: none"> – SEC Commissioner called for a comprehensive review of U.S. markets which should examine the role of speed in the markets
	• SEC/CFTC (2010)	<ul style="list-style-type: none"> • Market-makers and liquidity providers <ul style="list-style-type: none"> – Whether HFT market-makers should be subject to regulations that would require them to stay active in volatile markets, rather than deserting the markets en masse and damaging liquidity
	• Westbrook [27]	<ul style="list-style-type: none"> • Concerns <ul style="list-style-type: none"> – Lawmakers questioned whether the HFT practice is benefiting wall street at the expense of individual investors

[28, 29]. Since it is a new, emerging, ever changing and rapidly evolving area for the markets, regulators, and the public, this study try to provide a quantitative analysis of global empirical and theoretical HFT papers.

In this study, a traditional bibliometric method will be used to describe the latest advances in HFT. The Web of Science (WOS), which includes Science Citations Index Expanded (SCIE) and Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI) from the Institute of Scientific Information (ISI) Web of Science databases, is the most important and frequently used source for a broad review of scientific accomplishment in all fields.

Expected findings from these investigations can help researchers to realize the breadth of HFT research and to establish future research directions and to provide an entry point to any academic, regardless of their prior knowledge of the topic.

Specifically, this study will quantitatively analyze existing empirical and theoretical HFT papers to address the following objectives:

1. To reveal the developmental trends or future orientation of a new research field like HFT.
2. To get different point of view that emerge from a comprehensive review of existing HFT papers by future researchers before choosing their interested field.
3. To help to evaluate the need for regulatory intervention or regulatory purposes.

References

1. U.S. Securities and Exchange Commission (2010) Part III: concept release on equity market structure; proposed rule, 17 CFR Part 242, Federal Register vol 75, no 13, pp 3594–3614. 14 Jan. <https://www.sec.gov/rules/concept/2010/34-61358.pdf>
2. Duhigg C (2009) Stock traders find speed pays, in milliseconds. The New York Times, July 23. <http://www.nytimes.com/2009/07/24/business/24trading.html>
3. Goldstein MA (2014) Special issue on computerized and high-frequency trading: guest editor's note. *Financ Rev* 49(2):173–175
4. Goldstein MA, Kumar P, Graves FC (2014) Computerized and high-frequency trading. *Financ Rev* 49(2):177–202
5. Budish EB, Cramton P, Shim JJ (2015) The high-frequency trading arms race: frequent batch auctions as a market design response. Chicago Booth Research Paper, No. 14-03
6. Menkveld AJ (2014) High-frequency traders and market structure. *Financ Rev* 49(2):333–344
7. Schwartz R, Wu L (2013) Equity trading in the fast lane: the staccato alternative. *J Portfolio Manage* 39(3):3–6
8. Jarnecic E, Snape M (2014) The provision of liquidity by high-frequency participants. *Financ Rev* 49(2):371–394
9. Brogaard J, Hendershott T, Riordan R (2014) High-frequency trading and price discovery. *Rev Financ Stud* 27(8):2267–2306
10. Hendershott T, Riordan R (2013) Algorithmic trading and the market for liquidity. *J Financ Quant Anal* 48(4):1001–1024
11. Popper N (2012c), High-speed trading no longer hurtling forward. The New York Times, Oct 14
12. Baron M, Brogaard J, Kirilenko A (2012) The trading profits of high-frequency traders. Working paper, University of Washington
13. Jones C (2013a) What do we know about high-frequency trading? Columbia Business School Research Paper, no. 13–11
14. Hasbrouck J, Saar G (2013) Low-latency trading. *J Financ Markets* 16(4):646–679
15. Credit Suisse (2012) Who let the bots out? Market quality in a high frequency world. <https://www.managedfunds.org/wp-content/uploads/2012/11/HFT.pdf>
16. Aldridge I (2013) High-frequency trading: a practical guide to algorithmic strategies and trading systems, 2nd edn. Wiley, Inc., Hoboken
17. Kirilenko AA, Lo Andrew W (2013) Moore's Law versus Murphy's Law: algorithmic trading and its discontents. *J Econ Perspect* 27(2):51–72
18. Laughlin G, Aguirre A, Grundfest (2014) Information transmission between financial markets in Chicago and New York. *Financ Rev* 49(2):283–312
19. Rubenstein A (2015) This high-speed trader says thanks, regulators. *Wall Street J.* <http://www.wsj.com/articles/this-high-speed-trader-says-thanks-regulators-1429832042>. (April 23, C1)

20. Goldstein MA, Kumar P, Graves FC (2014) Computerized and high-frequency trading. *Financ Rev* 49(2):177–202
21. Angel JJ (2014) When finance meets physics: the impact of the speed of light on financial markets and their regulation. *Financ Rev* 49(2):271–281
22. Wissner-Gross AD, Freer CE (2010) Relativistic statistical arbitrage. *Phys Rev E* 82:056104
23. Brogaard J, Hendershott T, Hunt S, Ysusi C (2014) High-frequency trading and the execution costs of institutional investors. *Financ Rev* 49(2):345–369
24. Narang M (2010) Tradeworx, Inc. Public commentary on SEC market structure concept release. SEC Comment letter. <https://www.sec.gov/comments/s7-02-10/s70210-129.pdf>
25. Patterson S, Strasburg J, Plevin L (2013) High-speed traders exploit loophole. *Wall Street J*. <http://www.wsj.com/articles/SB10001424127887323798104578455032466082920>. (May 1)
26. Piwowar M (2013) The benefit of hindsight and the promise of foresight: a proposal for a comprehensive review of equity market structure. U.S. Securities and Exchange Commission, Speech in London, England. <http://www.sec.gov/News/Speech/Detail/Speech/1370540470552>. (Dec. 9)
27. Westbrook J (2010) SEC vote shows scope of high-frequency trading rules (Update 2). *Bloomberg*. <http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aiP2T5Sx9Nq4>. (Jan 13)
28. Fahimnia B, Sarkis J, Davarzani H (2015) Green supply chain management: a review and bibliometric analysis. *Int J Prod Econ* 162:101–114
29. Hsu C, Chiang C (2015) A bibliometric study of SSME in information systems research. *Scientometrics* 102(3):1835–1865

Interactive Performance Using Wearable Devices: Technology and Innovative Applications

Tzu-Chieh Tsai, Gon-Jong Su and Chung-Yu Cheng

Abstract This paper presents wearable computing in the Art and interactive performance. With AR/VR technology, human can play with virtual characters. However, people might feel bored to type keyboard or joysticks to play with computer and video game characters. We develop our motion sensors to capture user's motion. In this way, it becomes more interesting and fresh to swing your hands or shake your hips to interact with virtual characters. In the future, the wearable computing will be more and more popular, so we try to merge these sensors to let everyone can wear these tiny and cute sensors to experience. We believe this will be a new bright spot of the world.

1 Introduction

In recently years, wearable items grow more and more popular, such as iWatch, google glass, and so on. With these wearable items, many interesting things can be fulfilled. For example, there are many art performances using virtual characters and various lights to create amazing performance [1]. To perform this amazing show, the performers need a lot of time to practice. It is innovative to let performers wear sensors to make the effect on acousto-optic interaction. This inspires us to decide to combine wearable sensors with art. In this paper, we present our idea about how to use wearable sensors to combine with art performance. We surveyed many wearable sensors in the market to ensure whether the sensors can work or not, and implemented a platform to deliver sensing data in real time. We named the platform as Wearable Item Service runtimeE (WISE) which coordinates all messages. As long

T.-C. Tsai (✉) · G.-J. Su · C.-Y. Cheng
Department of Computer Science National Chengchi University, Taipei, Taiwan
e-mail: ttsai@cs.nccu.edu.tw

G.-J. Su
e-mail: 102971022@nccu.edu.tw

C.-Y. Cheng
e-mail: 102753022@nccu.edu.tw

as wearing our wearable sensors, the skeleton (posture) information from the user will be delivered to WISE in real-time to render animation. Our other team members use the AR/VR to present the virtual characters which was designed by professionals. However, the wearable sensors still have some problems about its appearance and instability. To this end, we tried step by step to miniature our sensors and overcome the problems. These problems will be introduced and solved in following sections.

2 Related Work

A successful design of interactive performance is not only the stability of wearable sensors but also the interactive experiment from users. For interactive story, the users' feeling is very important for overall benefit assessing. In 2000, Csikszentmihalyi [2] proposed the flow theory, the immersive moment when a person is completely involved in an activity for its own sake. Therefore we design wearable sensors to hope to let user in the "flow"—a state of heightened focus and immersion in activities such as art, play and work. Also, we try to let audiences wear our sensors to participate and perform in digital live. This concept is based on Sheridan [3] who proposed the ideals to make audiences participate in the performance.

MIThril (2003) is the wearable computing research platform [4] from MIT. The design core of the architecture is "Enchantment Whiteboard" which is used to integrate and coordinate the wearable sensors from users. Our team designed the platform, Wearable Item Service runtime (WISE), to integrate wearable sensors and coordinate to work in the performance.

What the performance needs is the posture of the performances. Therefore, we decide to survey sensors with gyro and accelerate to get skeleton (postures) information. Then, we find the 6-axis accelerometer and gyroscope sensors, named MPU6050, with the library of the scholar Rowberg [5], to be used in our performance, and the arduino nano with BLE [6] to transmit signals. Besides, the location of the user is another issue in the performance. Based on Erin-Ee-Lin Lau's [7] experiment, the Received Signal Strength Indicator (RSSI) can track user's location in real-time for indoor and outdoor environments. Finally, we decide to use the RSSI from Bluetooth low energy and raspberry pi to serve as the coordinator in our platform.

3 Proposed Approach

With the wearable computing, many interactive performances can be fulfilled. By using our platform to connect sensors, the performance will be more interesting and exciting. However, there are still some limitations, we try step by step to overcome.

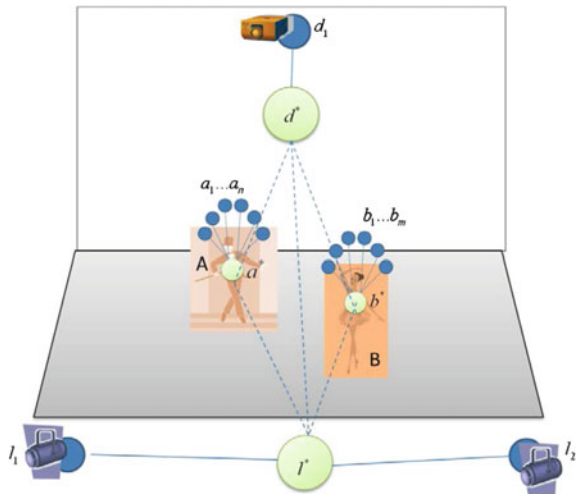
The research is divided into three parts: (A) system architecture, (B) wearable sensors, and (C) interactive storytelling. Details are as follows:

3.1 System Architecture

In the beginning, human wear many small sensors, named WISE Items, as $a_1 \dots a_n$ and $b_1 \dots b_m$ in Fig. 1. In addition, these WISE Items can be used for various applications to meet the needs of the performance. Besides, there is a WISE Coordinator used to integrate all the sensors. Due to these wearable sensors must be tiny and light, so we use low energy BLE to transmit messages.

From the implementation-level, the difference between WISE item and coordinator is its computation. These tiny WISE items are limited on the resource and size such as Arduino. On the contrary, the coordinator with more ability to compute can operate in the IP layer such as Raspberry pi. In live digital script, there are many issues need to be solved even in the easy interactive scenes. Therefore, we implemented the WISE to integrate message and finish the most important things to deliver data as soon as possible. As shown in Fig. 2, the WISE platform consists of wearable mo-cap sensors called WISE Items (see Fig. 2), a set of protocol gateway devices called the WISE Coordinators, and a message bus called the WISE Broker. The wearable mo-cap sensors detect dancer's motion and deliver the message to the WISE Coordinator by BLE, then send motion messages to the Unity from WISE Broker.

Fig. 1 The architecture of wearable performance



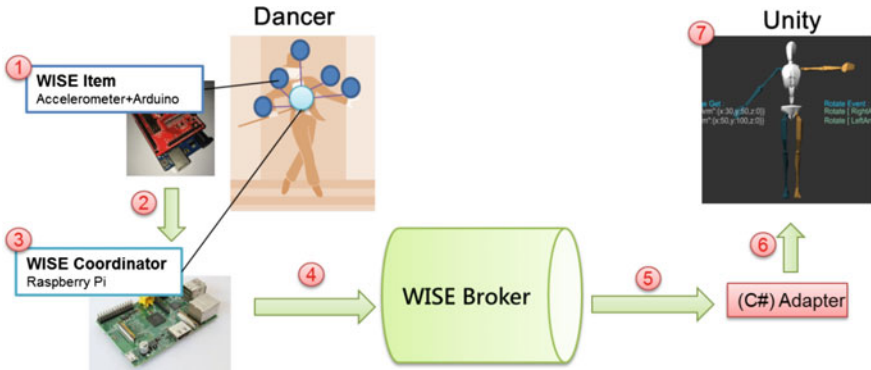


Fig. 2 The integration of wearable sensors

Followings are some items used in the platform:

1. WISE Broker

Broker is the core of the message delivery in the WISE for connecting WISE items to other computers. The message queue is designed to receive and send data as soon as possible to meet the needs of the performance in real-time. Currently, a WISE Item is able to transfer mo-cap data up to 55 messages per second. In live performance, we set the data rate of 33 messages per seconds.

2. WISE API, Adapter

To offer the connection of many different soft and hardware, the WISE must develop relative APIs and Adapters. For example, our team creates animation in Unity, so the WISE have to implement relative Adapter to deliver message between Unity and WISE Broker.

3. WISE platform

In the test, we encountered many difficulties in whether messages are sent to the Unity from WISE or not. In this way, the WISE offers a platform to monitor and simulate the delivery of messages. The platform implements three functions.

First, the MQTT Simulator, which provides graph interface, lets users transmit messages to WISE and receive messages in some topics you are interested as Fig. 3.

Second, the virtual data player, which can replay the real messages from the pre-recorded model, lets users test the efficiency of WISE as Fig. 4.

Finally, the skeleton poster simulator can monitor the effect of the animation under some conditions as Fig. 5.

4. WISE message protocol

The protocol formulates some message formats to deliver well and filters unnecessary data to ensure messages delivery in real-time.

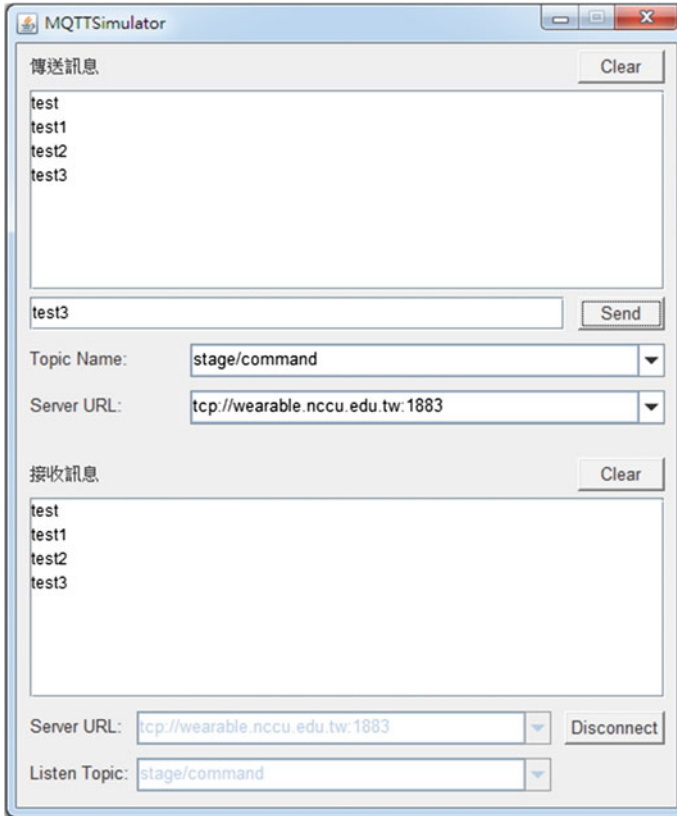


Fig. 3 MQTT simulator

3.2 Motion Capture

At the beginning, we try to get accelerate and gyro signals from the IMU (Inertial Measurement Unit). The raw data of IMU is unstable and has accumulative errors apparently, so we must filter the signals to reduce these errors. Fortunately, Rowberg [5] offers the library of digital motion processor (DMP) to filter noise and provide some useful APIs. In this way, we get the pitch, roll and yaw from IMU to detect the rotation from users.

Second, we chose the raspberry pi (RPi) to serve as our WISE coordinator, which can integrate signals of the IMU. The RPi is tiny, only 85.60×53.98 mm, and high efficient, CPU up to 900 MHz. But how can the IMU connect to RPi? According to the Internet Data Center (IDC) report “Worldwide Bluetooth Semiconductor 2008–2012 Forecast,” [8] Bluetooth low energy (BLE), will link wireless sensors up to the 70 % of cell phones and computers likely to be fitted with the Bluetooth wireless technology. Because of low energy, we use the 3.7 V lithium cell for its power as Fig. 6.

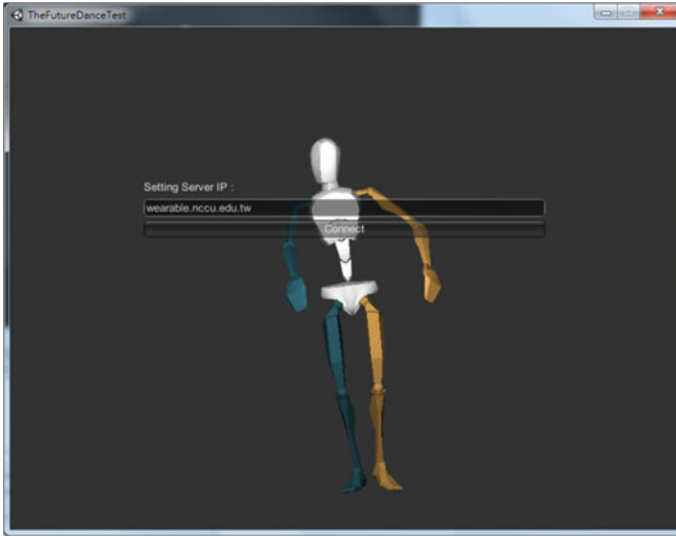


Fig. 4 The virtual data player

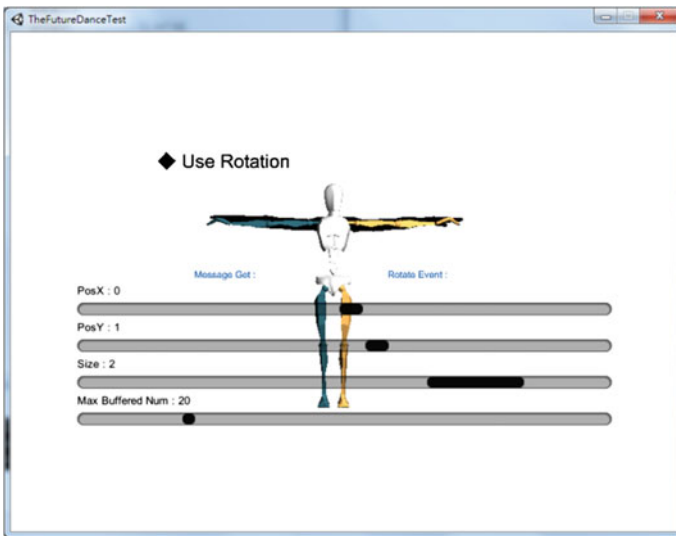


Fig. 5 Skeleton poster simulator

Third, we take the Arduino Uno, BLE and MPU6050 sensors in the market. However, the Arduino Uno is so large that we replace Arduino Uno of Arduino Nano, and use printing circuits board (PCB) layout to merge Arduino Nano and BLE together as the Fig. 7. Besides, it is important to protect the electronic equipment from wet skin, so we design its 3D modeling to create its home by 3D printer.

Fig. 6 3.7 V lithium cell



Fig. 7 The motion sensors: MPU6050, BLE, Arduino Nano, 3.7 V lithium



Finally, there are still some problems from wearable sensors, such as the cumulative errors from MPU6050. These errors are as following values.

The gyroscope signal can be modeled as:

$$Gyro_{total} = G_v + G_b + G_n.$$

where G_v , G_b and G_n are the angular velocity vector, the related angular velocity vector bias, and a white noise.

The acceleration signal can be modeled as:

$$Acc_{total} = A_v + A_b + A_n.$$

where A_v , A_b and A_n are the angular velocity vector, the related angular velocity vector bias, and a white noise.

The DMP from Jeff Rowberg uses the value of the acceleration to get the rotation and merges with the gyroscope to reduce the noise from the bias and white noise. However, it still suffers from cumulative errors even using the library of DMP. To solve these problems, we implemented the calibrated system to map sensors signals to the user's posture as the following: $Y = A \times X + B$

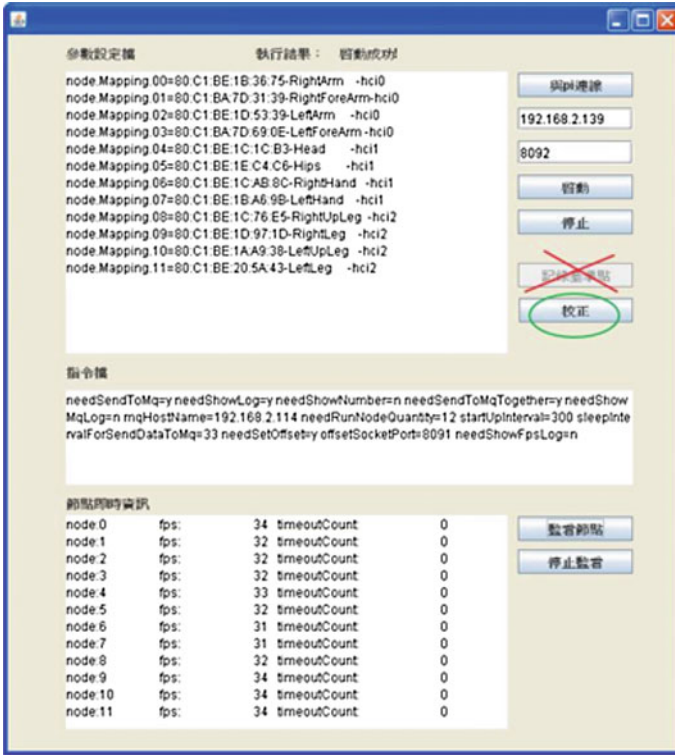


Fig. 8 Graph user interface of Motion monitor

$$\begin{bmatrix} Y_w \\ Y_p \\ Y_r \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \times \begin{bmatrix} X_w \\ X_p \\ X_r \end{bmatrix} + \begin{bmatrix} B_w \\ B_p \\ B_r \end{bmatrix}$$
, which w, p, and r represent yaw, pitch, and roll.

X is the value from our DMP and Y is the value from user's rotation. In addition, we implement a graph user interface to monitor the sensors as Fig. 8.

However, there are still some limitations of hardware that we can't solve, such as some angles can't be detected. Therefore, we put our sensors sticky to users and calibrate sensors after few minutes to reduce the noise from sensors as the Fig. 9.

After getting the user's posture, we still have another issue about how to get user's location.

To avoid the cumulative errors, we use the feature of the RSSI from BLEs, and propose our algorithm to detect the location of users. Due to the RSSI is more precise in 30 cm, we deploy 21 BLEs in 6 m * 2 m stage as in the Fig. 10. Owing to the frame rate of BLE is just only 5–20 fps, the gyro sensors are used to improve its frame rates and offer the credibility of RSSI.

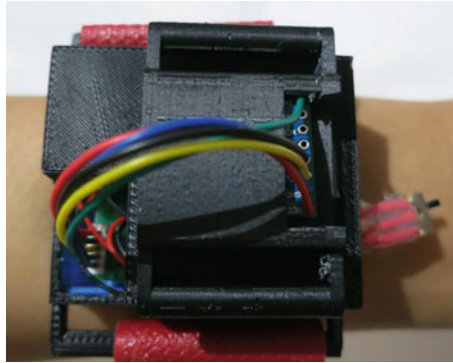


Fig. 9 IMU sensor put sticky on the arm

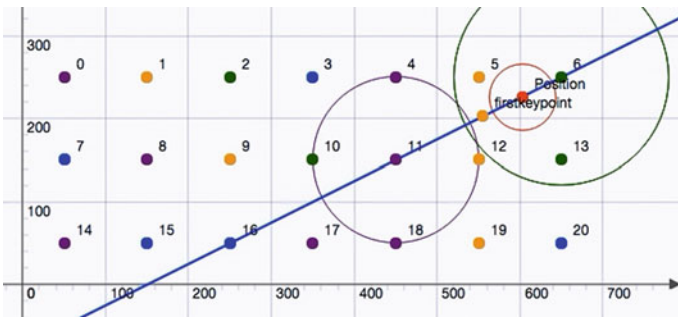


Fig. 10 The 21 BLEs illustration

With the RSSI, our algorithm gives the weight of each BLE from sorted node to get the user position as follows.

First, we sort all beacons by RSSI and get three closest beacons, named BLE₁, BLE₂, BLE₃. Then, we get the reference node (RN) by using weighting on BLE₂ and BLE₃ as following.

$$RN = BLE_2 + \text{Weighting} \times \text{Vector}_{2,3}, \text{ where } \text{Vector}_{2,3} \text{ is from } BLE_2 \text{ to } BLE_3$$

$$\text{Weighting} = \frac{(BLE_2)}{(BLE_2 + BLE_3)}$$

Finally, the nearest BLE is most precise, so we combine its RSSI with RN to ensure user's position. As Fig. 11 and Table 1, using following functions to calculate the intersection of them.

Circle: $(x - x_0)^2 + (y - y_0)^2 = \text{RSSI}^2$, which RSSI is from BLE₁

Line: $y - y_1 = m(x - x_1)$, which Line is from RN to BLE₁

Fig. 11 Localize user's position

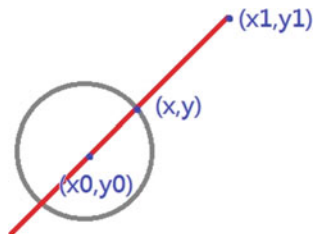


Table 1 The results of the 21 BLEs

Number of BLE	21		
Distance from BLE (cm)	50	30	15
Real position	(100, 250)	(80, 250)	(65, 250)
Calculate value	(85.1, 203.8)	(81, 196.8)	(61.4, 222)
Error (cm)	48.54	53.20	28.23
Graph	<p>The graph shows a 2D coordinate system with x and y axes ranging from -100 to 400. Multiple overlapping circles of various colors (purple, blue, green, yellow, orange, red) represent the localization ranges of 21 different BLE beacons. A central point is marked with a red dot, representing the user's estimated position.</p>		

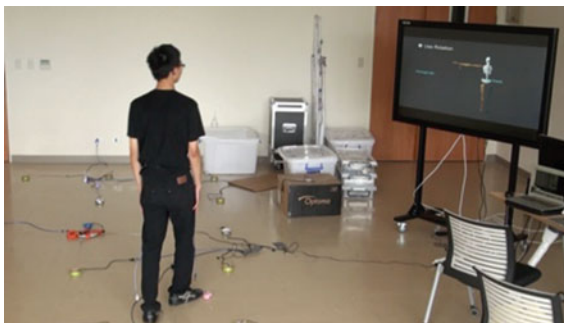
$$\text{Intersection: } X = x_0 + \frac{\text{rssi}(x_0 - x_1)}{\sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}}, Y = \left(\frac{y_0 - y_1}{x_0 - x_1}\right)(X - x_1) + y_1.$$

Besides, the value of the variation from gyro can ensure the user's movement as follows.

$$\text{Position}_{\text{final}} = \text{credibility} * \text{Position}_{\text{RSSI}} + (1 - \text{credibility}) * \text{Position}_{\text{final-1}},$$

where $\text{credibility} = \text{Gyro}_t - \text{Gyro}_{t-1} / (\text{Max_movement})$ as the Fig. 12.

Fig. 12 Human really move indoor and show its results



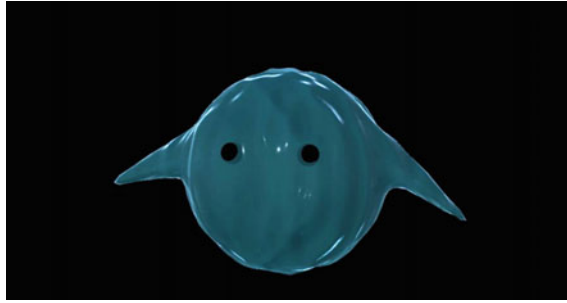


Fig. 13 3D model virtual character—water






<input type="checkbox"/>	Target Name	Type	Rating	Status
<input type="checkbox"/>	 waterco	Single Image	☆☆☆☆☆	Active
<input type="checkbox"/>	 BlueQ	Single Image	★★★★☆	Active
<input type="checkbox"/>	 BlackQ	Single Image	★★★★★	Active
<input type="checkbox"/>	 waterko	Single Image	★★★★★	Active
<input type="checkbox"/>	 waterkoicon	Single Image	★★★★☆	Active

Fig. 14 Design our target figure to improve the recognition

3.3 Interactive Storytelling

To interact with virtual characters, there needs three steps to finish this storytelling. First, our team used 3D model to design a cute virtual role, named Water, in Fig. 13, which presents fresh start and hope from water.

Second, using Vuforia and Unity 3D to make animations and let the user interact via smartphones, 3D smart glasses and our wearable sensors. To improve the recognition from AR, our team have tried many different ways and found that it is hard to recognize from symmetric figures. Therefore, adding the water decoration on one side and water ripple on the bottom can get high recognition. As in Fig. 14, more starts means better.

Third, our team designed following patterns to let user interact with Water and create short films to introduce the story in Table 2.

Table 2 Patterns and reaction of the AR

Step	Action	Water react	Describe
1	Hand push down	Water is pushed in the water	By doing this to trigger the animation
2	Hands push left and right	Water spreads	Use the ripple from unity to create animation (flat and spread)
3	Hand push toward	Water moves back	Interact with users by user's action (this action means water wants to play with user)
4	Hand push up	Water swept up (such as tornado)	Let water rise to the surface
5	Body wag from side to side	Water swings around	Water follows users curiously

4 Simulation Results

We use the Raspberry pi 2 Model B with OS of the Raspbian as our central controller, and Arduino Nano with MPU6050 and BLEs as wearable sensors. The raspberry pi 2 with two BLE dongles to receive the RSSI from beacons and we use 21 beacons to locate user's position. In addition, the frame rate of MPU6050 is 33 fps, and our wise server can serve up to 55 messages per second. Our team use MQTT (MQ Telemetry Transport) as the protocol in the WISE platform and test WISE Broker to ensure it can transmit message efficiently. We simulate to send 50 messages per second last 20 min, so the wise transmit 60,000 messages. This test proves that average delay time is 0.39 ms and the worst case is just 61 ms via WISE Broker. These delay time is acceptable and the WISE transmits messages high efficiently.

As mentioned before, our WISE can connect to different hardware and programs including, Unity, Arduino, Java, and C#, coordinate to make our performance work successfully. SensorDataRetriever is the API of Java, and it can be used to transmit data from BLE. The JAVA Adapter and the Unity Adapter(C#) can connect from JAVA to WISE and from Unity to WISE. With these adapters, we can send data to WISE, so Unity can receive data from WISE as in Fig. 15.

Finally, we use our wearable sensors to interact with the virtual character. Audiences can use tablets or smart glasses to see the reaction of the virtual character in real time. We perform on campus and invite audiences to wear our sensors to experience in Fig. 16.

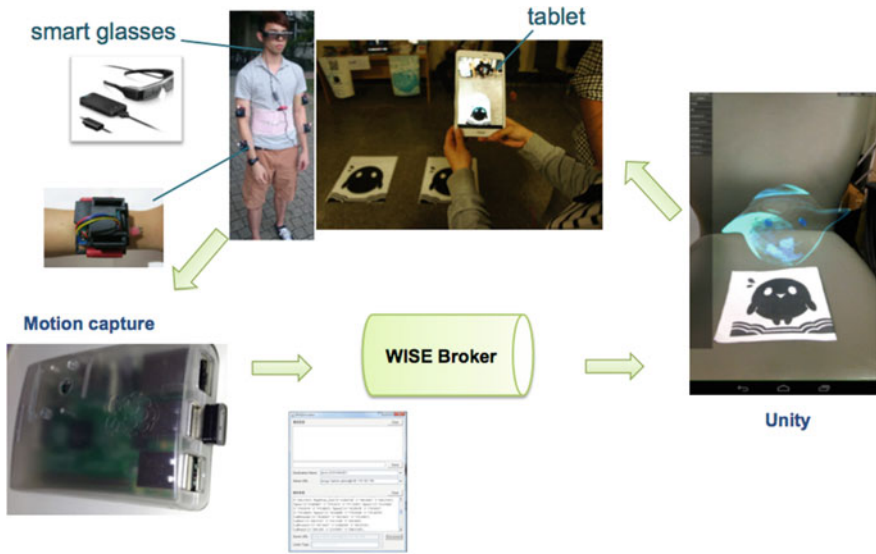


Fig. 15 System architecture: interactive performance is realized by our wearable sensors and the WISE, a platform for interacting our virtual character developed by our research team

Fig. 16 Audiences use our wearable sensor to interact



5 Conclusions and Future Work

We combine many wearable items and implement the WISE server to perform our interactive performance in campus successfully. Besides, we encountered some problems such as error cumulative from sensors, so we develop the calibration to solve. In the future, many users can wear our sensors and interact with not only AR/VR but also other people. We hope that for one day, users can play together by our sensors in different places. Finally, special thanks to our team members.

References

1. Britain's Got Talent. Available: <https://www.youtube.com/watch?v=A7IMKWvyBn4>
2. Csikszentmihalyi M (2000) Beyond boredom and anxiety: Jossey-Bass
3. Sheridan JG, Bryan-Kinns N, Bayliss A (2007) Encouraging witting participation and performance in digital live, Published by the British Computer Society Art
4. DeVaul R, Sung M, Gips J, Pentland A (2003) MITHril 2003. In: Proceedings of seventh IEEE international symposium on applications and architecture, wearable computers, pp 4–11
5. Rowberg J (2013) I2Cdevlib. MPU-6050 6-axis accelerometer/gyroscope. Accessed 28 May 2014. <http://www.i2cdevlib.com/devices/mpu6050#source>
6. Omre AH, Keeping S (2010) Bluetooth low energy: wireless connectivity for medical monitoring. *Jo Diab Sci Technol* 4(2):457–463
7. Lau EEL, Lee BG, Lee SC, Chung WY (2008) Enhanced RSSI-based high accuracy real-time user location tracking system for indoor and outdoor environments. *Int J Smart Sens Intell Syst* 1(2):534–548
8. Worldwide Bluetooth Semiconductor Revenue to Nearly Double by 2012. Available: <http://www.idc.com/groups/semiconductorfocus/issue15/index.html>

Usability Evaluation of Acoustic-Oriented Services on Mouse Manipulation: Can Manipulation with Dual Senses Be Good?

Chi Nung Chu

Abstract This study explored a dual-sensory user interface design that met the requirements of elderly people with low vision in mouse manipulation. The study showed significant efficiencies in integration of dual senses assisting the difficulties of the unimodal input, specifically the hearing input. The results demonstrate a new conceptualization of integration that is more than just the combination of ability to move with sight, hear, and integrate. The study focused on assessing the implications of difficult unimodal inputs on mouse manipulation as the physiologies of elderly participants are degenerating. The Acoustic Assistance on Cursor Navigation design works with an aural assistance environment to allow the elderly people to select either the appropriate functions of Talking Aid or Cursor Positioning to recognize what objects they are pointing at and to navigate where the mouse cursor they are moving within the computer windows. An advanced level of GPS-like loudspeaker in the computer windows environment, the Acoustic Assistance on Cursor Navigation design can help guide the elderly people with low vision by hearing sensory information to facilitate identifying and navigating where they are in the computer world.

Keywords Acoustic assistance on cursor navigation design · Dual coding theory · Text-to-Speech engine · Acoustic-oriented services

1 Introduction

With increasing computer usage in elderly population, the hindrances resulting from age-related eye disorders to computer mouse manipulation will also increase, which will affect the quality of life in cyberspace [13, 15, 23]. People with vision degeneration, such as cataract and presbyopia, are closely tied to the aging process

C.N. Chu (✉)

Department of Management of Information System, China University of Technology,
No. 56, Sec. 3, Shinglung Rd, Wenshan Chiu, Taipei, Taiwan116
e-mail: nung@cute.edu.tw

that no longer focus light well enough to produce clear visual images and even to distinguish from anything else printed in a small type [1, 24]. The low vision thus makes it difficult for the elderly users to perform tasks of mouse pointing and clicking that require sharp near vision. Elderly computer users have the difficulty in mouse manipulation as facing the small target on the window [14, 17].

The mouse input manipulations are still more accurate than touch screen or gestural commands for the task requiring only single point interaction on windows display settings [7, 8, 10]. It is a challenge for the elderly people with age-related eye disorders to accurately distinguish and locate the objects that are smaller on the graphical interface environment of a computer. How to accommodate the characteristics of elderly people to ensure the accessibility of information and services is important to their quality of computer usage.

Changing the elderly people to adopt the novel input technologies is a challenging task [4, 5]. The critical point of technology acceptance for the elderly people is to make them perceive benefits of use and ease of use for the novel technologies [3, 21, 29]. Although graphical interfaces intrinsically contribute to the ease of use of computers. Interfaces based on windows, icons, menus, and pointing allow fairly non-trivial operations to be difficultly distinguished for the elderly people with low vision from manipulation of simple mouse clicks.

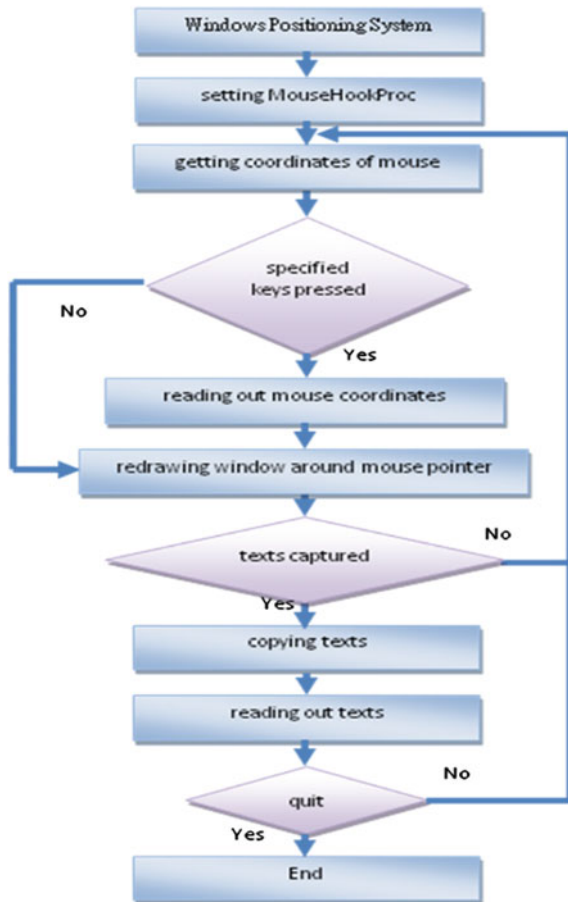
The performances of human behavior, perception and cognition are significantly faster and more accurate in multisensory processes than in only one sensory modality [6, 16, 19, 22]. Paivio [27] established the dual coding theory referring both visual and verbal information as to be processed differently and along distinct channels with the human mind. It is a theory of cognition. Both visual and verbal codes for representing information are used to organize incoming information into knowledge that can be acted upon, stored, and retrieved for subsequent use. The combination of vocalization cues with the visual signal adds another effective sensory source upon object identification. Consequently, the efficiency for the access to a target object with the integration of auditory and visual information occurs, as the receiver produces responses with alternative senses through the least ambiguous signal [11, 12, 20, 25, 28].

For assisting the elderly people with vision disordered in computer windows navigation, the information from auditory perception other than visual perception has become another important orientation messages for the very visually demanding mouse manipulations in computer windows environment [2, 9, 18, 26]. Visual rehabilitation with acoustic-oriented aids can play a major role in helping the elderly people with vision disordered improve their visual perception and increase their behavior effectiveness reflected on the mouse manipulations. As there are some cues telling where the mouse stays, where the mouse is going, and reading out to identify what the object is pointed by the cursor. The concept of alternative visual rehabilitation combines the use of assistive acoustic mechanism along with strategies for environmental modification to improve the quality of communication for the elderly people with vision disordered.

2 System Architecture

The design of Acoustic Assistance on Cursor Navigation (AACN) (Fig. 1) which provides the acoustic oriented services with mouse manipulations was developed by the Delphi 8.0 that integrated the IBM Chinese Text-to-Speech Engine and technology of capturing texts from the screen. With the technology of capturing texts from the screen, the AACN can copy the texts associated with an object pointed by the mouse cursor, such as a message box, a windows label, any tags going with a dialogue box, a program menu or help text, a graphic picture, a listing folders within a windows container, or some words displayed within an application, to the clipboard and subsequently read out by the IBM Chinese Text-to-Speech Engine as requested.

Fig. 1 System design flowchart



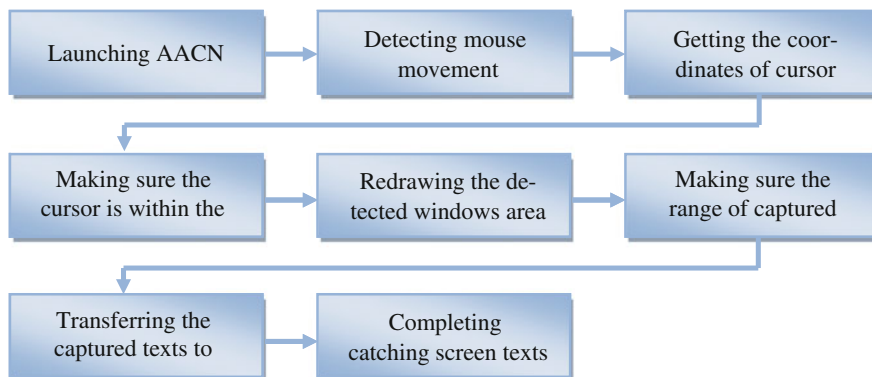


Fig. 2 Working flowchart of acoustic assistance on cursor navigation (AACN)

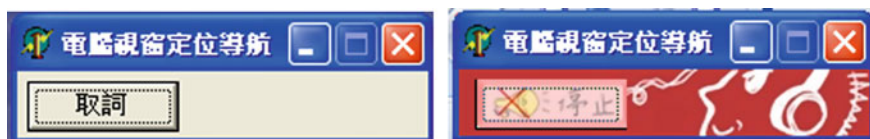


Fig. 3 Switching trigger function of talking aid

The working flowchart of AACN is shown as Fig. 2. The AACN mainly provides two core functions: Talking Aid (Fig. 3) and Cursor Positioning. For the Talking Aid function once triggered as requested, the AACN is able to read out the texts associated with the object where the cursor was just moved on. This would confirm user's mouse navigation along with the auditory information from the IBM Chinese Text-to-Speech Engine. The read-out texts could always associate with any windows object near the mouse cursor. Therefore, the Talking Aid function could identify users with the aural information of any target object during computer windows navigation with mouse. At the same time users could also highlight the texts of contents in the opened application or a windows area for reading out by the Talking Aid function.

For the Cursor Positioning function, the AACN can bring users to specifying cursor position by the readout of coordinates of present mouse cursor in x-coordinate and y-coordinate. It aims at facilitating the positioning of mouse cursor in need by clicking the right button of mouse. Thus while referring to the Cursor Positioning function of AACN, users can distinguish where the mouse cursor is in the computer windows environment.

3 Benefits Evaluation

The performances of high Acoustic-oriented Services with Talking Aid and Cursor Positioning, low Acoustic-oriented Services with Talking Aid, and low Acoustic-oriented Services with Cursor Positioning were evaluated for each subject during the tasks of portal website manipulations, WORD application, and operating system operations. Statistical analyses were completed on these performances. There were 52 elderly people aged above 55 years old who were classified into control group and experimental group.

3.1 *High Acoustic-Oriented Services/Talking Aid and Cursor Positioning*

A two-factor ANOVA with repeated measures comparing treatment groups across the first three tasks for portal website manipulations was used to analyze the conducted experiment data. The results showed that treatment had little effects on the increase of mouse manipulating time with the acoustic-oriented services via the functions of Talking Aid and Cursor Positioning ($F = 3.31$, $df = 1.510$, $p > .05$). However, there was a significant increasing accuracy of mouse manipulation across the three tasks for portal website surfing ($F = 11.98$, $df = 1.510$, $p < .05$). The accuracy of mouse manipulation was further analyzed and described a greater efficiency of acoustic-oriented services via the functions of Talking Aid and Cursor Positioning during the third task of portal website surfing than the previous two tasks ($p < .05$).

The differences in accuracy of mouse manipulation among the third task of portal website surfing, the task of WORD application, and the task of operating system operations for both groups were determined by the dependent t tests. The experimental group had a significant gain in accuracy of mouse manipulation via the functions of Talking Aid and Cursor Positioning from a mean of 9.8 during the third task of portal website surfing to a mean of 22.3 during the task of operating system operations [$t(25) = 3.59$, $p < .05$]. This gain in accuracy of mouse manipulation seems to be an extension of familiarity to the task of operating system operations. The accuracy of mouse manipulation differences between third task of portal website surfing, the task of WORD application was not significant [$t(25) = .54$, $p > .05$], while the mean accuracy of the task of WORD application, and the task of operating system operations was significantly different [$t(25) = 4.12$, $p < .05$]. The accuracy of mouse manipulation differences between the experimental and control groups on the task of WORD application showed that the experimental group with the functions of Talking Aid and Cursor Positioning had a greater accuracy of correct times than did the control group [$t(50) = 2.43$, $p < .05$].

3.2 *Low Acoustic-Oriented Services/Talking Aid*

A two-way ANOVA with repeated measures comparing the accuracy of mouse manipulation in both groups across the first three tasks of portal website surfing was used to obtain information in reducing difficulties about identification of widows objects. The results showed significant differences as the result of both the effect of acoustic-oriented services with Talking Aid ($F = 4.13$; $df = 1.47$; $p < .05$) and the effect of the task of WORD application ($F = 16.01$; $df = 1.510$; $p < .05$) as well as a significant interaction between acoustic-oriented services with Talking Aid and the task of WORD application ($F = 4.22$, $df = 1.150$; $p < .05$). It is obvious that an increased identification via acoustic-oriented services with Talking Aid for the tasks of portal website surfing is better for the experimental group.

The t tests were used to check the differences between the third task of portal website surfing and the task of WORD application with regard to the difficulties in identification of widows objects. The experimental group significantly increased the accuracy of correct times labeled acoustic-oriented services with Talking Aid between the third task of portal website surfing and task of WORD application [$t(25) = 2.38$, $p < .05$] and from task of WORD application to the task of operating system operations [$t(25) = 2.31$, $p < .05$].

3.3 *Low Acoustic-Oriented Services/Cursor Positioning*

A two-factor ANOVA with repeated measures was used to compare the differences made between treatment groups with regard to mouse positioning across the first three tasks of portal website surfing. Results showed that treatment had no significant effect on improving mouse positioning problems that subjects demonstrated while manipulating the mouse during the first portal website surfing task ($F = 2.29$, $df = 1.47$, $p > .05$). There was a difference as a result of the effect of portal website surfing task ($F = 12.05$; $df = 2, 100$; $p < .05$). All subjects improved the mouse positioning aspect of their portal website surfing from the first two tasks to the third, $p < .05$.

The t tests comparing the third task of portal website surfing and task of operating system operations were used as statistical analyses to observe the subjects' ability in mouse positioning. The experimental group made significant improvement from the third task of portal website surfing and task of operating system operations. The experimental group decreased by nearly 40 % amount of time engaged in looking for mouse behavior [$t(25) = 3.22$, $p < .05$]. A significant improvement between the task of WORD application and the task of operating system operations was also demonstrated [$t(25) = 7.1$, $p < .05$].

4 Conclusion

There are three anticipated effects from the evaluation of Acoustic Assistance on Cursor Navigation design in the mouse manipulation process for the elderly people as follows:

1. Smoothing away the obstructions of elderly people with low vision for manipulating computer system with traditional mouse input
The aural information assisted function of Acoustic Assistance on Cursor Navigation design would facilitate elderly people in the identification process of mouse cursor position, pointed objects by the mouse cursor, such as a message box, a windows label, any tags going with a dialogue box, a program menu or help text, a graphic picture, a listing folders within a windows container, or some words displayed within an application.
2. Providing an easier use of mouse manipulating environment
As the Acoustic Assistance on Cursor Navigation design launched, no more additional connected hardware is needed in comparison with traditional assisted hardware tool designs. To avoid the interferences of sight with hearing sense, the acoustic-oriented services of Talking Aid and Cursor Positioning functions are triggered only as requested by the right button of mouse. The elderly people would not be annoyed continually by the noises made the acoustic-oriented services as they have the ability to manipulate the computer system. The elderly people are always kept dominant with mouse input in a pure sight with hearing sense manipulating environment.
3. Encouraging the health care and the quality of life related industries developing digitalized web accessing materials
The access interfaces of Acoustic Assistance on Cursor Navigation design are simplified to reduce the learning curve and increase the mouse manipulating accuracy for the elderly people with low vision. It is a big opportunity for the health care and the quality of life related industries to extend and expand their services on the Internet as the aging era comes. And the elderly people with low vision will no longer be excluded from the digital world as their physiologies are degenerating.

The study showed significant efficiencies in integration of dual senses assisting the difficulties of the unimodal input, specifically the hearing input. The results of dual senses integration for mouse manipulation do not have any support of current input designs. The results do, however, support a new conceptualization of integration in which it is theorized that there are more contributing ingredients to integration than just the combination of the ability to move with sight, hear, and integrate. Further study is also necessary to assess the implications of difficult unimodal inputs on integration ability as we age, with future research including a wider age range of participants.

The Acoustic Assistance on Cursor Navigation design works with an aural assistance environment to allow the elderly people to select either the appropriate

functions of Talking Aid or Cursor Positioning to recognize what objects they are pointing at and to navigate where the mouse cursor they are moving within the computer windows. An advanced level of GPS-like loudspeaker in the computer windows environment, the Acoustic Assistance on Cursor Navigation design can help guide the elderly people with low vision by hearing sensory information to facilitate identifying and navigating where they are in the computer world.

References

1. Ah-Chan JJ, Downes S (2006) The aging eye. *Rev Clin Gerontol* 16(02):125–139
2. Baker CM, Milne LR, Scofield J, Bennett CL, Ladner RE (2014) Tactile graphics with a voice demonstration. In: Proceedings of the 16th international ACM SIGACCESS conference on computers and accessibility, ACM, pp 321–322
3. Boot WR, Charness N, Czaja SJ, Sharit J, Rogers WA, Fisk AD, Nair S (2015) Computer proficiency questionnaire: assessing low and high computer proficient seniors. *The Gerontologist* 55(3):404–411
4. Charness N, Boot WR (2009) Aging and information technology use potential and barriers. *Curr Dir Psychol Sci* 18(5):253–258
5. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, Sharit J (2006) Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychol Aging* 21(2):333
6. Durick J (2014) Reciprocal habituation: a study of older people and the Kinect. *ACM Trans Comput Hum Interact (TOCHI)* 21(3):18
7. Forlines C, Wigdor D, Shen C, Balakrishnan R (2007) Direct-touch vs. mouse input for tabletop displays. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 647–656
8. Fraser J, Gutwin C (2000) A framework of assistive pointers for low vision users. In: Proceedings of the fourth international ACM conference on assistive technologies, ACM, pp 9–16
9. Gao Q, Sun Q (2015) Examining the usability of touch screen gestures for older and younger adults. *Hum Factors J Hum Factors Ergon Soc* doi:[10.1177/0018720815581293](https://doi.org/10.1177/0018720815581293)
10. Grant K, Seitz P (1998) Measures of auditory-visual integration in nonsense syllables and sentences. *J Acoust Soc Am* 104:2438–2450
11. Granta KW (2002) Measures of auditory-visual integration for speech understanding: a theoretical perspective. *J Acoust Soc Am* 112(1):30–33
12. de Haas B, Schwarzkopf DS, Unger M, Rees G (2013) Auditory modulation of visual stimulus encoding in human retinotopic cortex. *Neuroimage* 70:258–267
13. Hanson VL, Crayne S (2005) Personalization of Web browsing: adaptations to meet the needs of older adults. *Univ Access Inf Soc* 4(1):46–58
14. Hanson VL (2001) Web access for elderly citizens. In: Proceedings of the 2001 EC/NSF workshop on universal accessibility of ubiquitous computing: providing for the elderly, ACM, pp 14–18
15. Hanson VL (2009) Age and web access: the next generation. In: Proceedings of the 2009 international cross-disciplinary conference on web accessibility (W4A), ACM, pp 7–15
16. Iain S, Edgar G, Florian P, Karl JH, Gerhard E, Andreas KE (2015) Auditory and visual interactions between the superior and inferior colliculi in the ferret. *Eur J Neurosci* 41(10):1311–1320
17. Jacko JA, Barreto AB, Marnet GJ, Chu JYM, Bausch HS, Scott IU, Rosa RH (2000) Low vision: the role of visual acuity in the efficiency of cursor movement. Proceedings of the fourth

- international acm conference on assistive technologies, ASSETS 2000. ACM, New York, NY, pp 1–8
18. Jochems N, Vetter S, Schlick C (2013) A comparative study of information input devices for aging computer users. *Behav Inf Technol* 32(9):902–919
 19. Kayser C, Logothetis NK, Panzeri S (2010) Visual enhancement of the information representation in auditory cortex. *Curr Biol* 20:19–24
 20. Kochkin S (2010) MarkeTrak VIII: consumer satisfaction with hearing aids is slowly increasing. *Hear J* 63:19–32
 21. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, Sharit J (2010) Older adults talk technology: technology usage and attitudes. *Comput Hum Behav* 26(6):1710–1721
 22. Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb Cortex* 14(4):452–465
 23. Mookherjee S, Bhattacharjee A, Sengupta M (2015) The aging eye. *J Ophthalmol* 2015 (832326):1–2
 24. Nagaraj RH, Linetsky M, Stitt AW (2012) The pathogenic role of Maillard reaction in the aging eye. *Amino Acids* 42(4):1205–1220
 25. Naue N, Rach S, Strüber D, Huster RJ, Zaehle T, Körner U et al (2011) Auditory event-related responses in visual cortex modulates subsequent visual responses in humans. *J Neurosci* 31:7729–7736
 26. Norman DA (2010) Natural user interfaces are not natural. *Interactions* 17(3):6–10
 27. Paivio A (1986) *Mental representations: a dual coding approach*. Oxford University Press, Oxford
 28. Tye-Murray N, Sommers M, Spehar B, Myerson J, Hale S (2010) Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear Hear* 31(5):636–644
 29. Vines J, Blythe M, Lindsay S, Dunphy P, Monk A, Olivier P (2012) Questionable concepts: critique as resource for designing with eighty somethings. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, pp 1169–1178

Effect of We-Intention on Adoption of Information System Embedding Social Networking Technology: A Case of Cloud Drive

Jerome Chih-Lung Chou

Abstract Collective intention, known as we-intention, should play a role in the adoption of cloud services that provide social networking or collaborative functions. In this research, the author explores the effects of we-intention on adoption of cloud drive, including both direct and moderation effects. The result shows that we-intention is a cause of adoption, but if the effect of usability on adoption is controlled, we-intention has little direct and moderation effects.

Keywords We-intention · Usability · Cloud drive

1 Introduction

Recently, due to the wide spread of concept of Web 2.0, many social technology-incorporated information systems are emerging in forms of cloud services, such as network album, social bookmark, audio and video station, hard drive, etc. Because the adoption of these systems relate to the decision of a group of people, the collective intention (we-intention) resulted from the perception of others' behavior or social influencing processes may play a role in the adoption. In the past, most studies on information system adoption took personal intention approach, neglecting the effect of we-intention. In this study, the author asserts we-intention should have an effect on adoption of cloud services, and takes the adoption of cloud hard drive as a representative example to explore the effect of we-intention. The author first examines the direct effect of we-intention, and then examines the moderation effect of it, speculating we-intention will strengthen the effect of usability on system adoption.

J.C.-L. Chou (✉)

Department of Management Information Systems, Hwa Hsia University of Technology,
New Taipei, Taiwan

e-mail: jerome@cc.hwh.edu.tw

2 Literature Review

2.1 Definition of Usability

The conceptual definition of usability by ISO (International Organization for Standardization) 9241-11 is the extent to which a product can be used to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [1, 2]. Usability means the use quality of a system for a user to achieve his/her own purpose after a series of tasks. Higher system usability helps users attain their challengeable objectives. The operational definition of usability by Nielsen dominates [3]. This definition includes visibility of system status, match between the system and the real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, and helping users recognize, diagnose, and recover from errors [4]. In short, five usability goals could be identified as a measurement framework: learnability, efficiency, memorability, error prevention, and satisfaction [5]. Usability has a positive effect on adoption, and this study uses Nielsen's framework to measure usability.

2.2 Definition of We-Intention

We-intention is often regarded as an individual's intention to perform a collective action with a group of people who are jointly committed to doing something as a body [6]. Different from the traditional individual intention, the concept of we-intention relies more on one's perception as a group member or an agent of a group, and taking the group as the target for intention formation. The concept of we-intention was initially explored by philosophic scholars who primarily focused on the conceptual and logical aspects [7]. Later, some studies in social psychology and marketing started to concentrate on measurement and hypothesis testing, and adopted this concept to explain online social behaviors [8–10]. Prior studies have suggested that social influence is especially important in predicting the successful adoption and use of interactive communication technology [11–13]. On this thread of research, Kelman's social influence processes [14] are often used as a theoretical base for developing knowledge in this area. Kelman has distinguished three aspects social influence processes, including compliance, internalization and identification. This study adopts Kelman's framework to measure we-intention.

3 Experiment Design

This study uses a convenient sample of university students who take the course of Electronic Commerce instructed by the author. In order to cultivate we-intention among students, the author assigned term projects to students and required them use Google Drive, a leading cloud service of hard drive and tools for team collaboration, to do their projects. Those teams who could manage their works with Google Drive could get better grades on team basis. At the end of semester, students were tested for their evaluation of usability of Google Drive, adoption of Google Drive, and we-intention to adopt Google Drive.

The author applied hierarchical regression to test the direct effect of we-intention on adoption. Furthermore, in order to examine the interaction effect of usability and we-intention, the scores of variables of usability, we-intention and the interaction term were centered and standardized before entering the model.

4 Results

The descriptive statistics are shown in Tables 1 and 2. The standard deviation of we-intention is higher than those of usability and adoption, indicating we-intention is more diverse among students. We-intention is highly correlated with usability and moderately correlated with adoption, showing it is a candidate factor of adoption.

The Cronbach's Alpha of usability, we-intention, and adoption are .881, .885, and .913 respectively, showing good reliability of constructs.

The hierarchical regression result is shown in Table 3. Model 1 uses usability and we-intention as independent variables, and its R square is .658, indicating good predictive power. Model 2 adds the mean centralized product of usability and we-intention to independent variables, but the increase of R square is small and insignificant, so the moderation effect is not obvious.

Read from Table 4, the coefficient of we-intention in Model 1, representing the direct effect on adoption, is small and insignificant while usability is already in the model.

However, if we only consider we-intention as independent variable, the regression result shows a significant effect of we-intention on adoption in Table 5. That means we-intention is a possible cause of adoption.

Table 1 Descriptive statistics

	Mean	Standard deviation
Usability	3.1897	.64609
We-intention	3.1839	.70965
Adoption	3.2672	.64609

Table 2 Correlation

	Usability	We-intention	Adoption
Usability	1	.752**	.804**
We-intention	.752**	1	.674**
Adoption	.804**	.674**	1

***p* < .01

Table 3 Model summary

Model	R ²	R ² change	F change	Significance
1	.658	.658	52.981	.000
2	.659	.000	.044	.835

Table 4 Coefficients

Model	B estimate	t	Significance
1 Constant	3.267	56.151	.000
Usability	.509	5.722	.000
We-intention	.120	1.343	.185
2 Constant	3.267	55.661	.000
Usability	.505	5.497	.000
We-intention	.116	1.261	.213
Interaction	-.014	-.209	.835

Dependent variable: adoption

Table 5 Coefficients

Model	B estimate	T	Significance
3 Constant	3.267	44.860	.000
We-intention	.502	44.860	.000

Dependent variable: adoption

5 Conclusion and Suggestion

Basing on the above analysis, the author found that when controlling usability, we-intention has a small direct effect on adoption, and no moderation effect. That is probably because usability is a strong determinant of adoption in this case. The sample’s evaluation of Google Drive usability is averagely high, and this is the main reason of adopting Google Drive. Future studies could be designed to add more brands of cloud hard drive for evaluation to gain wider range of usability score. If the sample includes more instances with low usability evaluation, the adoption mainly due to we-intention or the interaction between we-intention and usability might be more observable. The managerial implication is that we-intention is correlated with adoption and could be a significant factor of adoption of information system embedding social networking technology, in this case, Google Drive. For managers, in addition to improving system usability, finding ways of social network perspective to increase we-intention might be an alternative to increase system adoption.

References

1. Agarwal R, Venkatesh V (2002) Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability. *Inf Syst Res* 13(2):168–186
2. Peevers G, Douglas G, Jack M (2008) A usability comparison of three alternative message formats for an SMS banking service. *Int J Hum Comput Stud* 66(2):113–123
3. Gray W, Salzman M (1998) Damaged merchandise? A review of experiments that compare usability evaluation methods. *Hum Comput Interact* 13(3):203–261
4. Nielsen J (1994) Enhancing the explanatory power of usability heuristics. In: *Proceedings of the SIGCHI conference on human factors in computing systems celebrating interdependence—CHI '94*
5. Nielsen J (1993) *Usability engineering*. Academic Press, Boston
6. Bagozzi R, Lee K (2002) Multiple routes for social influence: the role of compliance internalization, and social identity. *Soc Psychol Q* 65(3):226
7. Tuomela R (2005) We-Intentions revisited. *Philos Stud* 125(3):327–369
8. Dholakia U, Bagozzi R, Pearo L (2004) A social influence model of consumer participation in network- and small-group-based virtual communities. *Int J Res Mark* 21(3):241–263
9. Bagozzi R, Dholakia U, Mookerjee A (2006) Individual and group bases of social influence in online environments. *Media Psychol* 8(2):95–126
10. Bagozzi R, Dholakia U, Pearo L (2007) Antecedents and consequences of online social interactions. *Media Psychol* 9(1):77–114
11. Koo C, Wati Y, Jung J (2011) Examination of how social aspects moderate the relationship between task characteristics and usage of social communication technologies (SCTs) in organizations. *Int J Inf Manage* 31(5):445–459
12. Cho H (2011) Theoretical intersections among social influences, beliefs, and intentions in the context of 3G mobile services in Singapore: decomposing perceived critical mass and subjective norms. *J Commun* 61(2):283–306
13. Dholakia U, Bagozzi R, Pearo L (2004) A social influence model of consumer participation in network- and small-group-based virtual communities. *Int J Res Mark* 21(3):241–263
14. Kelman H (1958) Compliance, identification, and internalization three processes of attitude change. *J Conflict Resolut* 2(1):51–60

Improving Project Risk Management of Cloud CRM Using DANP Approach

You-Shyang Chen, Chien-Ku Lin and Huan-Ming Chuang

Abstract Reducing information system project risks and improving organizational performance has become an important research issue. In this study, a research framework is constructed from the Stimulus-Organism-Response (S-O-R) framework, comprising the stimulus of project risk, the organism of project management, and the response of organizational performance. Cloud CRM experts projects management experience has many years in this study for the interview sample. DEMATEL-Based ANP (DANP) is MCDM analysis tool that haven't any pre-suppositions under the premise to explore dynamic relationship among project risk, project management, and organizational performance. The following empirical results were obtained: (a) effective project management reduced project risk and enhanced organizational performance; (b) of all of the types of project risk, organizational environment risk is the most challenging; (c) support from senior managers is crucial in project management; and (d) the multidimensional aspects of a organizational performance have garnered equal amounts of attention, indicating that financial performance is not the only important target.

Keywords DEMATEL based-ANP (DANP) · Cloud CRM · Project risk management · Organizational performance

Y.-S. Chen (✉)

Department of Information Management, Hwa Hsia University of Technology,
111, Gong Jhuan Rd, Chung Ho, New Taipei City 235, Taiwan, ROC
e-mail: ys_chen@cc.hwh.edu.tw

C.-K. Lin · H.-M. Chuang

Department of Information Management, National Yunlin University
of Science and Technology, 123, University Road, Section 3,
Douliou, Yunlin 64002, Taiwan, ROC
e-mail: g9923808@yuntech.edu.tw

H.-M. Chuang

e-mail: chuanghm@yuntech.edu.tw

1 Introduction

Information system projects are inherently associated with various types of risks, including those stemming from information technology (IT), human resources, usability, the project team, the project, and organization, as well as strategic and political risks [1]. Cloud CRM applies to any customer relationship management (CRM) technology where the CRM software, CRM tools and the organization's customer data together reside in the cloud and are delivered to end-users through Internet. In this study, we investigate the dynamic relationships between the project risk, project management, and organizational performance goals of cloud CRM projects. In sum, the research objectives of this study can be listed as follows. (1) Investigating the variables of the effects on cloud CRM project of project risk, project management, and organizational performance. (2) Exploring the relationship among project risk, project management, and organizational performance in cloud CRM project. (3) Establishing and improving cloud CRM projects in terms of risk management, project management and organizational performance mode, and thus make specific recommendations.

2 Literature References

This section introduces relevant literature on areas including project risks, project management, organizational performance measurement, risks of cloud computing service, cloud customer relationship management, and the DANP method.

2.1 *Project Risks*

Risk and uncertainty are different concepts [2], with Knight [3] being the first economist to propose the difference between risk and uncertainty. Risk means that although we do not know the outcome of events, we have the means to understand the results of the many different possible events and the likelihood of their occurrence. Project risk is an uncertain event that has a negative effect on project objectives [1]. Risks can be categorized as technical, external, human resource, cost, sponsor, and schedule-related.

2.2 *Project Management*

Project management is the flexible and effective application and coordination of various resources to meet project objectives and demands. Kerzner [4] asserts that

project management involves planning, organizing, and instruction in the use of controllable resources to attain concrete goals. In addition system path management is applied to assign specific project tasks to the functional employees in a department. Project management has been widely researched in various fields, such as the related literature [5–9].

2.3 Organizational Performance Measurement: A Balanced Scorecard

The balanced scorecard, or BSC, is a strategic management tool for establishing strategic indicators that facilitate the implementation of strategies to attain organizational goals. Kaplan and Norton [10] maintain that financial experts are responsible for overseeing performance measurement systems without the involvement of senior managers. Developing a BSC can help managers transcend traditional views, thereby converting strategies into measurable goals related to financial goals, customer satisfaction, internal business process, and learning and growth, while examining the performance of various domains.

2.4 Risks of Cloud Computing Service

Most of the enterprises providing data through cloud services exist at a higher level than local businesses. Therefore, cloud providers usually recommended doing security checks for enterprise cloud services in order to prevent malicious users from obtaining unauthorized access to data. Special attention should be paid to ensure that the clarity of cloud contracts definitions and protocol services meet customer needs [11]. Cloud CRM SaaS service model is one of the applications, with which a user needs only a computer, smart phone, or tablet PC's web browser in order to use the SaaS CRM.

2.5 Cloud Customer Relationship Management

Carr [12] asserted that cloud computing represents a transformation of the ways corporations perform computing, as evidenced by the shift in business computing from a private data centers into “the cloud.” According to the top 40 CRM software vendors rated by business-software.com, Microsoft Dynamics and Salesforce.com are the two leading cloud CRM software packages available for small and medium-sized businesses in Taiwan.

2.6 The DANP Method

Decision Making Trial and Evaluation Laboratory (DEMATEL)-based analytical network process (ANP) model to determine the relationships among and weights of criteria. This was done because the hybrid model can be used across various fields, such as outsourcing [13], Internet stores [14], and smart phone [15].

3 Research Methodology

This section is to introduce the procedures, including research framework, data collection, and data analysis and results.

3.1 Research Framework

Based on the literature review and the S-O-R model, which posits that environmental factors act as stimuli that affect individuals' cognitive reactions and then their behavior [16], Fig. 1 shows the research framework used to investigate the relationship among the 3 dimensions of project risk, project management, and organizational performance including 14 related constructs. Major dimensions and criteria of this study can be summarized as shown in Table 1.

3.2 Data Collection

We conducted a survey of 18 experts who were currently employed in cloud CRM experts with multiple years of practical experience in project management. This was

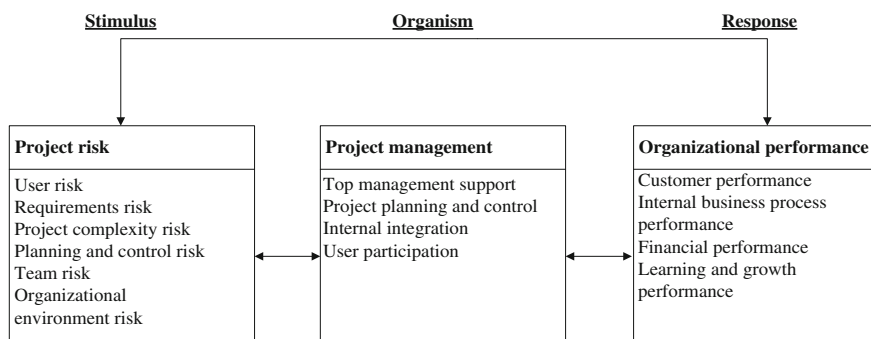


Fig. 1 Research framework of this study

Table 1 Dimensions and influential criteria of the research framework

Dimension	Influential criterion
A-Project risk	a1 User risk
	a2 Requirements risk
	a3 Project complexity risk
	a4 Planning and control risk
	a5 Team risk
	a6 Organizational environment risk
B-Project management	b1 Top management support
	b2 Project planning and control
	b3 Internal integration
	b4 User participation
C-Organizational performance	c1 Customer performance
	c2 Internal business process performance
	c3 Financial performance
	c4 Learning and growth performance

based on the recommendation of Northcutt and McCoy [17] that a focus group should comprise 12–20 experts who are (a) knowledgeable and highly experienced in the research topic; (b) capable of contemplating this topic and transcribing their thoughts into text; (c) motivated and available to participate in the research; (d) homogeneous regarding distance and power; and (e) able to exhibit excellent teamwork and not overly dominant or excessively shy in expressing their opinions.

3.3 Data Analysis and Results

This study adopts the steps 1–3 of the DANP model for building an IRM using the DEMATEL technique are summarized, as follows: (1) Generate the initial direct-relationship matrix; (2) normalize the initial relationship matrix to attain total-relationship matrixes; and (3) produce the impact relationship map. Steps 4–6, which are the latter part of the DANP model, to find influential weights using the ANP technique, as follows: (4) Normalize the total criteria relationship matrix; (5) normalize the total dimensions relationship matrix; and (6) build the weighted super-matrix and obtain influential weights of elements.

After steps 1–3, the attained results were adopted to produce 4 IRMs: dimensions of IRM, a dimension A-criterion of IRM, a dimension B-criterion of IRM, and a dimension C-criterion of IRM, all of which are displayed in Figs. 2, 3, 4 and 5, respectively. Major results of this study can be summarized as Tables 2 and 3 described as follows.

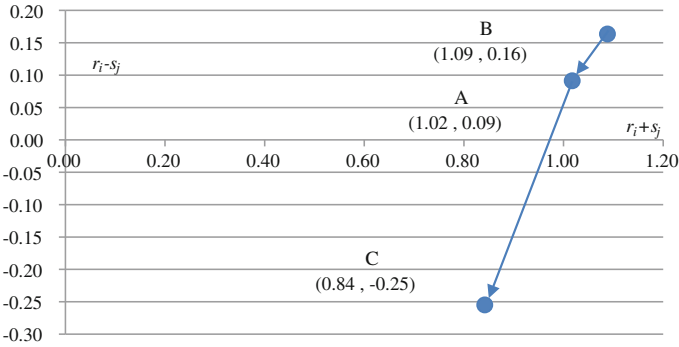


Fig. 2 Dimensions of IRM

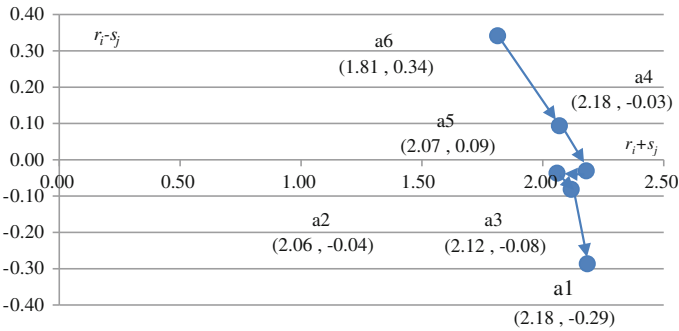


Fig. 3 Dimension A-criterion of IRM

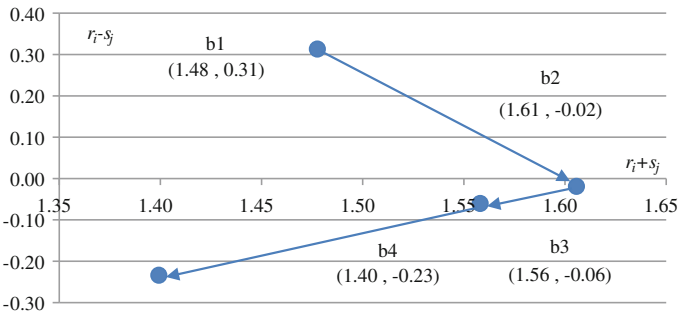


Fig. 4 Dimension B-criterion of IRM

In this study, using DEAMTEL can affect the known dimensions order of a cloud CRM project, as follows: “Project management (B),” “Project risk (A),” “Organizational performance (C).” For dimension A, the order of the criterion impact degrees are as follows: “Organizational environment risk (a6),” “Team risk

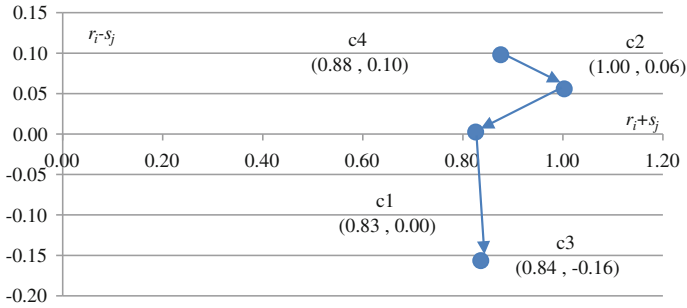


Fig. 5 Dimension C-criterion of IRM

Table 2 Key driving factors of dimensions and criteria

Primary factor	Secondary factors
B (Project management)	A (Project risk) > C (Organizational performance)
a6 (Organizational environment risk)	a5 (Team risk) > a4 (Planning and control risk) > a2 (Requirements risk) > a3 (Project complexity risk) > a1 (User risk)
b1 (Top management support)	b2 (Project planning and control) > b3 (Internal integration) > b4 (User participation)
c4 (Learning and growth performance)	c2 (Internal business process performance) > c1 (Customer performance) > c3 (Financial performance)

Table 3 Weights of each dimension, criterion and weight rank

Dimension	Criterion	Dimension		Criterion	
		Weight	Weight rank	Weight	Weight rank
A	a1	0.440	1	0.067	12
	a2			0.071	8
	a3			0.074	3
	a4			0.086	1
	a5			0.072	5
	a6			0.070	10
B	b1	0.297	2	0.071	9
	b2			0.082	2
	b3			0.072	6
	b4			0.072	7
C	c1	0.264	3	0.070	11
	c2			0.073	4
	c3			0.060	14
	c4			0.061	13

(a5),” “Planning and control risk (a4),” “Requirements risk (a2),” “Project complexity risk (a3),” “User risk (a1).” For dimension B, the order of the criterion impact degree is as follows: “Top management support (b1),” “Project planning and control (b2),” “Internal integration (b3),” “User participation (b4)”; in dimension C, the order of the criterion impact degree, following: “Learning and growth performance (c4),” “Internal business process performance (c2),” “Customer performance (c1),” “Financial performance (c3).” Furthermore, calculation of DANP weights indicates that “Project risk (A)” is the most important in cloud CRM project. “Internal business process performance (c2)” and “Financial performance (c3)” and “User risk (a1)” the influence of the highest, and “Customer performance (c1)” and “Project planning and control (b2)” were ranked fourth and fifth weights.

4 Conclusion

According to the results of cloud CRM experts study, cloud CRM project managers can consider making relevant improvements based on the following three criteria. (a) Internal business process performance, through effective project quality planning and problem solving can improve internal processes of the cloud CRM project planning, execution, and control, then support effective tracking milestone. The purpose of monitoring the project results is to determine project quality standards. Project results include the results of project and project management performance that can be delivered. (b) Financial performance during execution of the project and periodic review of financial performance can ensure the best use of limited resources. When the budget is added that should be monitored strictly. Financial audits should be regularly implemented at each stage of project in order to find problems in the early stages. (c) If the user does not have sufficient knowledge of the cloud CRM system, a project knowledge management system can be created, so the project management experience and the best practices can be transferred to the relevant members of project and staff training results can be significantly improved.

References

1. Gido J, Clements JP (2012) Successful project management. Cengage Learning, Boston
2. Zimmermann HJ (2000) An application-oriented view of modeling uncertainty. *Eur J Oper Res* 122(2):190–198
3. Knight FH (1921) Risk, uncertainty, and profit. Hart, Schaffner and Marx, New York
4. Kerzner H (1984) Project management: a systems approach to planning, scheduling, and controlling. Van Nostrand-Reinhold, New York
5. Davis K (2014) Different stakeholder groups and their perceptions of project success. *Int J Project Manage* 32(2):189–201
6. Hornstein HA (2015) The integration of project management and organizational change management is now a necessity. *Int J Project Manage* 33(2):291–298

7. Kaiser MG, El Arbi F, Ahlemann F (2015) Successful project portfolio management beyond project selection techniques: understanding the role of structural alignment. *Int J Project Manage* 33(1):126–139
8. Serra CEM, Kunc M (2015) Benefits realisation management and its influence on project success and on the execution of business strategies. *Int J Project Manage* 33(1):53–66
9. Todorović ML, Petrović DČ, Mihić MM, Obradović VL, Bushuyev SD (2015) Project success analysis framework: a knowledge-based approach in project management. *Int J Project Manage* 33(4):772–783
10. Kaplan RS, Norton DP (1992) The balanced scorecard-measures that drive performance. *Harvard Bus Rev* 70(1):71–79
11. Fan CK, Chen TC (2012) The risk management strategy of applying cloud computing. *Int J Adv Comput Sci Appl* 3(9):18–27
12. Carr N (2008) *Cloud computing: the new spice trails, the big switch rewiring the world from Edison to Google*. WW. Norton, New York
13. Hsu CC, Liou JJ, Chuang YC (2013) Integrating DANP and modified grey relation theory for the selection of an outsourcing provider. *Expert Syst Appl* 40(6):2297–2304
14. Chiu WY, Tzeng GH, Li HL (2013) A new hybrid MCDM model combining DANP with VIKOR to improve E-store business. *Knowl Based Syst* 37:48–61
15. Hu SK, Lu MT, Tzeng GH (2014) Exploring smart phone improvements based on a hybrid MCDM Model. *Expert Syst Appl* 41(9):4401–4413
16. Mehrabian A, Russell JA (1974) *An approach to environmental psychology*. MIT Press, Cambridge
17. Northcutt N, McCoy D (2004) *Interactive qualitative analysis: a systems method for qualitative research*. Sage, Thousand Oaks, California

Improving Project Risk Management by a Hybrid MCDM Model Combining DEMATEL with DANP and VIKOR Methods—An Example of Cloud CRM

Chien-Ku Lin, You-Shyang Chen and Huan-Ming Chuang

Abstract In recent years, cloud technology has been widely used, including enterprise CRM systems. To reduce the cloud CRM project risk, that improves organizational performance, so it has become an important issue. In this study, a research framework is constructed by the Stimulus-Organism-Response (S-O-R) framework, that through DANP (DEMATEL based ANP) and VIKOR research methods explore the relationship in project risk, project management and organizational performance. The findings of this research can provide a valuable reference for minimizing project management risk and enhancing organizational performance through effective project management in cloud CRM project.

Keywords DEMATEL based-ANP (DANP) · VIKOR · Cloud CRM · Project risk management · Organizational performance

1 Introduction

Project risk management positively influences project selection, the determination of project scope, the development of realistic schedules, and cost estimations [1]. As the most comprehensive and the most advanced cloud computing model is considered the SaaS model. SaaS model of cloud computing provides complete functionality for

C.-K. Lin (✉) · H.-M. Chuang

Department of Information Management, National Yunlin University of Science and Technology, 123, University Road, Section 3, Douliou, Yunlin 64002, Taiwan, ROC
e-mail: g9923808@yuntech.edu.tw

H.-M. Chuang

e-mail: chuanghm@yuntech.edu.tw

Y.-S. Chen

Department of Information Management, Hwa Hsia University of Technology, 111, Gong Jhuan Rd. Chung Ho, New Taipei 235, Taiwan, ROC
e-mail: ys_chen@cc.hwh.edu.tw

companies, and applications include e-mail or other communication tools such as web, video or office suite. In this study, we investigated the dynamic relationship among the project risk, project management, and organizational performance goal of cloud CRM project. In summation, the research objectives of this study can be listed as follows: (1) Explore the relationship among project risk, project management, and organizational performance in cloud CRM project; (2) empirical cases findings and views of cloud CRM experts verify each other that will provide suggestions for improvement of cloud CRM project risk management.

2 Literature References

This section introduces relevant literature on areas including project risks, project management, organizational performance measurement, risks of cloud computing service, cloud customer relationship management, the DANP method, the Delphi method, and the VIKOR method.

2.1 Project Risks

Uncertainty means that we cannot know or anticipate the outcome of possible events and that we are therefore unable to obtain the information necessary to make qualitative or quantitative judgments [2]. Earlier risk research [3] defined risk as a condition that can exert a serious threat to the successful delivery of an IT project. Risk management has been widely thought to improve the approaches to IT project performance [4, 5].

2.2 Project Management

Recent studies regarding project management in the related literature [6, 7] conclude that the key factors determining the success of a project include cost, performance, time to completion, and scope; satisfying the requirements of these factors indicates that a project has been successfully implemented.

2.3 Organizational Performance Measurement: A Balanced Scorecard

Thus, with the involvement of senior managers, employees across the entire company can increase their understanding of the company's visions and goals.

Similar to the dials and indicators in an airplane cockpit, a BSC can help managers fully comprehend complex information [8].

2.4 Risks of Cloud Computing Service

Reducing costs is benefits of cloud computing, CRM through the Internet and Web browser service are saving in the cost of building, maintenance, IT staff training. However, SaaS problems need to be considered, which mainly includes the company's integration with other information systems and stored in the service provider's sensitive data [9]. In a traditional part of the scope of information security policies (e.g. Cyber Security Liability Insurance, Cyber breach Insurance, Privacy-Data Breach Insurance, and Network Security and Privacy Insurance), provide cloud computing services applications.

2.5 Cloud Customer Relationship Management

SaaS model is currently the best approach to automate for customer relationships management through CRM systems. Cloud CRM system saves the cost of maintenance and IT personnel. It does not require each computer to install CRM software.

2.6 The DANP Method

The DEMATEL-based ANP model (named the DANP model in this study) was adopted to verify the research framework and confirm the dependence relationship between dimensions and criteria. Particularly, the DEMATEL and ANP techniques do not need any presumptions; thus, to use them in exploring the dependence relationship is suitable in this study. The DANP model, whose steps are organized in two parts: (1) building an IRM using the DEMATEL technique, and (2) finding influential weights using ANP technique.

2.7 The Delphi Method

The Delphi method is a research technique that is used to address complex problems by using a structured communication process of a panel of experts to forecast [10], make decisions, and solve complex problems. With objective application of the Delphi method, we explore creative ideas and produce valuable information.

Knowledge collected during the Delphi study is synthesized and distilled from the use of a series of questionnaires. Responses to questionnaires were collected on site and were reviewed directly [11].

2.8 The VIKOR Method

The positive and negative ideal points are on basic concept of VIKOR technique, which was first put forth by Opricovic [12] and Opricovic and Tzeng [13]. The VIKOR (the Serbian name, VlseKriterijumska Optimizacija I Kompromisno Resenje) method is based on the compromise programming of MCDM. The various alternatives are denoted as $A_1, A_2, \dots, A_i, \dots, A_m$. For an alternative A_i , the merit of the j th aspect is denoted by f_{ij} , that is, f_{ij} is the value of j th criterion function for the alternative A_i .

3 Research Methodology

This section is to introduce the procedures, including research framework, data collection, and data analysis and results.

3.1 Research Framework

Based on the literature review and the S-O-R model [14] that Fig. 1 shows the research framework to investigate the relationship among the 3 dimensions of project risk, project management, and organizational performance and including 14 related criteria.

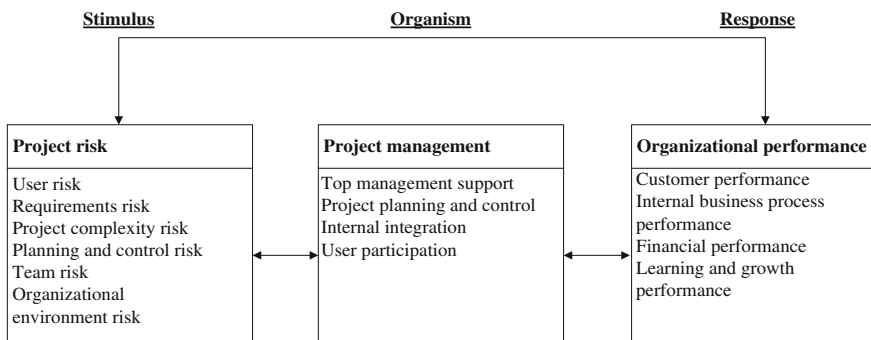


Fig. 1 Research framework of this study

3.2 Data Collection

In the cloud CRM project that investigate 18 cloud CRM experts, and through DANP, Delphi and VIKOR to analyze questionnaire data. It was based on the recommendation of Northcutt and McCoy [15] that the focus group should comprise 12–20 experts.

3.3 Data Analysis and Results

This study adopts the steps 1–3 of the DANP model for building an IRM using the DEMATEL technique are summarized, as follows: (1) Step 1: the initial direct-relationship matrix A was normalized to obtain matrix X . All elements x_{ij} in X must satisfy $0 \leq x_{ij} \leq 1$ and the principal diagonal element must equal 0; (2) step 2: subsequently, the total criteria relation matrix T_c and total dimensions relation matrix T_d can be obtained through matrix X ; and (3) step 3: the degree to which each criterion and dimension influences and is influenced by all others was determined. Steps 4–6, which are the latter part of the DANP model, to find influential weights using the ANP technique, as follows: (1) Steps 4 and 5: the total criteria relationship matrix and total dimension relationship matrix were normalized to yield T_{c^*} and T_{d^*} , respectively, and were subsequently multiplied with each other to produce the original weighted super-matrix; and (2) step 6: matrix S was transposed to S^* such that S^* satisfies the column-stochastic principle.

Through analysis of the results of Delphi Round 2, which dimensions and criteria are in compliance with the $IQR < 0.6$ and between 0.6 and 1.0. More than 85 % IQR question is less than 1.0 that indicates expert opinion has converged and the Delphi method is completed questionnaires in this study. Then taking the data of Mean of round 2 as the VIKOR's performance data.

In this study, using DEAMTEL can affect the cloud CRM project to known dimensions order, following: "Project management (B)," "Project risk (A)," "Organizational performance (C)"; in dimension A, the order of the criterion impact degree, following: "Organizational environment risk (a6)," "Team risk (a5)," "Planning and control risk (a4)," "Requirements risk (a2)," "Project complexity risk (a3)," "User risk (a1)"; in dimension B, the order of the criterion impact degree, following: "Top management support (b1)," "Project planning and control (b2)," "Internal integration (b3)," "User participation (b4)"; in dimension C, the order of the criterion impact degree, following: "Learning and growth performance (c4)," "Internal business process performance (c2)," "Customer performance (c1)," "Financial performance (c3)." Further, through DANP calculate weights, "Project risk (A)" is the most important in cloud CRM project. "Internal business process performance (c2)" and "Financial performance (c3)" and "User risk (a1)" the influence of the highest, and "Customer performance (c1)" and "Project planning and control (b2)" were ranked fourth and fifth weights.

Cloud CRM project’ priority gaps for improvement can be listed as follows. The priority index for improving the dimensions includes project risk, organizational performance, and project management. The priority index for improving the criteria includes user risk, requirements risk, organizational environment risk, team risk, planning and control risk, financial performance, learning and growth performance, internal business process performance, user participation, project planning and control, internal integration, project complexity risk, customer performance, and top management support.

The study result, which ranked three dimensions’ performances, determined that project management surpassed organizational performance, which surpassed project risk [project management (B) > organizational performance (C) > project risk (A)]. Major results of this study can be summarized as Tables 1 and 2 described as follows.

Table 1 Performance values combined with the influential weights of the criteria according to the DANP for the cloud CRM experts

	Global weight (DANP)	Local weight (DANP)	Cloud CRM experts		
			Aspiration value	Performance (Delphi)	Gap (VIKOR)
A		0.403		6.215	0.348
a1	0.082	0.204	9.000	5.780	0.402
a2	0.069	0.170	9.000	5.892	0.388
a3	0.071	0.177	9.000	7.160	0.230
a4	0.069	0.172	9.000	6.228	0.347
a5	0.063	0.157	9.000	6.184	0.352
a6	0.049	0.120	9.000	6.041	0.370
B		0.275		7.151	0.231
b1	0.058	0.210	9.000	7.528	0.184
b2	0.075	0.273	9.000	7.044	0.244
b3	0.070	0.254	9.000	7.083	0.240
b4	0.072	0.263	9.000	7.028	0.247
C		0.322		6.672	0.291
c1	0.075	0.234	9.000	7.333	0.208
c2	0.089	0.277	9.000	6.667	0.292
c3	0.084	0.260	9.000	6.306	0.337
c4	0.074	0.229	9.000	6.403	0.325
Total	1.000	1.000		6.618	0.298

Table 2 The ranking indexes of performance for the Cloud CRM project

	S	Q	R
A. Project risk	2.089	0.402	1.246
B. Project management	0.915	0.247	0.581
C. Organizational performance	1.161	0.337	0.749

4 Conclusion

In this study, using VIKOR know cloud CRM project performance for experts, the order following: “Project management (B),” “Organizational performance (C),” “Project risk (A)”; to improve criterion first, following: “User risk (a1)” and “User participation (b4)” and “Financial performance (c3).”

Cloud CRM expert consensus in the relevant project, the following are suggestions for improvement: (a) Project risk: from project risk planning needs to improve, including planning, analysis, processing and monitoring; how risk tracking, filing how to perform, and the project reserve and the use of time, etc., that it needs a clear description of the records, and strengthen education about risk management knowledge then really do good risk management; (b) user risk: if user does have not confidence for the system that will affect the progress of the project’s final. User’s responsibilities need clear; you can reduce the conflict, but also enhance the project’s team strength; (c) user participation: the teams need to continue to inform the user about the project’s progress. End of the project related information can be used to pass information to all the relationship people by delivery system, allows users assesses project team to complete the task; and (d) financial performance: recording implementation process of the financial cost and change approval process can control the development costs within the budget range.

References

1. Schwalbe K (2010) Information technology project management. Cengage Learning, Boston (Revised)
2. Zimmermann HJ (2000) An Application-oriented view of modeling uncertainty. *Eur J Oper Res* 122(2):190–198
3. Schmidt R, Lyytinen K, Keil M, Cule P (2001) Identifying software project risks: an international Delphi study. *J Manage Inf Syst* 17(4):5–36
4. De Bakker K, Boonstra A, Wortmann H (2010) Does risk management contribute to IT project success? A meta-analysis of empirical evidence. *Int J Project Manage* 28(5):493–503
5. Spears JL, Barki H (2010) User participation in information systems security risk management. *MIS Q* 34(3):503–522
6. Lawry K, Pons DJ (2013) Integrative approach to the plant commissioning process. *J Indust Eng* 2013:1–12
7. Vanhoucke M (2013) An overview of recent research results and future research avenues using simulation studies in project management. *ISRN Comput Math* 2013:1–19
8. Kaplan RS, Norton DP (1992) The balanced scorecard-measures that drive performance. *Harvard Bus Rev* 70(1):71–79
9. Němeček J, Vaňková L (2011) CRM and cloud computing. In: Proceedings of the 2nd international conference on applied informatics and computing theory. World Scientific and Engineering Academy and Society (WSEAS), pp 255–259
10. Mattingly-Scott M (2006) Delphi method. http://www.12manage.com/methods_helmer_delphi_method.html

11. Adler M, Ziglio E (1996) *Gazing into the oracle: the Delphi method and its application to social policy and public health*. Jessica Kingsley Publishers, London
12. Opricovic S (1998) Multicriteria optimization of civil engineering systems. *Fac Civ Eng Belgrade* 2(1):5–21
13. Opricovic S, Tzeng GH (2004) Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. *Eur J Oper Res* 156(2):445–455
14. Mehrabian A, Russell JA (1974) *An approach to environmental psychology*. MIT Press, Cambridge
15. Northcutt N, McCoy D (2004) *Interactive qualitative analysis: a systems method for qualitative research*. Sage, Thousand Oaks, California

Using VIKOR to Improve E-Service Quality Performance in E-Store

Chien-Ku Lin, You-Shyang Chen, Huan-Ming Chuang
and Chyuan-Yuh Lin

Abstract With the rise of the Internet, e-commerce vigorously grows and more and more websites gradually emerge. If the websites want to keep competitive advantage and sustainable development in the highly competitive environment, they inevitably need to provide consumers with high-quality service to create excellent experience for consumers and win the customers' heart to establish mutually beneficial and long-term relationship. In this study, we chose three well-known domestic e-stores, use VIKOR method, in the optimal solution way to compare performance of dimension and criteria, and then propose improving suggestions and strategies to reduce the gap.

Keywords Service quality · VIKOR

1 Introduction

The rapid rise of e-services, to subvert the traditional business model, but also changed the shopping habits of consumers. In addition to physical retail stores to buy goods, consumers can purchase products through online shopping.

C.-K. Lin (✉) · H.-M. Chuang · C.-Y. Lin

Department of Information Management, National Yunlin University of Science and Technology, 123, University Road, Section 3, Douliou 64002, Yunlin, Taiwan, ROC
e-mail: g9923808@yuntech.edu.tw

H.-M. Chuang
e-mail: chuanghm@yuntech.edu.tw

C.-Y. Lin
e-mail: g9923807@yuntech.edu.tw

Y.-S. Chen
Department of Information Management, Hwa Hsia University of Technology, 111, Gong Jhuan Rd., Chung Ho District, New Taipei City 235, Taiwan, ROC.
e-mail: ys_chen@cc.hwh.edu.tw

The homogeneity of the product on the Internet is easy to cause competitive price, the seller if you want in a competitive environment to stand out; you need to establish long-term partnerships with customers, while also giving consumers innovative services. Service is an important key that the firm and its competitors to produce differentiated. Different companies can offer the same products, but the service was not the same content. Therefore, enterprises can be highlighted differences by providing services. For consumers, enterprise provided outcome of services through service quality to assess. The service quality will affect the merits of consumer willingness to buy [1]. If companies provide good service quality, customer will have positive behavioral intentions [2]. Scholars also pointed out that service quality is the main factor that influences consumer trust and satisfaction to the website [1]. In addition, loyalty is one of the most common factors to investigate commercial activity which is the key success factor in the competitive market. Fornell and Wernerfelt [3] mentioned that the cost of business create for a new source of customer, companies spend far more than maintaining an existing customer [3]. Therefore, we adopt SERVQUAL's and QES's service quality dimension as we survey affecting service quality dimensions [4]. In this study, we chose three well-known domestic e-stores, use VIKOR (Vlse Kriterijumska Optimizacija I Kompromisno Resenje), in the optimal solution way to compare performance of dimension and criteria, and then propose improving suggestions and strategies to reduce the gap. The purpose of this study is summarized in the following points: (1) Use cases study of VIKOR to investigate dimensions and criteria performance assessment; and (2) make recommendations on the research results, as a reference for when e-store development strategy.

2 Literature Review

This section introduces relevant literature on areas including service quality, e-service quality, service quality models, customer loyalty and its antecedents, and the VIKOR method.

2.1 Service Quality

The construct of service quality as conceptualized in the service marketing literature centers on perceived quality, defined as a consumer's judgment about an entity's overall excellence or superiority [5]. The most popular and consistent definition of "perceived service quality" likens the concept to an attitude. Parasuraman et al. [6, 7] defined perceived service quality as "a global judgment, or attitude, relating to the superiority of the service." Service quality perceptions result from a comparison of customer expectations with actual service performance. McIntyre [8] defined perceived service quality as "a belief (or attitude) about the degree of excellence of

a service...” that means, the definition of service quality also come from the viewpoints of customers’ requirements, namely, the consistency between result of service and customer expected service specification [8].

2.2 *E-Service Quality*

Electronic services have recently received considerable attention in academic research. Rust [9] defined the concept as “the provision of service over electronic networks.” Indeed, contributions to research on electronic services mainly originate from the fields of services marketing [10] and electronic commerce [11], as well as information systems research [12] and they tend to be multidisciplinary.

2.3 *Service Quality Models*

PZB SERVQUAL Model. As to measure the service quality perceived by the customers, they proposed ten service quality dimensions: (1) reliability (2) responsiveness (3) competence (4) access (5) courtesy (6) credibility (7) communication (8) security (9) tangible (10) understanding/knowing the customer [6]. Later, in order to avoid the overlaps between the dimensions, the aforementioned ones were simplified into five dimensions: (1) reliability (2) responsiveness (3) assurance (4) empathy (5) tangible, the SERVQUAL model was also developed [7]. The SERVQUAL model has caught the attention of scholars of the service marketing field. It is widely applied in different disciplines, and its versatility [13] and the measurement method [14] have also heated a fierce debate within the academia.

Quality of Electronic Services Model (QES). Fassnacht and Koese [15] use the three main dimensions-environment quality, process quality and outcome quality to development a hierarchical model for the measurement of online service quality. Scholars believe that the environment quality refers to the appearance of the user interface; the process quality, which is the delivery quality, refers to the interaction of the consumers with the website during the service process (such as searching information, selecting the product or going through the transaction); outcome quality refers to the measurement of the service result after accepting the service [15].

2.4 *Customer Loyalty and Its Antecedents*

Customer Loyalty. Oliver [16] defines customer loyalty as “a deeply held commitment to re-buy or re-patronize a preferred product or service consistently in the future, despite situational influences and marketing efforts having the potential to

cause switching behavior” [16]. Action loyalty relates to the conversion of intentions to action, supported by a willingness to overcome obstacles to achieve the action. This study aims to discuss the “repurchase intention” and the “advocacy intention” formed among the consumers after having traded with the e-shopping.

Repurchase Intention. This study believes that consumers’ loyalty intention shall include both the attitudinal and the behavioral levels. Such concept is applicable on e-business [17]: when the e-loyalty is higher, the possibility of re-purchasing the product or service provided by the enterprise is also higher in the future, so the profit of the e-retailer would also increase.

Advocacy Intention. This study advocates that advocacy intention is a loyalty representation of the customers. Business administrators can measure the willingness of providing positive recommendations as the assessment of the customers for the enterprise [18]. Based on the obtained information and compilation, the customers can maintain their superior purchase experience as well as the advocacy intention creating a win-win situation.

Satisfaction. Satisfaction is necessary but not sufficient component of loyalty [19] Satisfaction is an antecedent of brand loyalty, with increases in satisfaction leading to increases in brand loyalty [20].

Affective Commitment. The consumers are connected with the business partners through psychological acceptance and attachment [21]. Moreover, Gundlach et al. [22] believe that affective commitment is interlinked by several affections, including loyalty, involvement and attachment [22]. Overall, affective commitment is originated from the pleasurable purchase experience or outstanding service, so a positive feeling is generated and the consumer would will to keep the linkage with its partner.

2.5 The VIKOR Method

The basic concept of VIKOR technique lies in defining the positive and negative ideal points, which was first put forth by Opricovic [23] and Opricovic and Tzeng [24]. The VIKOR (The Serbian name, VlseKriterijumska Optimizacija I Kompromisno Resenje) method is based on the compromise programming of MCDM. The various alternatives are denoted as a_1, a_2, \dots, a_m . For an alternative a_i , the merit of the j th aspect is denoted by f_{ij} , that is, f_{ij} is the value of j th criterion function for the alternative a_i .

3 Research Methodology

This section is to introduce the procedures, including research framework, data collection, and data analysis and results.

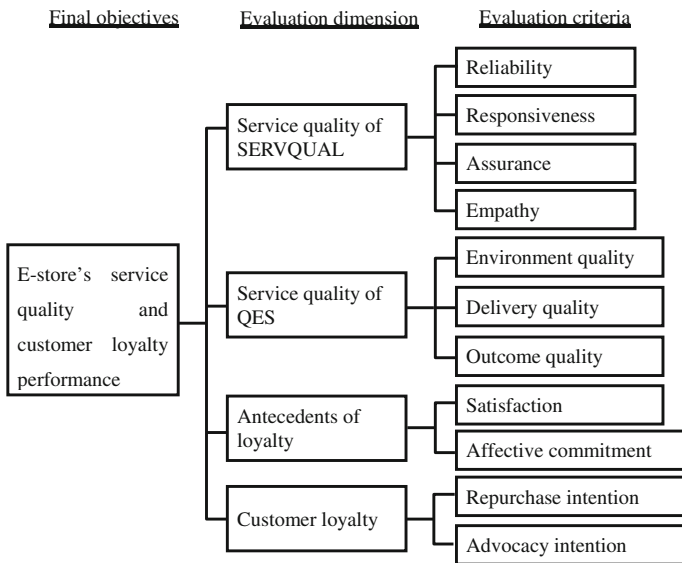


Fig. 1 Research framework of this study

3.1 Research Framework

This study investigates the relationship between E-store's service quality and customer loyalty evaluation as shown in Fig. 1.

3.2 Data Collection

The central issue of this study is the elements within the service quality that influence the consumers' loyalty behaviors during an online shopping process. Thus, the objects of this study are Internet users who actually purchase in shopping websites in the Taiwanese region. Through surveys, we would target only on consumers who have online shopping experiences and a certain quantity of samples is being needed to reach the requirement of our study. This study expects to collect a wide range of information, so the gender, age and educational background are not restricted in this study. Two questionnaires were delivered: first, convenience sampling would be executed for pilot study. Based on the result, the items were adjusted minimally and a formal questionnaire was given. Online questionnaire was given for convenience sampling as to collect the necessary amount of samples. The formal questionnaire was designed with Google forms. The link of the questionnaire was posted in BBS and social networks so that it can be completed voluntarily. Last, invalid questionnaires were deleted when recovering the samples as to

increase the reliability and validity of the information. Through Facebook and PTT Bulletin Board System, 389 online questionnaires were recycled with 11 of them as invalid samples, which accounts for 2.8 % of the total amount of questionnaires; that is to say, there are 378 valid samples accounting for 97.2 % of the total questionnaires. As to have an overall understanding of the samples, a descriptive statistical analysis was done for the 378 valid samples. Items of analysis include gender, age, educational level, profession, average monthly income, online shopping experience, average shopping frequency, daily average time of using shopping website and the most visited shopping website.

The three e-store are Yahoo! Shopping Mall (272, 59.8 %), Bookline (65, 14.3 %) and Pchome Online Shopping (41, 9.0 %). Having surveyed several Taiwan EC websites, Chang and Chen [25] found that, according to users, Yahoo.com, Yahoo auction, PChome.com, and Books.com were the four top online shopping sites [25]. Another study [26] found that the most popular website was Yahoo (45.1 %), followed by PChome (7.9 %), Books.com (6.8 %), and Ezfly (5.3 %), which accounted for 65 % of all responses. These studies are consistent with the results of our survey. Therefore, we choose the three top e-stores to demonstrate the proposed method. We will survey the three e-store website performance by service quality, satisfaction, commitment and loyalty factors.

3.3 *Data Analysis and Results*

In the VIKOR technique processing, Steps 1-3 were executed accordingly. In the experimental results, Table 1 presents the results of criterion performance. Table 2 presents the ranking indexes of three e-stores performances.

4 Conclusion

In Table 2, this study's result which ranked three e-stores' performances, determined that Books surpassed Pchome, which surpassed Yahoo (Books > Pchome > Yahoo). In Table 1, the integration of the performance index scores of Yahoo further demonstrated that the dimension of the "customer loyalty" gap is 0.384 and the gap for the "advocacy intention" criterion is 0.429 constitution the largest gaps, which the Yahoo e-store should improve as a priority. Thus, the priority for Yahoo is to enhance their advocacy intention to promote their customer's loyalty. The integration of the performance index scores of Books further demonstrated that the dimension of the "antecedents of loyalty" gap is 0.319 and the gap for the "affective commitment" criterion is 0.389 constitution the largest gaps, which the Books e-store should improve as a priority. Thus, the priority for Books is to enhance their affective commitment to promote their antecedents of loyalty. The integration of the performance index scores of Pchome further

Table 1 Performance values combined with the influential weights of the criteria according to the factor loading from SEM factor analysis results

Alternatives		Yahoo		Books		Pchome	
Dimensions	Criteria	Performance	Gap	Performance	Gap	Performance	Gap
A. Service quality of SERVQUAL		4.855	0.358	5.437	0.260	5.152	0.308
	REL	4.952	0.341	5.580	0.237	5.406	0.266
	RES	4.894	0.351	5.583	0.236	5.421	0.263
	ASS	5.080	0.320	5.630	0.228	5.433	0.261
	EMP	4.492	0.418	4.956	0.341	4.349	0.442
B. Service quality of QES		5.370	0.272	5.609	0.232	5.320	0.280
	ENQ	5.207	0.299	5.542	0.243	4.941	0.343
	DEQ	5.611	0.231	5.775	0.204	5.657	0.224
	OUQ	5.291	0.285	5.511	0.248	5.362	0.273
C. Antecedents of loyalty		4.765	0.372	5.083	0.319	4.851	0.358
	SAT	4.918	0.347	5.497	0.250	5.390	0.268
	ACM	4.613	0.398	4.669	0.389	4.312	0.448
D. Customer loyalty		4.698	0.384	5.132	0.311	4.790	0.368
	REP	4.973	0.338	5.249	0.292	4.996	0.334
	ADV	4.424	0.429	5.015	0.331	4.584	0.403

Table 2 The ranking indexes of performances for the empirical cases

	S	Q	R
Yahoo	3.758	0.429	2.093
Books	2.999	0.389	1.694
Pchome	3.525	0.448	1.986

demonstrated that the dimension of the “customer loyalty” gap is 0.368 and the gap for the “affective commitment” criterion is 0.448 constitution the largest gaps, which the Pchome e-store should improve as a priority. Thus, the priority for Pchome is to enhance their affective commitment to promote their antecedents of loyalty, and improve customer loyalty. The e-store strategies were defined by using the data in Table 1. This process demonstrated that the priorities of each e-store’s strategy were dissimilar. The results indicated that Yahoo, Pchome needed to enhance their customer loyalty, Books needed to enhance its antecedents of loyalty. Therefore, Yahoo, Pchome and Books should efforts to build customer relationship and promote customer loyalty. In addition, providing the best service quality to customers at service quality dimensions of QES is an essential dimension because this consideration primarily determines the customers’ choice of e-store.

The e-store strategy, which emphasizes the e-store business goal of satisfying customers' needs. The results of this research indicate that no e-store business strategy is the same; consequently, managers of e-stores must use this method to determine their customers' wants and needs to define the gap and improve it to achieve the ideal solution or aspiration level.

References

1. Zhou T, Lu Y, Wang B (2009) The relative importance of website design quality and service quality in determining consumers' online repurchase behavior. *Inf Syst Manag* 26(4):327–337
2. Zeithaml VA, Berry LL, Parasuraman A (1996) The behavioral consequences of service quality. *J Mark* 60(2):31–46
3. Fornell C, Wernerfelt B (1987) Defensive marketing strategy by customer complaint management: a theoretical analysis. *J Mark Res* 337–346
4. Brady MK, Cronin JJ Jr (2001) Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. *J Mark* 65(3):34–49
5. Zeithaml VA, Berry L (1987) The time consciousness of supermarket shoppers. Texas A&M University Working Paper
6. Parasuraman A, Zeithaml VA, Berry LL (1985) A conceptual model of service quality and its implications for future research. *J Mark* 49(4):41–50
7. Parasuraman A, Zeithaml VA, Berry LL (1988) SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *J Retail* 64(2):12–40
8. McIntyre G (1993) Sustainable tourism development: guide for local planners. World Tourism Organization (WTO)
9. Rust R (2001) The rise of e-service. *J Serv Res* 3(4):283–284
10. Janda S, Trocchia PJ, Gwinner KP (2002) Consumer perceptions of internet retail service quality. *Int J Serv Ind Manag* 13(5):412–431
11. Sultan F, Urban GL, Shankar V, Bart IY (2002) Determinants and role of trust in e-business: a large scale empirical study. *Ebus Res Center* 1–44
12. Aladwani AM, Palvia PC (2002) Developing and validating an instrument for measuring user-perceived web quality. *Inf Manag* 39(6):467–476
13. Carman JM (1990) Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions. *J Retail* 66(1):33–55
14. Peter JP, Churchill GA, Brown TJ (1993) Caution in the use of difference scores in consumer research. *J Consum Res* 19(4):655–662
15. Fassnacht M, Koese I (2006) Quality of electronic services conceptualizing and testing a hierarchical model. *J Serv Res* 9(1):19–37
16. Oliver RL (1999) Whence consumer loyalty? *J Mark* 63(4):33–44
17. Reichheld F, Scheffer P (2000) E-Loyalty your secret weapon on the web. *Harv Bus Rev* 78(4):105–113
18. Reichheld F (2006) The microeconomics of customer relationships. *MIT Sloan Manag Rev* 47(2):73–78
19. Agustin C, Singh J (2005) Curvilinear effects of consumer loyalty determinants in relational exchanges. *J Mark Res* 42(1):96–108
20. Bennett R, Rundel-Thiele S (2005) The brand loyalty life cycle: implications for marketers. *J Brand Manag* 12(4):250–263
21. Fullerton G (2011) Creating advocates: the roles of satisfaction, trust and commitment. *J Retail Consum Serv* 18(1):92–100
22. Gundlach GT, Achrol RS, Mentzer JT (1995) The structure of commitment in exchange. *J Mark* 59(1):78–92

23. Opricovic S (1998) Multicriteria optimization of civil engineering systems. Faculty of Civil Engineering, Belgrade. 2(1):5–21
24. Opricovic S, Tzeng GH (2004) Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. *Eur J Oper Res* 156(2):445–455
25. Chang HH, Chen SW (2009) Consumer perception of interface quality, security, and loyalty in electronic commerce. *Inf Manag* 46(7):411–417
26. To PL, Liao C, Lin TH (2007) Shopping motivations on internet: a study based on utilitarian and hedonic value. *Technovation* 27(12):774–787

Study on the Intellectual Capital and Firm Performance

Chiung-Lin Chiu, You-Shyang Chen and Mei-Fang Yang

Abstract The purpose of this study is to investigate the relationship between intellectual capital and firm performance for pharmaceutical industry in Taiwan. As we know, the intellectual capital has becoming the main resource for creating companies' value and the competitiveness of an enterprise. R&D productivity is often considered as the key success factor for creating firms' value of pharmaceutical industry. The 252 firm-year observations of this research are collected from companies of pharmaceutical industry listed on Taiwan Stock Exchange and Gre Tai Securities Market dated from 2007 through 2013. Financial data of this research is collected from Taiwan Economic Journal (TEJ) database. I find evidence revealing that there is a positive correlation between intellectual capital and firm performance.

Keywords Intellectual capital · Firm performance · Pharmaceutical industry

1 Introduction

In knowledge economy ear, intellectual capital is the key resource for enterprises to obtain profits and create competitive advantages. Intellectual capital in new economy era gradually replaces land, capital and materials in old economy ear and becomes important sources to create value. In recent years, firm in intelligence

C.-L. Chiu (✉)

Department of Business Administration, Hwa Hsia University of Technology,
New Taipei City, Taiwan
e-mail: jolene@cc.hwh.edu.tw

Y.-S. Chen · M.-F. Yang

Department of Information Management, Hwa Hsia University of Technology,
New Taipei City, Taiwan
e-mail: ys_chen@cc.hwh.edu.tw

M.-F. Yang

e-mail: hugoing67@gmail.com

intensive and knowledge intensive industry create higher market value by intellectual capital. For pharmaceutical industry, there can be the differences of at least 10 times. Hence, in new economy era, enterprises' capacity to control and use intangible assets such as employees' skills, R&D innovation and business process becomes critical of corporate competitiveness.

As some labor intensive industries have relocated to other counties, the industrial structure in Taiwan has changed significantly. Intellectual capital has become the success key factors of enterprises' transformation as high value added industry in Taiwan. This study uses pharmaceutical industry as an example. The pharmaceutical industry has paid special attention to intangible assets. According to the "Taiwan Pharmaceuticals and Healthcare Report Q2 2015" from Market Research.com, pharmaceuticals expenditure is TWD167.74bn (USD5.53bn) and TWD175.07bn (USD5.56bn) in 2014 and 2015 respectively.

Many literatures suggested that intellectual capital is the factor to drive and create corporate value and it positively influences corporate performance [1, 4, 5, 8–11, 14, 15, 17]. They found that intellectual capital leads to competitive advantages for enterprises, and increases corporate value. It is the most valuable asset and useful tool of competition. In this research, we examine the relationship between Intellectual Capital and firm performance for pharmaceutical industry in Taiwan. Our proxy for intellectual capital is Tobin's Q ratio. Our results indicate that there is a positive association between intellectual capital and firm performance.

2 Literature Review

2.1 *Intellectual Capital*

According to the Wikipedia, "*Intellectual Capital is the combined value of its people (Human Capital), the value inherent in its relationships (Relational capital), and everything that is left when the people go home (Structural capital), of which Intellectual property is but one sub-component.*" Figure 1 represents the concept of intellectual capital [13]. Intellectual capital can be separated into strategic and measurement stream. Table 1 represents the intellectual capital framework of Skandia's value scheme. As shown in Table 1, intellectual capital is the sum of human capital and structural capital. Structural capital also included customer capital and organization capital. Edvinsson and Malone [4] recommended 112 metrics to measure five areas (financial focus; customer focus; process focus; renewal and development focus, and human focus) of the navigator model. These metrics are summarized as shown in Table 2.

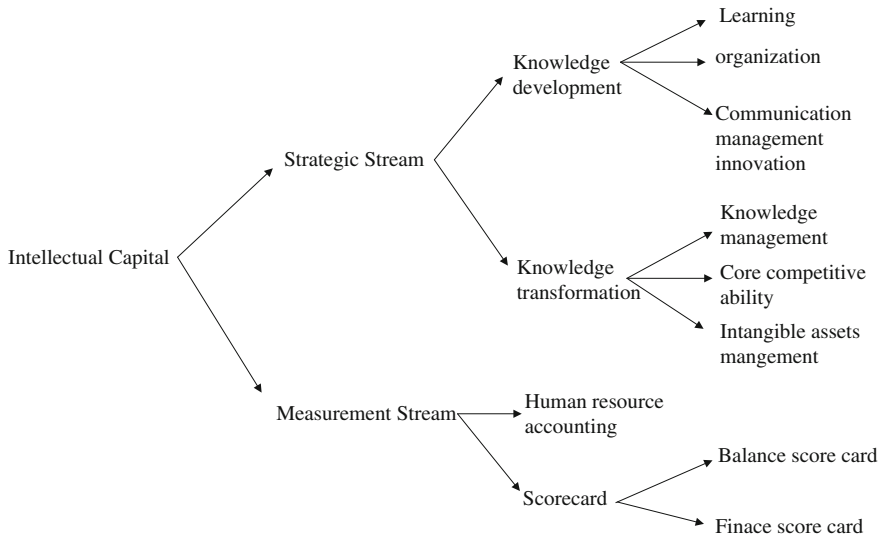


Fig. 1 The concept of intellectual capital. *Resource* Roos et al. [13]

Table 1 Sample of Skandia IC measures

Financial focus	• Revenues/employee (\$)
	• Revenues from new customers/total revenue (\$)
	• Profits resulting from new business operations (\$)
Customer focus	• Days spent visiting customers (#)
	• Ratio of sales contacts to sales closed (%)
	• Number of customers gained versus lost (%)
Process focus	• PCs/employee (#)
	• IT capacity—CPU (#)
	• Processing time (#)
Renewal and development focus	• Satisfied employee index (#)
	• Training expense/administrative expense (%)
	• Average age of patents (#)
Human focus	• Managers with advanced degrees (%)
	• Annual age of patents (#)
	• Leadership index (%)

Resource Botis [2]

2.2 Intellectual Capital and Firm Performance

According to research findings of Hirschey and Weygandt [6], Hall [7], Chauvin and Hirschey [3], R&D expenditure and corporate performance have significant and positive correlation with market value. Sougiannis [16] treated 573 large-scale

Table 2 Descriptive statistics

Variable	Mean	Std. Dev.	Min.	Max.
$RDI_{i,t}$	0.784	2.616	0.000	25.603
$IC_{i,t}$	0.929	1.000	0.186	11.283
$Adverstising_{i,t}$	0.182	0.193	0.000	2.188
$Lev_{i,t}$	0.337	0.327	0.008	4.853
$Size_{i,t}$	14.160	1.060	10.990	16.364
$Gw_{i,t}$	2.582	24.615	-1.000	406.750

enterprises in the U.S. from 1975 to 1985 as samples and probed into effect of R&D expenditure on long-term profits and market value of enterprises. Based on research results, increase of R&D expenditure with 1 NTD will enhance 2 NTD of profits and 5 NTD of market value in the following 7 years. That study not only supported the positive relationship between R&D expenditure and corporate performance, but also found that the effect of R&D expenditure on market value is based on profits. In other words, investors consider the profits led by R&D expenditure for enterprises; hence, they believe that operational performance of enterprises has high market value and their investment is worthy. Lev and Sougiannis [12] tried to find if R&D expenditure of enterprises has delay effect on future profits and if capital of R&D expenditure increases stock prices and profits. According to the findings, an increase of R&D expenditure with 1 NTD leads to profits of 2.328 NTD. Capitalization of R&D expenditure enhances explained power of stock price and profits (Fig. 2).

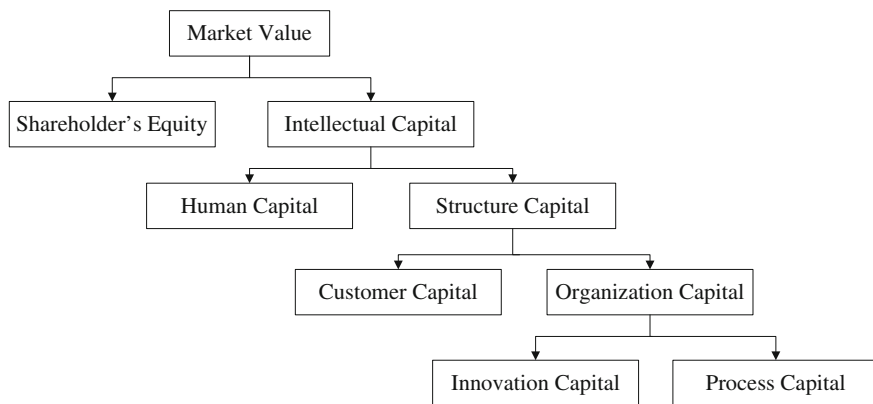


Fig. 2 Intellectual capital framework of Skandia. Edvinsson and Malone [4]

3 Methodology

3.1 Hypothesis

In modern times of knowledge economy, the main factor to drive and create corporate value is intellectual capital. Hence, it is inferred that results by creating and accumulating intellectual capital will be reflected on corporate performance. The hypothesis is as follows:

H: There exists a positive correlation between intellectual capital and firm performance.

3.2 Sample and Data

Data of this research were collected from the Taiwan Economic Journal (TEJ) database for pharmaceutical companies listed on Taiwan Stock Exchange (TSE) and Gre Tai Securities Market (GTSM) dated from 2007 to 2013. There are 252 firm-year observations used in this research.

3.3 Empirical Model

Our empirical model tests the association between intellectual capital and firm performance. The regression model is shown in Eq. (1).

$$RDI_{i,t} = \beta_0 + \beta_1 IC_{i,t} + \beta_2 Advertising_{i,t} + \beta_3 Lev_{i,t} + \beta_4 Size_{i,t} + \beta_5 Gw_{i,t} + \varepsilon_{i,t} \quad (1)$$

The operational definition of the variables are described as follows:

IC = (market value of common stock + Book value of preferred stock)/Book value of total assets

The proxy for intellectual capital is Tobin's Q ratio. Tobin's Q is a dummy variable, set equal to 1 if the Tobin's Q ratio is more than 1, set to 0 otherwise.

$RDI_{i,t}$ stands for the expenditure of research and development divided by sales.

$$(RDI_{i,t}) = \frac{R\&D \text{ Expenditure}}{Sales}$$

Adverstising $_{i,t}$ stands for the expenditure of advertisement divided by sales income.

$$(\text{Adverstising}_{i,t}) = \frac{AD \text{ Expenditure}}{Sales \text{ Income}}$$

$Lev_{i,t}$ stands for the sum of short-term and long-term borrowing divided by total assets.

$$(Lev_{i,t}) = \frac{Sum\ of\ short - term\ borrowing\ and\ long - term\ borrowing}{Total\ Assets}$$

$Size_{i,t}$ is the log of total assets.

$$(Size_{i,t}) = Total\ Assets(Ln)$$

$Growth_{i,t}$ is the sales of year t minus sales of year t – 1 divided by sales of year t – 1.

$$(Growth_{i,t}) = \frac{Sales_t - Sales_{t-1}}{Sales_{t-1}}$$

4 Empirical Results

4.1 Descriptive Statistics

Table 2 presents the means, standard deviation, minimum and maximum of the regression variables.

4.2 Pearson Correlation

Table 3 presents the Pearson correlation between the variables used in Regression (1).

Table 3 Pearson correlation

Variable	$RDI_{i,t}$	$IC_{i,t}$	Adverstising $_{i,t}$	$Gw_{i,t}$	$Size_{i,t}$	$Lev_{i,t}$
$RDI_{i,t}$	1					
$IC_{i,t}$	0.501**	1				
Adverstising $_{i,t}$	0.230**	-0.80	1			
$Gw_{i,t}$	0.313**	0.009	-0.038	1		
$Size_{i,t}$	-0.157**	-0.377**	0.058	0.050	1	
$Lev_{i,t}$	0.430**	0.207**	-0.066	0.028	0.053	1

Table 4 Regression result

Variable	β	P value
$IC_{i,t}$	0.231	0.001***
$Adverstising_{i,t}$	0.356	0.000***
$Lev_{i,t}$	-0.133	0.017**
$Size_{i,t}$	-0.070	0.253
$Gw_{i,t}$	0.354	0.000***

$$RDI_{i,t} = \beta_0 + \beta_1 IC_{i,t} + \beta_2 Adverstising_{i,t} + \beta_3 Lev_{i,t} + \beta_4 Size_{i,t} + \beta_5 Gw_{i,t} + \varepsilon_{i,t}$$

4.3 Regression Results

Our empirical model investigates the association between intellectual capital and firm performance, after controlling for other determinants of firm performance. Table 4 reports the results of regression of R&D intensity, Tobin's Q (proxy of intellectual capital), and other controls. We find that Tobin's Q is positive and statistically significant associated with the R&D intensity.

5 Conclusion

Strategy and organizational culture are important in the adoption of knowledge management. Intellectual capital plays an important role in it. From the perspective of enterprises, with vigorous development of information technology and communication network, environment of business organizations becomes complicated and international and cross-departmental integration increases. In order to deal with highly competitive environment, short life cycle of product and diversity of goods, effectively enhance competitiveness and satisfy customers' needs, enterprises' proper management of intellectual capital is an important issue. In addition, the key asset of enterprises is knowledge. Knowledge asset which is unique and can hardly be imitated by colleagues is their competitive advantages. Therefore, intellectual capital is the key to measure unique knowledge of enterprises.

This study has determined that the pharmaceutical industry has paid special attention to benchmarking strategy. In this study, we examine the firm performance and intellectual capital. Our proxy for intellectual capital is Tobin's Q ratio. Our results indicate that there is a positive association between intellectual capital and firm performance.

Our results are subject to several caveats. One caveat pertains to our use of the R&D intensity to measure firm performance. There are other proxies such as ROA, ROE. Second, There are several models used to measure intellectual capital, such as MVA, EVA, citation-weighted patents, balanced score card and human resource accounting. Future researchers could extend our study by using other proxies as stated above.

References

1. Amir E, Lev B (1996) Value-relevance of nonfinancial information: the wireless communications industry. *J Account Econ* 22:3–30
2. Bontis N (2001) Assessing knowledge assets: a review of the models used to measure intellectual capital. *Int J Manag Rev* 3(1):41–60
3. Chauvin KW, Hirschey M (1993) Advertising, R&D expenditures and the market value of the firm. *Financ Manag* 22(Winter):128–140
4. Edvinsson L, Malone MS (1997) Intellectual capital—realizing your company's true value by finding its hidden roots. Harper Business, New York
5. FASB (2001) Improving business reporting: insight into enhancing voluntary disclosures. Steering Committee Business, Reporting Research Project Financial Accounting Standard Board, Norwalk, CT
6. Hirschey M, Weygandt J (1985) Amortization policy for advertising and research and development expenditures. *J Account Res* 23(Spring):326–335
7. Hall R (1993) A framework linking intangible resource and capabilities to sustained competitive advantage. *Strateg Manag J* 14(November):607–618
8. Johanson U, Martensson M, Skoog M (2001a) Mobilizing change through the management control of intangibles. *Account Organiz Soc* 26(October–November):715–733
9. Johanson U, Martensson M, Skoog M (2001) Measuring to understand intangible performance drivers. *Eur Account Rev* 10(September):407–437
10. Kaplan RS, Norton DP (1992) The balanced scorecard *Harvard Business Review*, Jan–Feb. 71–79
11. Ittner CD, Larcker DF, Rajan MV (1997) The choice of performance measures in annual bonus contracts. *Account Rev* 72(April):231–255
12. Lev B, Sougiannis T (1996) The capitalization, amortization and value relevance of R&D. *J Account Econ* 21(February):107–138
13. Roos J, Roos G, Dragoetti N, Edvinsson L (1997) Intellectual capital: navigating in the New Business Landscape, Macmillan Business, London
14. Stewart TA (1997) Intellectual capital: the new wealth of organizations, Bantam Doubleday Dell Publishing Group, Inc
15. Sullivan PJ (2000) Value-driven intellectual capital: how to convert intangible corporate assets into market value. Wiley, New York, NY
16. Sougiannis T (1994) The accounting based valuation of corporate R&D. *Account Rev* 69 (January):44–68
17. Wallman S (1995) The future of accounting and disclosure in an evolving world: the need for dramatic change. *Account Horiz* 9(3):81–91

Voluntary Disclosure and Future Earnings

Chiung-Lin Chiu and You-Shyang Chen

Abstract This study investigates whether voluntary disclosure provides future earnings information to investor in the capital market of Taiwan. Using 165 firm-year observations collected from the Taiwan Economic Journal (TEJ) financial database for companies listed on the Taiwan Stock Exchange and Gre Tai Securities Market in 2005 with December as the fiscal year-end. The result shows that voluntary disclosure improves the association between current stock returns and future earnings.

Keywords Voluntary disclosure · Future earnings

1 Introduction

As we know, there is an inevitable existence of the information risk in the capital market. The information asymmetry has long existed in the capital market [2, 6, 10]. It is also existed between the management and the investors. Sloan [13] found that the agency cost can be reduced by information disclosure.

Voluntary disclosure is mainly driven by the company itself in an attempt to disclose related financial forecast and other performance information through annual report or website. The Securities and Futures Institute (hereafter SFI), entrusted by the Taiwan Stock Exchange Corporation and the Gre Tai Securities

C.-L. Chiu (✉)

Department of Business Administration, Hwa Hsia University of Technology,
New Taipei City, Taiwan
e-mail: jolene@cc.hwh.edu.tw

Y.-S. Chen

Department of Information Management, Hwa Hsia University of Technology,
New Taipei City, Taiwan
e-mail: ys_chen@cc.hwh.edu.tw

Market, launched “Information Disclosure and Transparency Rankings System” (hereafter IDTRS) to evaluate the level of transparency for all listed companies in Taiwan since 2003. SFI has released annual ratings on its web site, making such data publicly available for investors.

Since 2005, the information transparency ratings announced by IDTRS and in addition, with companies willing to share more transparency brought forward by voluntary disclosure. Do investors pay more importance to such voluntary disclosure? This research employs the methodology of Lundholm and Myers [11] and intends to examine whether the anticipation of future earnings will be reflected in the current stock prices of voluntary disclosure in Taiwanese stocks market.

The rest of the paper is organized as follows. Section 2 reviews the related literature and develops the hypothesis. Section 3 describes my research sample and methodology. Section 4 reports the empirical results. Section 5 concludes the research.

2 Literature Review and Hypothesis Development

2.1 Future Earnings

In this study, Future Earnings Response Coefficient (FERC) is used to measure the future earnings. Future Earnings Response Coefficient was originally developed by Collins et al. [4]. Lundholm and Myers [11] adapted Collins’s model to study the effect of voluntary disclosure on the return-earnings relation. This study shows disclosure (as measured by AIMR corporate disclosure ratings) can “bring the future forward”, that means more future earnings news can be reflected in current stock prices.

Gelb and Zarowin [7] found that voluntary corporate disclosure results in more future earnings information which is impounded in current returns. Chu and Wu [3] investigated the impact of product market competition and public opinion pressure on the informativeness of stock prices using an international dataset. They find that FERC is higher in countries with strong public opinion pressure for loss firms. Yu [14] uses the cross-country data to examine the relationship between corporate governance and stock price informativeness. The result shows that FERC increases with the quality of a firm’s corporate governance. Haw et al. [9] use data from 32 economies and find that greater financial disclosure, higher earnings quality and greater information dissemination are associated with stock prices that are more informative about future earnings.

2.2 Voluntary Disclosure

Information Disclosure and Transparency Ranking (IDTRS) identified 114 disclosure items and searched them from annual report, regulatory filing via Internet as evaluation criteria grouped into the following five categories: (1) Compliance with the mandatory disclosures; (2) Timeliness of reporting; (3) Disclosure of financial forecast; (4) Disclosure of annual report; (5) Corporate website disclosure. (http://weblinesfi.org.tw/top/Ranking_System/2010RankingResults.pdf).

Companies listed in Taiwan Stock Exchange Corporation and Gre Tai Securities Market were ranked as “Grade A⁺”, “Grade A”, “Grade B”, “Grade C” and “Grade C⁻” according to the evaluation criteria by IDTRS. Ranking results published by IDTRS include a company list of voluntary disclosure with higher information transparency. Voluntary disclosure is mainly driven by the company itself in an attempt to disclose related financial forecast and other performance information through annual report or website.

2.3 Voluntary Disclosure and Future Earnings

The company management that practices voluntary forecast is mainly driven to prevent major surprises from happening when releasing earnings [1], or to intimidate competitors from entering the industry [5, 8, 12]. Companies of voluntary disclosure with higher transparency will have higher FERC. I formulate the hypothesis as follows:

H: Firms with higher voluntary disclosure transparency level have a greater FERC.

3 Research Methodology

3.1 Sample and Data

Financial data of this research were obtained from the Taiwan Economic Journal (TEJ) finance database for companies listed on Taiwan Stock Exchange (TSE) and Gre Tai Securities Market (GTSM) in 2005 with December as the fiscal year-end. The information transparency and disclosure ranking data were obtained from the website of The Securities and Futures Institute (<http://www.sfb.gov.tw/en/>).

3.2 Empirical Model

Model in Eq. (1) is used as our standard model in our analyses.

$$R_t = b_0 + b_1X_{t-1} + b_2X_t + b_3X3_t + b_4R3_t + \varepsilon_t \quad (1)$$

In Eq. (1), R_t is the annual stock return for year t ; X_{t-1} and X_t are annual earnings per share for fiscal year $t - 1$ and t , deflated by the stock price at the beginning of year t , respectively; $X3_t$ is the sum of earnings per share for fiscal years $t + 1$ through $t + 3$, all deflated by the stock price at the beginning of year t ; $R3_t$ is the annual returns for the three-year period following year t .

To test my hypothesis, I expand Eq. (1) by adding the voluntary disclosure measure VD and its interaction with the other variables of Eq. (2). VD is a dummy variable, set equal to 1 if firms were ranked as the better voluntary disclosure by the Information Disclosure and Transparency Rankings System and set to 0 otherwise. Equation (2) is my empirical model to test hypothesis:

$$R_t = b_0 + b_1X_{t-1} + b_2X_t + b_3X3_t + b_4R3_t + b_5VD_t + b_6VD_t * X_{t-1} + b_7VD_t * X_t + b_8VD_t * X3_t + b_9VD_t * R3_t + \varepsilon_t \quad (2)$$

4 Empirical Results

4.1 Descriptive Statistics

Table 1 provides descriptive statistics on returns, earnings and voluntary disclosure variables. Current stock returns R_t have a mean of 0.008, while the three year sum is 1.731. The mean of current earnings X_t is 0.023, while the three year sum is 0.086. For the voluntary disclosure variable, 50 % of the observations are classified as firms with high transparency level of voluntary disclosure.

Table 1 Descriptive statistics

Variables	Minimum	Maximum	Mean	Std. Dev.
R_t	-0.629	3.781	0.008	0.494
X_{t-1}	-3.245	0.788	0.016	0.231
X_t	-1.456	0.482	0.023	0.152
$X3_t$	-6.868	1.616	0.086	0.453
$R3_t$	-11.481	81.365	1.731	5.664
VD_t	0	1	0.5	0.5

Table 2 Regressions of voluntary disclosure

$R_t =$	$b_0 +$	$b_1 X_{t-1} +$	$b_2 X_t +$	$b_3 X3_t +$	$b_4 R3_t +$	$b_5 VD_t +$	Adj- R^2
	0.081	-0.240	0.312	0.084	-0.162	0.317	0.503
	(0.002)	(0.006)	(0.000)	(0.411)	(0.015)	(0.000)	
	$b_6 VD_t * X_{t-1} +$	$b_7 VD_t * X_t +$	$b_8 VD_t * X3_t +$	$b_9 VD_t * R3_t + \varepsilon_t$			
	-0.505	0.453	0.282	-0.371			
	(0.000)	(0.000)	(0.000)	(0.000)			

Table 2 reports the regression results from Eq. (2). The coefficient on $VD_t * X3_t$ is significantly positive, which supports the hypothesis of a positive association between voluntary disclosure and FERC, as predicted in Hypothesis.

5 Conclusion

With financial fraud scandals arise in Taiwanese Stocks Market, investors are paying more attention to the investment information. Since 2003, The Securities and Futures Institute (hereafter SFI), entrusted by the Taiwan Stock Exchange Corporation and the Gre Tai Securities Market, launched “Information Disclosure and Transparency Rankings System” (hereafter IDTRS) to evaluate the level of transparency for all listed companies in Taiwan. Ranking results published by IDTRS include a company list of voluntary disclosure with higher information transparency since 2005.

This study investigates the association between voluntary disclosure with higher information transparency and future earnings. The result shows that voluntary disclosure improves the association between current stock returns and future earnings. The result further finds evidence that the future earnings will be reflected in the current stock prices of voluntary disclosure in Taiwanese stocks market.

References

1. Ajinkya BB, Gift MJ (1984) Corporate managers’ earnings forecasts and symmetrical adjustments of market expectations. *J Account Res* 22:425–444
2. Brown S, Hillegeist SA (2007) How disclosure quality affects the level of information asymmetry. *Rev Account Stud* 12:443–477
3. Chu EL, Wu W (2011) The informativeness of current return about future earnings and extra-legal institutions: international evidence. *NTU Manag Rev* 22(1):67–96
4. Collins DW, Kothari SP, Shanken J, Sloan RG (1994) Lack of timeliness and noise as explanations for the low contemporaneous return-earnings association. *J Account Econ* 18 (3):289–324
5. Darrrough MN, Stoughton NM (1990) Financial disclosure policy in an entry game. *J Account Econ* 12:219–244

6. Easley D, O'Hara M (1992) Time and the process of security price adjustment. *J Financ* 47 (2):577–604
7. Gelb D, Zarowin P (2002) Corporate disclosure policy and the informativeness of stock prices. *Rev Account Stud* 7(1):33–52
8. Gigler FB (1992) Self-enforcing public disclosures. Working paper. University of Minnesota
9. Haw I, Hu B, Lee J, Wu W (2012) Investor protection and price informativeness about future earnings: international evidence. *Rev Account Stud* 17(2):389–419
10. LaFond R, Watts RL (2008) The information role of conservatism. *Account Rev* 83 (2):447–478
11. Lundholm R, Myers LA (2002) Bringing the future forward: the effect of disclosure on the returns-earnings relation. *J Account Res* 40(3):809–839
12. Newman P, Sansing R (1993) Disclosure policies with multiple users. *J Account Res* 31:92–112
13. Sloan G (2001) Financial accounting and corporate governance: a discussion. *J Account Econ* 32(1–3):335–347
14. Yu J (2011) Stock price informativeness and corporate governance: an international study. *Int Rev Financ* 11(4):477–514

A Smart Design of Pre-processing Classifier for Impulse Noises on Digital Images

Jieh-Ren Chang, Hong-Wun Lin and Huan-Chung Chen

Abstract Even though many kinds of filters algorithms have been developed for cancelling impulse noise from digital images in the past few decades, the issue of image restoration and reconstruction is still regarded for many researchers nowadays. In this study, a smart design of pre-processing classifier is used to categorize several distribution types which were identified as noise by noise detection. This smart design can be used for extracting useful local information from the corrupted image and hopefully results more image details to preserve for image filter. The pre-processing classifier goes through four-phase detection procedures to determine the condition of central pixel of local image window by using the similarity between neighbouring pixels. Simulation results show that our algorithm has extraordinary effective and accurate ability for classifying the condition of noise type.

Keywords Image processing · Impulsive noise · Edge preservation · Fuzzy classifier

1 Introduction

Impulse noise is one of image noises, also called salt and pepper noise since its appearance as white and black dots imposed on an image. Images are corrupted by impulses stemming from signal acquisition (non-ideal sensors), bit errors in transmission (channel errors or decoding errors), or faulty hardware in storage media.

Numerous methods have been proposed to remove impulse noise from digital images in the literature. Linear filters are not able to effectively eliminate impulse noise, because they blur the edges of the image [1]. The median filter (MF) [2] is the most famous non-linear filter to suppress impulse noise. The performance of median filter is quite well but it falters when noise ratio exceeds 50 %. To overcome

J.-R. Chang (✉) · H.-W. Lin · H.-C. Chen
Department of Electronic Engineering, National Ilan University,
No. 1, Sec. 1, Shen-Lung Road, 26047 I-Lan, Taiwan, ROC
e-mail: jrchang@niu.edu.tw

this situation, adaptive median filter (AMF) [3] used variable window size to achieve effective removal of impulse. However, AMF increased windows size at higher noise densities will leads to blurring and distortion of the result of image processing. These types of filters can cause sharp corners and thin lines disappear after filtering.

Decision based algorithm (DBA) [4] replaced the corrupted pixels by their neighboring pixels using predefined rules by local features. The process of replacement of neighboring pixel subsequently will produce streaking effect for an image with higher noise ratio. The streaking effect usually occurs evidently in the image border, since there are lacks of referred pixels at those positions. The most recent version of DBA, “modified decision based asymmetric trimmed median filter” (MDBUTMF) [5] made use of the concept of simple logic classifier to choose two types of filters separately for high noise or low noise density to reduce the flaw of streaking and maintain detail.

Recently, directional filter and edge preservation algorithm [6–10] have been widely discussed because of its excellent noise reduction capabilities. It successfully overcome those filters cannot distinguish thin lines from impulses noised image by going through four one-dimensional Laplace operators. Each local region of image is convolved with four Laplace operators separately then the minimum absolute value of these four convolution results is used to compare with a threshold value for edge detection. So, the threshold will seriously affect the performance of edge detection. It is not easy to derive an optimal threshold value through analytical formulation [10, 11]. In addition, the condition is required that at least one pair of pixels in desired direction is noise free simultaneously for directional filter. It is quite hard to meet as the noise density is increased.

According to above overview of filter design, some exceptions cannot be avoided that were not expected by each filter designer. Under the circumstance of the different noise conditions, an intuitive solution is that each noise condition should be considered to select suitable corresponding filter. Therefore, a more precise preprocessing classifier is appropriate to handle each of noise conditions for integrating various filters. We look forward to the more delicate image will be obtained in an integrated filters system that is based on the preprocessing classifier.

The paper is organized as follows: Sect. 2 introduce Preliminary work, Sect. 3 presents the design of fuzzy pre-processing classifier, Sect. 4 shows simulations and results. Finally, Conclusions present in Sect. 5.

2 Noise Model

In this study, we consider grey-scale images which are corrupted by impulse noises. Consider the actual situation, salt-and-pepper is only a special case of “Noise Model 4” in [3]. So we choose the model 4 as the study object to make our algorithm more realistic and more general. The model is described in detail in [3]. Instead of two fixed values, impulse noise could be more realistically modelled by two fixed

ranges that appear at both ends with a length m_1 , m_2 respectively, except that the densities of low-intensity impulse noise and high-intensity impulse noise are unequal. That is, for each image pixel at location (i,j) with intensity value o_{ij} , the intensity of corresponding pixel of the noisy image is given by x_{ij} , in which the probability density function of x_{ij} is

$$f(x_{ij}) = \begin{cases} \frac{p_1}{m_1}, & \text{for } 0 \leq x_{ij} < m_1 \\ \mathbf{1} - p, & \text{for } x_{ij} = o_{ij} \\ \frac{p_2}{m_2}, & \text{for } 255 - m_2 < x_{ij} \leq 255 \end{cases} \quad (1)$$

where p is the noise density, p_1 is the noise density for $x_{ij} < m_1$, p_2 is the noise density for $x_{ij} > 255 - m_2$, $p = p_1 + p_2$ and $p_1 \neq p_2$ in general.

We used the effective BDND method [3] for detection of impulse noise in the stage of noise detection. Based on the same idea as [3], the first step of detect operation is to find the impulse noise boundaries m_1 and m_2 used to establish the model of salt and pepper noise. Then, in order to identify whether the pixel is corrupted, we build a binary matrix T with the same size ($M \times N$) of the input image X . I.e., at each pixel location (i,j) , the corresponding element of mark matrix T will be created by using the following equation:

$$t_{ij} = \begin{cases} \mathbf{1}, & 0 \leq x_{ij} < m_1 \\ \mathbf{1}, & 255 - m_2 < x_{ij} \leq 255 \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (2)$$

where x_{ij} is the pixel intensity value at location (i,j) and $1 \leq i \leq M$, $1 \leq j \leq N$. $t_{ij} = \mathbf{1}$ represents a corrupted candidate pixel, while $t_{ij} = \mathbf{0}$ represents an uncorrupted pixel to be retained.

3 A Smart Pre-processing Classifier

3.1 Definition of Classifier Input

The proposed Smart Pre-Processing Classifier adopts the directional-correlation dependent filtering technique which had been shown that has a good capability for preserving edges and fine details of images. However, this technique is modified as a double convolution operation here to achieve the goal of second detection. Each of the corresponding mark sub-matrix T_{ij} and the input image window X_{ij} are convolved with a set of convolution kernels respectively for the pixel location (i,j) . Four convolution results are obtained respectively correspond to each direction. We take the absolute values of four results of convolution values $V(s)$ as the intensity homogeneity information of the neighborhood pixel. This computation strategy can be formulated as follows:

$$V(s) = |X_{ij} \otimes \mathbf{Ker}(s)|, \quad s = 1, 2, 3, 4 \quad (3)$$

where $\mathbf{Ker}(s)$ is the s th directional kernel, and \otimes denotes the convolution operator.

In order to extract more features that hidden in the local information, these four values of $V(s)$ can be used in combination to provide reliable decisions. $V(s)$ is required to pass an ascent sorting operation, and we define Eqs. (5)–(7) as three groups of A , B , and C respectively, then the average value of $R(s)$ is calculated by (8).

$$R = \mathit{Sort}(V) \quad (4)$$

$$A(s - 1) = R(s) - R(1), \quad S = 4, 3, 2 \quad (5)$$

$$B(s - 2) = R(S) - R(2), \quad S = 4, 3 \quad (6)$$

$$C = R(4) - R(3) \quad (7)$$

$$dav = \mathit{mean}(R) \quad (8)$$

A , B , C , and dav are used as 7 input variables of the classifier as directional homogeneity information to estimate the distribution of impulse noise.

3.2 If-Then Rules for Classification

In order to preserve image details and use correct information for filter process, the pre-processing classifier is used to divide the distribution of impulse noise into five types as follows:

[Uniform region type]: In this type of noise, all elements' values of R should be below a certain level (dav is small) since current pixel value are similar to neighboring pixels, and the differences among all elements of R are close to each other (A , B , and C are small).

[Isolated type]: The neighboring pixels are noise free but only central pixel is corrupted in this type. In this case, all elements' values of R must be not small (dav is also not small). The differences among all elements of R are close to zero (A , B , and C are small).

[Diagonal line type]: In this type, the distribution of all elements' values in R is obviously divided into two groups. One group is small ($R(1)$ and $R(2)$), another group is not small ($R(3)$ and $R(4)$), and the difference between two elements' values in the same group of R is small or equal (C is small). This distribution leads to average value dav can be approximately calculated by using $\frac{R(3)-R(2)}{2}$ or $\frac{R(4)-R(2)}{2}$. In other words, $(\frac{1}{2}) \times B(1)$ is very close to dav and $(\frac{1}{2}) \times B(2)$ is very close to dav .

[Edge type]: The researchers classified all edges into three types which are strong edge, weak edge and false edge respectively [12–14]. There are some similar feature between weak edge and false edge. So the Edge type is considered as real

edge in image, which should be preserved in filtering process. The rule is extracted as same as the concept of “Diagonal line”. One group $R(2)$, $R(3)$ and $R(4)$ is not small (B and C are small), another group $R(1)$ is not large and dav can be calculated only using $(\frac{3}{4}) \times A(1)$, $(\frac{3}{4}) \times A(2)$ or $(\frac{3}{4}) \times A(3)$.

[Other type]: This type is default output for those unclassifiable conditions.

Finally, we use 7 input variables as directional homogeneity information to estimate the distribution of impulse noise. These input values indicate in which degree the local region can be classified as some types of detail. The output of classifier is calculated by the following If-Then rules:

Rule 1: IF dav is small AND group A is small AND group B is small AND C is small
THEN noise distribution type is uniform region type.

Rule 2: IF dav is not small AND group A is small AND group B is small AND C is small
THEN noise distribution type is isolated type

Rule 3: IF C is small AND group $|\frac{1}{2} * B - dav|$ is small
THEN noise distribution type is diagonal line type.

Rule 4: IF group B is small AND C is small AND group $|\frac{3}{4} * A - dav|$ is small
THEN noise distribution type is edge type.

Rule 5: IF not above conditions
THEN noise distribution type is other type.

Fuzzy logic uses the whole interval of real numbers between zero (False) and one (True) to develop logic as a basis for rules of inference. Fuzzy rule of inference enables computers to make decisions using fuzzy reasoning rather than traditional logic reasoning. Fuzzy relation in different product space can be combined with each other by the operation which contains some conjunctions (AND operators) and disjunctions (OR operators). There are a number of qualified methods can be formulated to calculate the fuzzy relation. The operation of “max-min composition” is adopted for easy and fast application in this research.

4 Simulation Results

In this section, the proposed scheme is tested with a total of eight standards 512×512 , 8-bit grayscale images (Lena, Boat, Cameraman, Plane, Lake, Baboon, Pepper, and Goldhill). This set of standard test images contains various characteristics and suitable to test the performance of our pre-processing classifier. The impulse noise with uneven distribution is added to these images, which means that the pixel has unequal probability of being corrupted by either a positive impulse or a negative impulse.

In order to evaluate the performance and robustness of classifiers for impulse noises, the classification accuracy rate is calculated for each distribution type of

Table 1 Classification accuracy rate of each impulse noise condition is calculated for the proposed pre-processing classifier on the Lena images

Noise %	Classification accuracy rate				
	Uniform region (%)	Isolated (%)	Diagonal line (%)	Edge (%)	Other (%)
20 %(5 + 15)	100	100	99.75	99.76	99.64
40 %(30 + 10)	100	99.94	98.87	98.85	97.99
60 %(20 + 40)	100	99.37	98.36	98.12	97.14
80 %(50 + 30)	100	97.52	97.42	96.99	96.39
90 %(40 + 50)	100	95.11	96.73	96.24	95.74

Table 2 Classification accuracy rate of each impulse noise condition is calculated in proposed pre-processing classifier on the Baboon images

Noise %	Classification accuracy rate				
	Uniform region (%)	Isolated (%)	Diagonal line (%)	Edge (%)	Other (%)
20 %(5 + 15)	100	100	99.63	98.76	99.10
40 %(30 + 10)	100	99.52	98.52	97.54	97.15
60 %(20 + 40)	100	99.01	97.12	96.75	96.74
80 %(50 + 30)	100	96.56	96.54	95.57	95.12
90 %(40 + 50)	100	94.17	93.45	93.56	94.75

impulse noise condition in Table 1. The evaluation selects the “Impulse Noise Model 2” in which grey level 0 represents pepper and 255 represents salt with unequal probabilities. It can prove the pre-processing classifier has good performance even in unequal noise probabilities. The measurements of the experimental results are based on the standard metrics for evaluations of accuracy rate in this study. The robustness of a classifier is defined as “the ability of insensitiveness of classification algorithms to noise corruptions”. Simulations results show that our proposed detection algorithm achieves very amazing accurate classification rate and robustness for each noise level from Tables 1, 2 and 3. From the test results of Lena

Table 3 Classification accuracy rate of each impulse noise condition is calculated in proposed pre-processing classifier on the eight test images (average)

Noise %	Classification accuracy rate				
	Uniform region (%)	Isolated (%)	Diagonal line (%)	Edge (%)	Other (%)
20 %(5 + 15)	100	100	99.70	99.66	99.63
40 %(30 + 10)	100	99.91	98.83	98.80	98.31
60 %(20 + 40)	100	99.50	98.17	98.23	97.65
80 %(50 + 30)	100	97.83	97.31	97.04	96.73
90 %(40 + 50)	100	95.56	96.17	96.14	95.89

and Baboon, the success rate of edge identification don't significantly decrease by the complexity of the lines such as human's hair and animal's whisker in the proposed classifier. More importantly for mention, the performance of edge detection in the proposed classifier is not dependent on unknown thresholds evaluations which were used in the directional filter.

5 Conclusions

In this paper, a smart pre-processing classifier is proposed to categorize corrupted pixel into five distribution types of impulse noise before filtering stage. Four-phase detection procedures are used by the similarity between neighbouring pixels to extract more information of local image in fuzzy classifier. The pre-processing classifier is used to control filtering methodology that is the most crucial stage of switching median filter framework. Since the different filters can deal with different types of noise of local image effectively, the pre-processing classifier is suitable to be used for image processing.

Simulation results show that our algorithm has extraordinary effective and accurate ability in classifying the condition of noise type. A filters' system that can use our smart pre-processing classifier hopefully will be recommended for image restoration and reconstruction. The proposed method has great superiority to provide a different idea for future filtering strategy.

References

1. Toh KKV, Isa NAM (2010) Noise adaptive fuzzy switching median filter for salt-and-pepper noise reduction. *Signal Process Lett IEEE* 17:281–284
2. Jafar IF, AlNa'mneh RA, Darabkh KA (2013) Efficient improvements on the BDND filtering algorithm for the removal of high-density impulse noise. *IEEE Trans Image Process* 22:1223–1232
3. Ng PE, Ma KK (2006) A switching median filter with boundary discriminative noise detection for extremely corrupted images. *IEEE Trans Image Process* 15:1506–1516
4. Srinivasan KS, Ebenezer D (2007) A new fast and efficient decision-based algorithm for removal of high-density impulse noises. *Signal Process Lett IEEE* 14:189–192
5. Esakkirajan S, Veerakumar T, Subramanyam AN, PremChand CH (2011) Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter. *Signal Process Lett IEEE* 18:287–290
6. Li Z, Liu G, Xu Y, Cheng Y (2014) Modified directional weighted filter for removal of salt & pepper noise. *Pattern Recogn Lett* 40:113–120
7. Lu CT, Chou TC (2012) Denoising of salt-and-pepper noise corrupted image using modified directional-weighted-median filter. *Pattern Recogn Lett* 33:1287–1295
8. Xuming Z, Youlun X (2009) Impulse noise removal using directional difference based noise detector and adaptive weighted mean filter. *Signal Process Lett IEEE* 16:295–298
9. Chen PY, Lien CY (2008) An Efficient Edge-Preserving Algorithm for Removal of Salt-and-Pepper Noise. *Signal Processing Letters, IEEE* 15:833–836

10. Lien CY, Huang CC, Chen PY, Lin YF (2013) An efficient denoising architecture for removal of impulse noise in images. *IEEE Trans Comput* 62:631–643
11. Duan F, Zhang YJ (2010) A highly effective impulse noise detection algorithm for switching median filters. *Signal Process Lett IEEE* 17:647–650
12. Khaire PA, Thakur NV (2012) Image edge detection based on soft computing approach. *Int J Comput Appl* 51:12–14
13. Salman N (2006) Image segmentation based on watershed and edge detection techniques. *Int Arab J Inf Technol* 3:104–110
14. Chen CM, Lu HHS, Chen YL (2003) A discrete region competition approach incorporating weak edge enhancement for ultrasound image segmentation. *Pattern Recogn Lett* 24:693–704

An Effective Machine Learning Approach for Refining the Labels of Web Facial Images

Jieh-Ren Changn and Hung-chi Juang

Abstract The technique of search-based face annotation is implemented by mining weakly labeled facial images that are freely collected from the internet web sites but is incompletely correct label data. In this study, the particle swarm algorithm and binary particle swarm algorithm are used to achieve the technique of Unsupervised Label Refinement (ULR) for refining the labels of web facial images. The experimental data is provided from IMDb website and with 45 % initial incorrect label mark rate. The results show that the particle swarm algorithm and binary particle swarm algorithm have the better correction rate and convergence performance than other approaches.

Keywords Web facial images · Auto face annotation · Unsupervised learning · Particle swarm algorithm

1 Introduction

With advanced improvements in science and technology, a variety of digital cameras and mobile phones are incredibly popular. Taking pictures, posting immediately on the social network and storing them on the cloud space are very common nowadays. So the requirement of available storage space is also growing quickly, human face tagged digital photos on demand are even more important. Auto-tagging face images in this century will be very valuable research and application. In general, it will be a very good contribution to small individual households, social networking sites (e.g. Facebook, VK), police investigators or the mass media [1].

J.-R. Changn (✉) · H. Juang
Department of Electronic Engineering, National Ilan University,
No. 1, Sec. 1, Shen-Lung Road, Yi-Lan 260, Taiwan, ROC
e-mail: jrchang@niu.edu.tw

H. Juang
e-mail: asd1038@yahoo.com.tw

The traditional methods have two trouble processes to automatically tag faces [2–4] on digital images. The first one is to collect large and completely correct facial images with a training process for a complicated human face mark system. The second one is the new training process that must be done again when new data is added into the database [5]. These all take a lot of computation time.

A search-based face annotation (SBFA) framework was proposed for auto face annotation architecture in order to effectively address time-consuming problems in data collection [5–9]. Compared with the traditional photos database, many marked face photos have been used from the internet on a search-based framework and these photos are readily available. The purpose of SBFA is to assign correct name labels to a given query facial image. Actually, we first retrieve a short list of the top K most similar facial images from a weakly labeled facial image database for a new facial image to be identified, and then annotate the facial image by taking vote on the labels associated with the top K similar facial images. But there are also many wrong labels which have been marked on face photograph in this database, called weakly labeled facial image database. An Unsupervised Label Refinement (ULR) [5] was proposed by Dayong Wang, Steven CH Hoi and Ying to correct and fix the weakly labeled facial image database.

A basic assumption of ULR technology is that there are similar facial images' feature values for the same person's photos. A wrongly labeled photo could be remarked by comparing the most similar photos in the traditional ULR technology. Gradient descent method or coordinate descent method was often used to get the optimum label refinement for face annotation in the traditional ULR technology. Both the methods of gradient descent and coordinate descent usually get trapped into a local optimal solution [10, 11].

Therefore, this study considers to get the global optimal solution by using the particle swarm algorithm (PSO) [12]. PSO is originally developed by Kennedy, Eberhart and Shi, and was intended for simulating social behavior, as a stylized representation of the movement of organisms in a bird flock. In this algorithm, each particle has its own speed, and make adjustments based on past experience and group behavior themselves in the searching process. PSO is a computational intelligence technique that optimizes a problem by iteratively trying to improve a seeker solution with regard to a given cost function, and it can be easily implemented with the computer program. There were many good applications in various fields, such as the traveling salesman problem [13] and the n queens problem [14], which were solved by using PSO. PSO can also be combined with other methods to solve many problems, such as the membership functions adjustment in fuzzy control [15], the parameters of neural network [16], the classification in data mining [17] and so on.

The original PSO algorithm is to maximize or minimize objective function of real numbers. In this study, a face mark matrix is built to label the name for each facial image. Every element of face mark matrix is only marked by 1 or 0, so the cost function of binary number is suitably built up in this problem. A novel binary PSO [18] with a new definition for the velocity vector is used for this study. Two velocity vectors represent the probability of the bits of the particle changing to zero

and the probability that bits of particle change to one for each particle are introduced, and the probability limitation is added in this study.

The rest of this paper is organized as follows. In Sect. 2, the related research works are introduced. In Sect. 3, it describes how to use PSO and binary PSO algorithm to achieve Unsupervised Label Refinement technique. Section 4 describes the results of this experiment and analysis. In Sect. 5, some conclusions are presented and several recommendations are suggested for future research.

2 The Related Work

The related work is divided into three parts to be described as follows, Unsupervised Label Refinement, the method of gradient descent and the coordinate descent method. In general, the method of multi-step gradient descent and the coordinate descent are used for solving the ULR optimization problem.

2.1 Unsupervised Label Refinement

We first define $X \in R^{n*d}$, the extracted facial image features matrix, a collection of all of the face characteristic values, where n represents the number of facial images and d represents the number of feature dimensions. For the automatic learning purposes, the terminology of graph-based learning methodology is used. A sparse graph is built by computing a weight matrix $W = [W_{ij}] \in R^{n*n}$, where W_{ij} represents the similarity between two facial images (X_i and X_j) defined as follows:

$$W_{ij} = \begin{cases} e^{-\frac{\|x_i-x_j\|_2^2}{2\sigma^2}} & \text{if } X_i \in X_j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Further, the definition of the initial raw label matrix $Y \in [0, 1]^{n*m}$ represents the initial weak label information which is collected from the internet web, where n denotes the number of facial images, m denotes the total number of human names, $Y_{ij} = 1$ represents that the i -th facial image X_i is marked the j -th name label. So the vector Y_{i*} only has one element tagged with 1.

In order to improve the initial raw label matrix Y , a key assumption of “label smoothness” is proposed, i.e., the more similar features of two facial images, the more likely they share the same labels. The purpose of the ULR is to learn a refined label matrix $F \in R^{n*m}$ to improve matrix Y . To achieve this challenge, a graph-based learning solution based on the label smoothness principle was proposed. It can be formally formulated as an optimization problem of minimizing the following loss function $E_s(F, W)$:

$$E_s(F, W) = \frac{1}{2} \sum_{i,j=1}^n W_{ij} \|F_{i*} - F_{j*}\|_F^2 = \text{tr}(F^T L F) \tag{2}$$

where $L = D - W$, $D_{ii} = \sum_{j=1}^n W_{ij}$. Although the original weakly labeled Y is not a fully correct mark matrix for all facial images, it still has many right mark elements for the reference. So a regularization term $E_p(F, Y)$ is defined as follow:

$$E_p(F, Y) = \|(F - Y) \circ S\|_F^2 \tag{3}$$

where $S = \text{sign}(Y_{ij})$ is a sign matrix, \circ represents the Hadamard product (i.e., the entry wise product) between two matrices.

Finally, taking into account the sparsity of real facial mark matrix, a sparsity regularizer $E_e(F)$ is introduced by following the “exclusive lasso” technique [19]:

$$E_e(F) = \sum_{i=1}^n (\|F_{i*}\|_1)^2 \tag{4}$$

Therefore, combining formulas (2), (3), and (4), the objective function is defined as follows:

$$O(F) = E_s(F, W) + \alpha E_p(F, Y) + \beta E_e(F) \tag{5}$$

$$F^* = \arg \min_{F \geq 0} A(F) \tag{6}$$

where α and β are the weight values, and must be greater than zero.

2.2 Multi-step Gradient Algorithm

A search technique based on gradient descent is to minimize an objective function that can find the best solution by following the steepest direction. Methods in which the next iterate depends not only on the current iterate but also on the preceding ones are called multi-step methods. The Multi-step gradient method is easy to implement and requires modest computations for ULR optimization problem. Local optimal solution is often gotten from this method by fixed-point iterations and achieves improved convergence rates.

2.3 Coordinate Decent Algorithm

The coordinate descent method is a non-derivative optimization algorithm in which only one coordinate of the current iterate is updated at a time. Each iteration of

coordinate descent consists of picking a random coordinate descent and then performing a descent step on that coordinate to find the best optimum solution.

3 PSO and Binary PSO for ULR

The function $O(F)$ is neither a simple concave nor a simple convex function, so the multi-step gradient descent and coordinate descent are easier to get stuck in a local minimum [10, 11]. In contrast, particle swarm algorithm can effectively solve this problem. Two methods will be studied, which are PSO and Binary PSO algorithm respectively.

3.1 Particle Swarm Optimization for ULR

Since the refined label matrix, $F \in R^{n \times m}$ is a sparse matrix and only one element is allowed to be labeled in each row, an encode skill is applied and the n elements of vector P is defined as following formula:

$$P = \{P_i | P_i = h, F_{ih} = 1; F_{ij} = 0, j \neq h\} \tag{7}$$

For example, F is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, P is converted to $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

The vector P is used in the PSO algorithm to find the optimum solution, and then it has to be reversed to F by the following formula:

$$F_{i*} = \{F_{ij} | F_{ij} = 1, j = P_i; F_{ij} = 0, j \neq P_i\} \tag{8}$$

In PSO algorithm, there are many initial particles provided randomly in the search space for starting point of all searching path. Each particle has their own speed, and to search and adjust they are based on experience of past experience and group behavior. The moving speed of the k th particle at time t , ${}^k v^t$ is calculated as:

$${}^k v^t = w * {}^k v^{t-1} + c_1 * rand() * ({}^k p - {}^k P^{t-1}) + c_2 * rand() * (g - {}^k P^{t-1}) \tag{9}$$

$${}^k P^t = {}^k P^{t-1} + {}^k v^t \tag{10}$$

where ${}^k P^t$ is the position of the k th particle in time t , ${}^k p$ is the best visited position of the particle for the k th particle, the best position explored so far is g , c_1 and c_2 are positive constants, and $rand()$ is random variable with uniform distribution between 0 and 1. In this formula, w is the inertia weight which shows the effect of previous

velocity vector on the new vector. The algorithm for the PSO in ULR can be summarized as follows:

1. Initialize the particle swarm position ${}^kP^0$, the position of particles are randomly initialized within integer 1 to m .
2. Using the formula (8) to get F .
3. Calculate $O(F)$ by formula (5).
4. Using the formula (7) with F to get P .
5. Using the formula (9) and (10) to update the ${}^kP^t$ then convert it into an integer.
6. Update kP and g .
7. Determine whether the speed of each particle tends to 0? If the answer is “Yes” then g is the best solution. If the answer is “No” then go to step 2.

3.2 Binary Particle Swarm Optimization for ULR

Using the PSO for ULR as above mention, it needs to encode the labeled matrix into the vector and the results need to be converted into integer numbers. It may possibly cause the truncation errors by above traditional PSO procedures. So the Binary particle swarm optimization (BPSO) can solve this problem for ULR. First a binary particle swarm matrix Z is created for BPSO algorithm. The relationship between Z and vector P is as follows:

$$P_i = \sum_{j=1}^m Z_{ij} * 2^{j-1} \tag{11}$$

For example, P is $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, then Z is $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$. And the next particle state is

computed as follows:

$$Z_{ij}(t+1) = \begin{cases} \bar{Z}_{ij}(t), & \text{if } r_{ij} < V_{ij}^c \\ Z_{ij}(t), & \text{if } r_{ij} > V_{ij}^c \end{cases} \tag{12}$$

where \bar{Z}_{ij} is the 2’s complement of Z_{ij} , and r_{ij} is a uniform random number between 0 and 1. V_{ij}^c is the probability of the bits of the particle to change to zero or the probability that bits of particle change to one. The relevant formula as follows:

$$V_{ij}^c = \begin{cases} V_{ij}^1 = wV_{ij}^1 + q_{ij,1}^1 + q_{ij,2}^1, & Z_{ij} = 0 \\ V_{ij}^0 = wV_{ij}^0 + q_{ij,1}^0 + q_{ij,2}^0, & Z_{ij} = 1 \end{cases} \tag{13}$$

$$\begin{cases} q_{ij,1}^1 = -c_1 r_1 q_{ij,1}^0 = c_1 r_1, P_{ibest}^j = 0 \\ q_{ij,1}^1 = c_1 r_1 q_{ij,1}^0 = -c_1 r_1, P_{ibest}^j = 1 \end{cases} \tag{14}$$

$$\begin{cases} q_{ij,2}^1 = -c_2 r_2 q_{ij,2}^0 = c_2 r_2, P_{gbest}^j = 0 \\ q_{ij,2}^1 = c_2 r_2 q_{ij,2}^0 = -c_2 r_2, P_{gbest}^j = 1 \end{cases} \tag{15}$$

where the best position visited so far for a particle is P_{ibest}^j and the global best position for the particle is P_{gbest}^j , c_1 and c_2 are two fixed variables which are determined by user, r_1 and r_2 are two random variables in the range of (0, 1). The algorithm for the BPSO in ULR can be summarized as follows:

1. Initialize the particle swarm position Z with each element 0 or 1.
2. Initialize V_{ij}^c with a random number from 0 to 1.
3. Using the formula (11) to get P from Z and using formula (8) to convert P into F
4. Calculate $O(F)$ by formula (5).
5. Using the formula (7) to get P from F and using formula (11) to get Z from P .
6. Update P_{ibest}^j and P_{gbest}^j .
7. Update Z_{ij} By (12), (13), (14), (15) to move the particles.
8. Determine whether the speed of each particle swarm tends to 0?
 If the answer is “Yes” then P_{gbest}^j is the best solution.
 If the answer is “No” then go to step 3.

Also in order to prevent the V_{ij}^c being greater than 1 or less than 0, so the restrictions are set as follows:

$$\begin{cases} V_{ij}^c = 0 \text{ if } V_{ij}^c < 0 \\ V_{ij}^c = 1 \text{ if } V_{ij}^c > 1 \\ w + c_1 + c_2 = 1 \end{cases} \tag{16}$$

4 Experimental Results and Analysis

In this study, a facial feature database is provided by Wang [5] in which a human name list is collected consisting of popular actor and actress names from the IMDb website <http://www.imdb.com>. The name list covers the actors and actresses who were born between 1950 and 1990.

For the experiment test of the reliability of ULR, a total of 200 pictures of 20 different peoples were selected. After that, we let the database with 45 % initial incorrect label mark rate to start this test. Figure 1 shows one example in the initial incorrect labeled database.

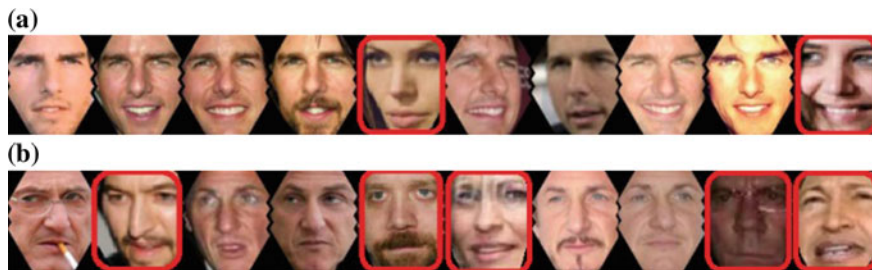


Fig. 1 Red box represents an error flag in the initial incorrect labeled database a Tom Cruise b Sean Penn

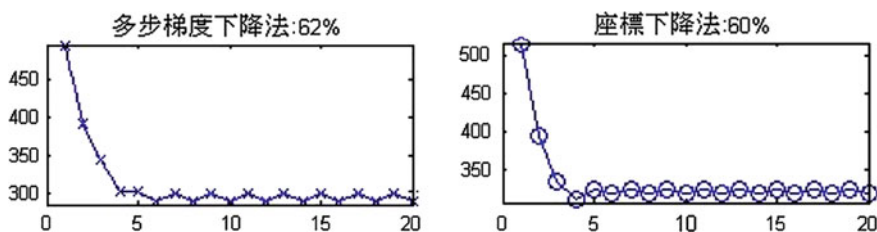


Fig. 2 Multi-step gradient and coordinate descent method convergence trend

4.1 Experiment by Multi-step Gradient and Coordinate Decent Algorithm

This experiment is to evaluate the convergence performance and correct labeled rate of the multi-step gradient and coordinate decent algorithms. Figure 2 show the evaluations on the objective functions of the two different algorithms. First of all, it is clear that all the algorithms converge quickly. Second, with the same formulation, we can see that the final objective values obtained by two different solvers are very close but the phenomenon of convergence there is an oscillation and tends to get a local minimum. The vertical axis represents the objective function value, and the horizontal axis represents the number of runs in Fig. 2. The correct labeled facial image rates were 62 and 60 % by the multi-step gradient descent and the coordinate descent methods respectively (Fig. 3).

4.2 Experiment by PSO and Binary PSO

This experiment is to evaluate the convergence performance and correct labeled rate of the PSO and the binary PSO algorithms. Figure 4 show the evaluations on the objective functions of the two different algorithms. The objective functions of both algorithms converge quickly, but the PSO for ULR needs a analog to digital

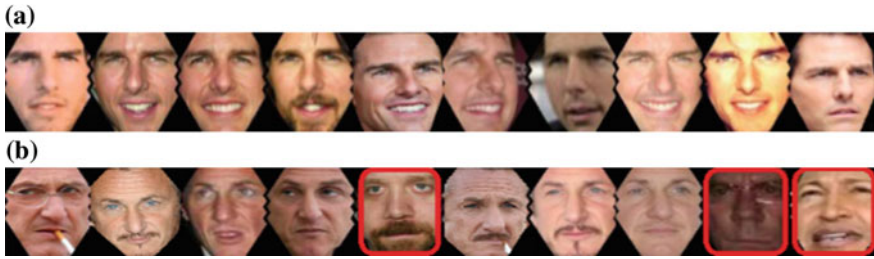


Fig. 3 Using the multi-step gradient method, the red boxes represent the error flag a Tom Cruise b Sean Penn



Fig. 4 Using the coordinate descent method, the red boxes represent the error flag a Tom Cruise b Sean Penn

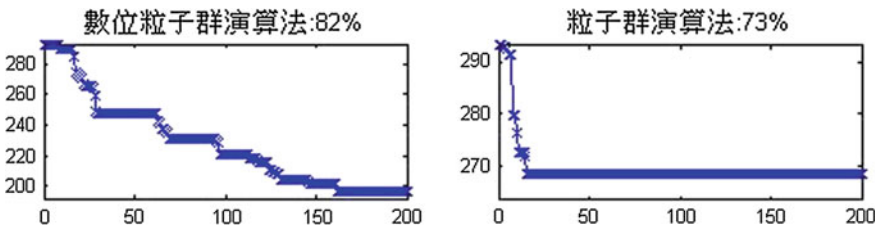


Fig. 5 Several particle swarm algorithm and particle swarm algorithm convergence trend

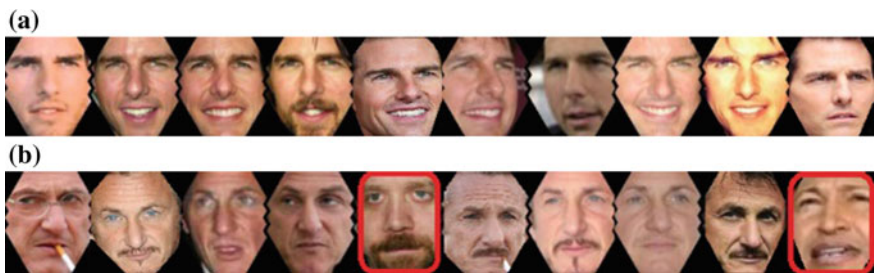


Fig. 6 The case of using particle swarm algorithm corrected, the red boxes represent the error flag a Tom Cruise b Sean Penn

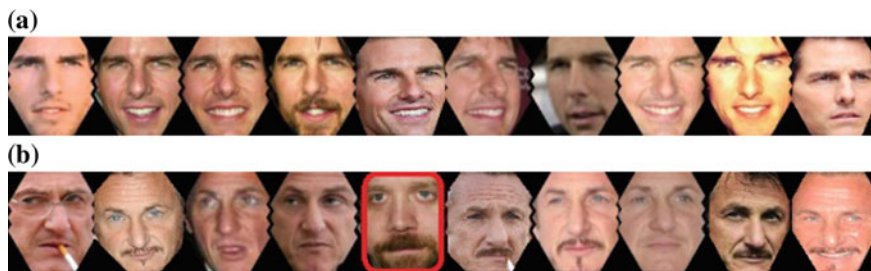


Fig. 7 The case of using digital particle swarm algorithm corrected, the *red boxes* represent the error flag **a** Tom Cruise **b** Sean Penn

conversion. It may possibly cause the truncation errors by traditional PSO procedures. So the BPSO can solve this problem for ULR with higher accuracy rate. The correct labeled facial image rates were 82 and 73 % by the BPSO and the PSO methods respectively (Figs. 5, 6 and 7).

5 Conclusions

In this study, the particle swarm algorithm and the binary particle swarm algorithm are applied to implement the technique of Unsupervised Label Refinement (ULR) for refining the labels of web facial images. A total of 200 pictures of 20 different peoples were selected from a internet facial feature database in which a human name list is collected consisting of popular actor and actress names from the IMDb website. With 55 % initial correct label mark rate, different algorithms are applied to start this experiment. The results showed that the binary PSO and PSO algorithm have the better solution compared with Multi-step gradient decent and coordinate decent for ULR technology. In addition the study also found that final results did not achieve very high correction rate in the final. It may be not considered due to the different camera angles, or different people may have similar facial characteristics value. We hope in the future some more factors will be added in this unsupervised mark trimming technology to improve the accuracy of the final result.

References

1. Satoh S, Nakamura Y, Kanade T (1999) Name-it: naming and detecting faces in news videos. *MultiMedia IEEE* 6:22–35
2. Berg TL, Berg AC, Edwards J, Maire M, White R, Yee-Whye T, Learned-Miller E, Forsyth DA (2004) Names and faces in the news. In: *Proceedings of the 2004 IEEE computer*

- society conference on computer vision and pattern recognition, vol 842. CVPR 2004, pp II-848–II-854
3. Yang J, Hauptmann AG (2004) Naming every individual in news video monologues. In: Proceedings of the 12th annual ACM international conference on multimedia. ACM, New York, pp 580–587
 4. Asthana A, Lucey S, Goecke R (2011) Regression based automatic face annotation for deformable model building. *Pattern Recogn* 44:2598–2613
 5. Dayong W, Hoi SCH, Ying H, Jianke Z (2014) Mining weakly labeled web facial images for search-based face annotation. *IEEE Trans Knowl Data Eng* 26:166–179
 6. Dayong W, Hoi SCH, Ying H, Jianke Z, Tao M, Jiebo L (2014) Retrieval-based face annotation by weak label regularized local coordinate coding. *IEEE Trans Pattern Anal Mach Intell* 36:550–563
 7. Wang D, Hoi SCH, He YH (2014) A unified learning framework for auto face annotation by mining web facial images. In: *CIKM*. ACM, pp 1392–1401
 8. Hoi SCH, Wang D, Cheng IY, Lin EW, Zhu J, He Y, Miao C (2013) FANS: face annotation by searching large-scale web facial images. In: Proceedings of the 22nd international conference on world wide web companion. International world wide web conferences steering committee, Rio de Janeiro, Brazil, pp 317–320 (2013)
 9. Wang D, Hoi SCH, Wu P, Zhu J, He Y, Miao C (2013) Learning to name faces: a multimodal learning scheme for search-based face annotation. In: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval. ACM, Dublin, pp 443–452
 10. Docherty PD, Schranz C, Chase JG, Chiew YS, Möller K (2014) Utility of a novel error-stepping method to improve gradient-based parameter identification by increasing the smoothness of the local objective surface: a case-study of pulmonary mechanics. *Comput Methods Programs Biomed* 114:e70–e78
 11. Patrascu A, Necoara I (2014) Random coordinate descent methods for ℓ_0 regularized convex optimization. *IEEE Trans Autom Control* 60:1811
 12. Shi Y, Eberhart R (1998) A modified particle swarm optimizer. In: The 1998 IEEE international conference on evolutionary computation proceedings, 1998. IEEE world congress on computational intelligence. IEEE, pp 69–73
 13. Wang K-P, Huang L, Zhou C-G, Pang W (2003) Particle swarm optimization for traveling salesman problem. In: 2003 international conference on machine learning and cybernetics. IEEE, pp 1583–1585
 14. Hu X, Eberhart RC, Shi Y (2003) Swarm intelligence for permutation optimization: a case study of n-queens problem. In: Proceedings of the 2003 IEEE swarm intelligence symposium, SIS'03. IEEE, pp 243–246
 15. Esmiri A, Aoki A, Lambert-Torres G (2002) Particle swarm optimization for fuzzy membership functions optimization. In: 2002 IEEE international conference on systems, man and cybernetics, vol 3. IEEE, p 6
 16. Salerno J (2012) Using the particle swarm optimization technique to train a recurrent neural model. In: 2012 IEEE 24th international conference on tools with artificial intelligence. IEEE Computer Society, pp 0045–0045
 17. Sousa T, Silva A, Neves A (2004) Particle swarm based data mining algorithms for classification tasks. *Parallel Comput* 30:767–783
 18. Khanesar MA, Teshnehlab M, Shoorehdeli MA (2007) A novel binary particle swarm optimization. In: Mediterranean conference on control and automation, MED'07. IEEE, pp 1–6
 19. Zhou Y, Jin R, Hoi S (2010) Exclusive lasso for multi-task feature selection. In: International conference on artificial intelligence and statistics. pp 988–995

Using the Data-Service Framework to Design a Distributed Multi-Levels Computer Game for Insect Education

Chih-Min Lo and Hsiu-Yen Hung

Abstract Data Services is a convenient mechanism to provide a service interface to access data from a database. Accordingly, providing data as a service not only encourages the access to data anywhere, at any time, but also reduces the cost of an application system implementation. The computer game is an interactive medium that can effectively to help us to learn some subjects such as a serious game. An interactive game may include a data exchange function to provide information to run the game in each different game level. This paper utilized a data service framework to build a game server for providing game information to the interactive game. And this paper also included a real case to demonstrate the feasibility and merits of this proposed computer game design approach.

Keywords Computer game · Serious game · Data service · Software framework · Interactive game · Educational game

1 Introduction

In recent years, computer games are a rapidly growing segment of the entertainment industry. Design and development of modern computer games can be a complex activity involving many participants from a variety of disciplines. And computer games design approaches typically appear to be less formalized than those used for other types of software application systems [1, 2]. In general, most of the game is needed to save the game playing information to the next level or stage using.

C.-M. Lo (✉)

Department of Digital Multimedia Design, National Taipei University of Business,
Taipei, Taiwan
e-mail: cmlotw@ntub.edu.tw

H.-Y. Hung

Department of Digital Media Design, Hwa Hsia University of Technology,
New Taipei, Taiwan
e-mail: yann@cc.hwh.edu.tw

The game playing information can be saved to a file or a database. A game server provides the game rules, communication among the players, and should limit includes the data provide service. An interactive computer game or real time game requires an efficient and parallel processing capability game server to serve game clients [3].

Data services are specifically tailored for information oriented tasks to deal with business service requirements, relying on the distributed architecture of consumer data processing mechanisms. The implementation of a data service is a server, can resolve inconsistencies between various applications involved in the exchange of data. A data services server differentiates data integration from business applications, which enhances the flexibility of processing and managing information [4]. In the past years, data services are used widely to implement to apply to the business information systems for providing data access services [5–8].

The data service framework is an application software framework. An application framework is an efficient approach to develop an information system, which integrates software modeling technique, code generators and software reuse technique. Most of the software developers have had it as their ultimate goal to achieve high quality software with time-saving development schedule and low development cost. Therefore, using an application framework to develop a software system is an efficient way [9–11].

In this paper, we utilize the data service framework into a game server for the data exchanging purposes. We use this method in the game software development, in order to reduce development costs, save development time and reduce the bug rates of data access programming. In addition, using the data service framework to build a game software system can be easy to set up game server, and good for testing of data access.

The remainder of this work is structured as follows. Section 2 presents the proposed game design framework for building a game server with data service functions. Section 3 provides a real case demonstrating the value of the proposed framework. The final section presents our conclusions.

2 Game Design

In this paper, we present a distributed multi-levels computer game with network to exchange game playing information. This game software system includes game client, game server and data service server. The game server implements by two parts, one is communication service, and other one is game rule service. Figure 1 shows the computer game software system architecture. A data service server can serve one or more game server to provide data access functions.

Most of software projects need to construct data manipulation code by programmers. In this way, a software project will increase the risk of software quality. In this paper, we decompose a computer game software system to three layers, includes game client, game communication service and data access service.

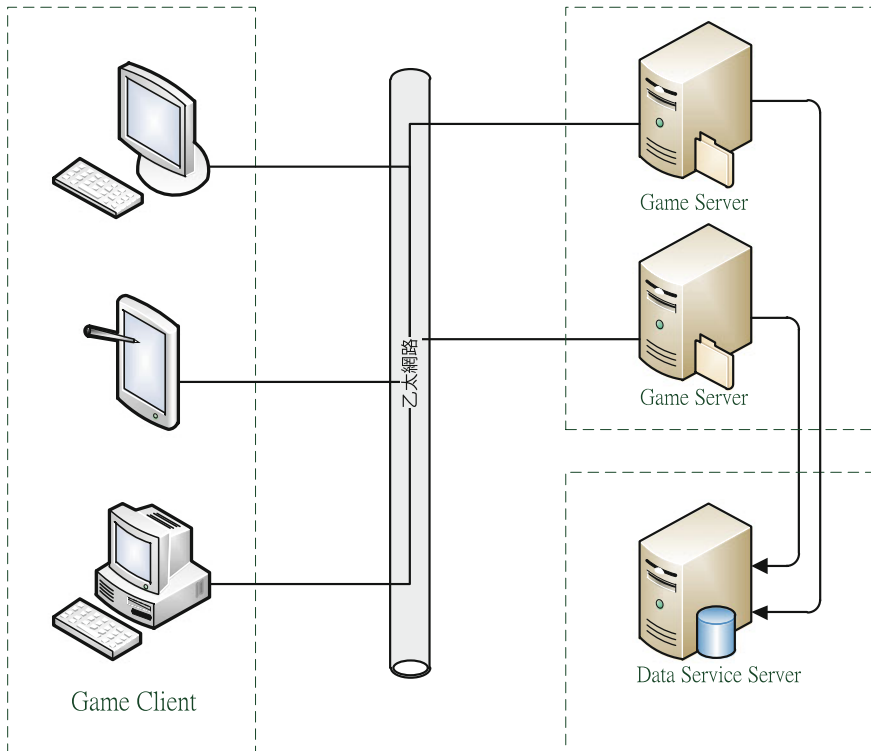


Fig. 1 The computer game system architecture

The data access service includes data access objects, data entities, data web services, data processing logic and data access helper (see Fig. 2). In the programming phase, we can use the code generator to generate code, which includes data access objects, data entities, data web services. And we use a data access framework to help system to manipulate the database. Therefore, we can save the programming time and reduce data access testing processes.

3 A Real Case

In this paper, we illustrate our framework with a real case, which the proposed computer game design approach. We implement an interactive computer game with a data service server for insect education that public in the Special Exhibition of Revealing the hidden code of insects in National Taiwan Science Education Center, Taipei, Taiwan, from Nov. 19, 2013 to Aug. 31, 2014. Figure 3 shows the game explanation of the Special Exhibition of Revealing the hidden code of insects.

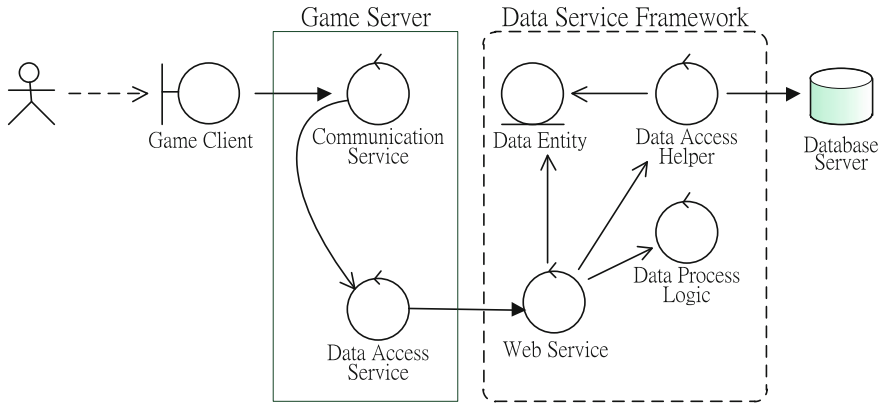


Fig. 2 The architecture of the data service framework

昆蟲對對碰
昆蟲的組成部分有軀體、觸角、口器、足部和翅膀，看看玻璃櫃中的各式各樣標本，牠們外型有什麼特色？展場內共有6個遊戲機台，請選擇正確的組件，完成心目中理想的昆蟲！答對者可向工作人員洽詢並領取昆蟲貼紙乙張
▲卡片與貼紙每日限量，發完為止

This is a puzzle game. There are 6 computers in total with different games in this exhibition. Please choose the right component to composite an insect. If you have the correct answer, you can win a sticker as a prize.

A-軀體Body B-觸角Antennae
C-口器Mouth D-足部Feet
E-翅膀Wings F-解答Answer

The poster also features a map of the exhibition area with stations labeled A through F, and illustrations of various insects and game components.

Fig. 3 The game explanation of the special exhibition of revealing the hidden code of insects

This game is a distributed multi-levels computer game for insect education, and it is a puzzle game. There are 6 game levels, and each level deploy to a computer. The player should choose the right component from each level, which executed in computer A, B, C, D, and E. In the game, players can composite an insect at the final level in computer F. If a player has the correct answer, he or she can win a

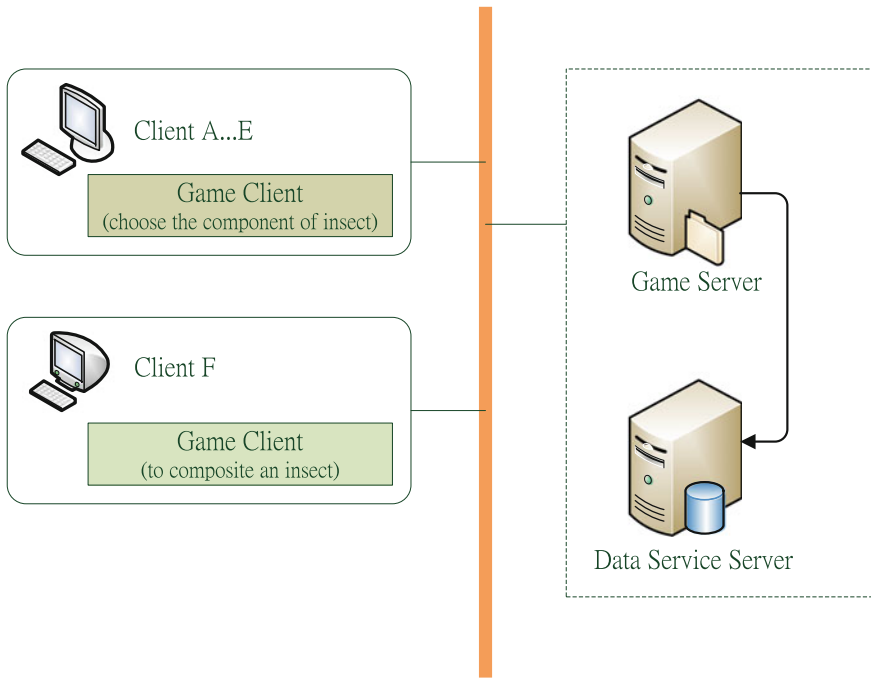


Fig. 4 Insect game software architecture

sticker as a prize, and if the player has an incorrect answer, he or she may return to the levels.

An insect will be broken down into five elements in the game, which are the body, antennae, mouth, feet, wings, and in the course of the game players need to understand the characteristics of the various parts of the insect and structure, and then choose to piece together parts of insects. Five parts of the insects and allows players the freedom to choose any component, so players in the process of interaction with the game in addition to experience the pleasure of creation can easily learn the relevant knowledge of insects.

In this case, we implement a data service mechanism into a game server as Fig. 4 shown. The data service includes all data manipulation functions and contact to a database server. In the programming phase, we use a code generator to create data access classes, data entity classes, web service classes and business logic class templates. At the run time, these generated class code is running on the data access framework. In this case, we found that to develop a game software system, using a data service framework can save a lot of time for the game development.

4 Conclusion and Future Work

Data services serve as a new form of application for the rapid implementation of distributed application systems that facilitate sharing and integration of data between applications. We recognize that it is the key for the rapid computer game software development to develop a mechanism which is capable of reducing the efforts, raising the quality, and advancing the software maintainability. In this real case we can reduce effort by over 80 % in the data access coding stage. On the contrary, the development effort without utilizing the data service framework should be 100 %. Accordingly, we can justify that this method is for computer game developers constructing distributed data services. Despite recent progress in the game server development with the data access integration, to develop game server with data services remains a challenge. In the nearly future, we hope to take more depth discussion for applying data service mechanism to game server development.

References

1. Taylor MJ, Gresty Liverpool D, Baskett M (2006) Computers in entertainment (CIE)—Theoretical and Practical Computer Applications in Entertainment, vol 4, no 1, Article No. 5, January 2006. ACM New York, NY
2. Amory A (2007) Game object model version II: a theoretical framework for educational game development. *Educ Tech Res Dev* 55:51–77
3. Yoo HS, Kim SW (2013) Simplified game specific description language for rapid game server development using LDD (Language Driven Development) framework. *Adv Sci Technol Lett* 39:123–130
4. Lo CM, Huang SJ (2012) Applying model-driven approach to building rapid distributed data services. *IEICE Trans Inf Syst* E95-D(12):2796–2809
5. Castro VD, Marcos E, Vara JM (2011) Applying CIM-to-PIM model transformations for the service-oriented development of information systems. *Inf Softw Technol* (53)1:87–105
6. Zhang XG (2008) Model driven data service development. In: *IEEE international conference on networking, sensing and control*, pp 1668–1673
7. Andukuri PV, Guo J, Pamula R (2011) Automatic integration of web services. In: *Proceedings of the 2011 international conference on internet technology and applications (iTAP)*, pp 1–4
8. Chung JY, Lin KJ, Mathieu RG (2003) Web services computing: advancing software interoperability, computer. IEEE Computer Society, pp 35–37
9. Hung HY, Lo CM (2013) A framework-based model-driven development approach to building information systems. *Adv Inf Sci Serv Sci* 5(10):1226–1233
10. Fayad M, Schmidt DC (1997) Object-oriented application frameworks. *Commun ACM* 40(10):32–38
11. Gachet A (2003) Software frameworks for developing decision support systems—a new component in the classification of DSS development tools. *J Decis Syst* 12(3):271–281

Financial Diagnosis System (FDS) for Food Industry Listed in the Taiwan Stock Exchange (TWSE)

Cheng-Ming Chang

Abstract This paper uses empirical financial data, which were announced on the website between 2007 and 2014, of food industry companies listed in the Taiwan Stock Exchange (TWSE) to verify the Financial Diagnosis System (FDS). The FDS sets up criteria by ranking the food industry companies' financial indicators and grouping each indicator into five classes from the best to the worst. A company, for those listed as well as unlisted in the stock exchange, will be evaluated its performance of the certain financial indicator by ranking it into one of the five classes as the best, the better, the average, the worse, the worst. The FDS also sets up five forces, which is stability, profitability, activity, Growth Potential, productivity, financial analysis criteria by classifying financial indicators into five forces, ranking each force into five classes from the best to the worst.

Keywords Financial diagnosis system (FDS) · Taiwan stock exchange (TWSE) · Empirical financial data · Financial indicator · Five forces

1 Introduction

Financial crisis of 2008 had injected the economic and stock market all over the world, especially in the electrical, car, construction, material and other non-living industries. On the other hand, the food industry had showed no obvious influences and had maintained more stability than the others' industries. These characteristics of food industry catch my eye to exploit the method to evaluate and to rank a company to find out how it was performed in the past and thereafter to predict how it will perform in the future. The food industry stability, high-quality companies which are listed in the Taiwan Stock Exchange (TWSE) were included, along with

C.-M. Chang (✉)

Department of Information Management, Hwa Hsia University of Technology,
Zhonghe District, Taiwan (ROC)
e-mail: cmchang@go.hwh.edu.tw

the long-term empirical financial data for 8 years will make the Financial Diagnosis System (FDS) more confidence and more reliable.

The FDS demonstrates the following characteristics:

Empirical financial data collected from open source website of Taiwan Stock Exchange Corporation for the period of 2007–2014.

Not just calculation of financial indicators but also ranking them into five classes from the best, the better, the average, the worse to the worst and analysis of the performance for certain company to find out where it stand on the five classes.

Sets up five forces, which is stability, profitability, activity, Growth Potential, productivity, financial analysis criteria by classifying financial indicators into five forces, then ranking each force into five classes from the best to the worst, and analysis of the performance for certain company to find out where it stand on the five classes.

Easy operation of the FDS by clicking on the selection menu to pick a company, to display diagnosis report. The diagnosis report will show by historical and industry.

Table 1 The five forces

Forces	Financial indicators
Stability	1. Fixed long-term suitable ratio = $\frac{\text{Fixed asset} + \text{long-term investment}}{\text{Net value} + \text{long-term debt}} * 100 \%$
	2. Debt ratio = $\frac{\text{total liabilities}}{\text{Net value}} * 100 \%$
	3. Equity capital ratio = $\frac{\text{Net value}}{\text{Total assets}} * 100 \%$
	4. Current ratio = $\frac{\text{Current asset}}{\text{Current liability}} * 100 \%$
Profit-ability	1. Gross profit margin = $\frac{\text{Gross margin}}{\text{Net sales}} * 100 \%$
	2. Operating profit margin = $\frac{\text{Operating profit}}{\text{Net sales}} * 100 \%$
	3. Profit margin = $\frac{\text{Income before tax}}{\text{Net sales}} * 100 \%$
	4. Return on total assets = $\frac{\text{Income before tax}}{\text{Total assets}} * 100 \%$
Activity	1. Inventory turnover ratio(times) = $\frac{\text{Sales cost}}{\text{Inventory}}$
	2. Receivable turnover rate(times) = $\frac{\text{Net sales}}{\text{Accounts receivable} + \text{notes receivable}}$
	3. Total assets turnover rate(times) = $\frac{\text{Net sales}}{\text{Total assets}}$
Growth potential	1. Sales growth rate = $\left(\frac{\text{Current net sales}}{\text{Last year net sales}} - 1 \right) * 100 \%$
	2. Sales grow rate per person = $\left(\frac{\text{Current sales per person}}{\text{Last year sales per person}} - 1 \right) * 100 \%$
Productivity	1. Output value per person (dollars) = $\frac{\text{Output value}}{\text{Number of direct employee}}$
	2. Output value per person (dollars) = $\frac{\text{Output value}}{\text{Total number of employee}}$

2 Data Description

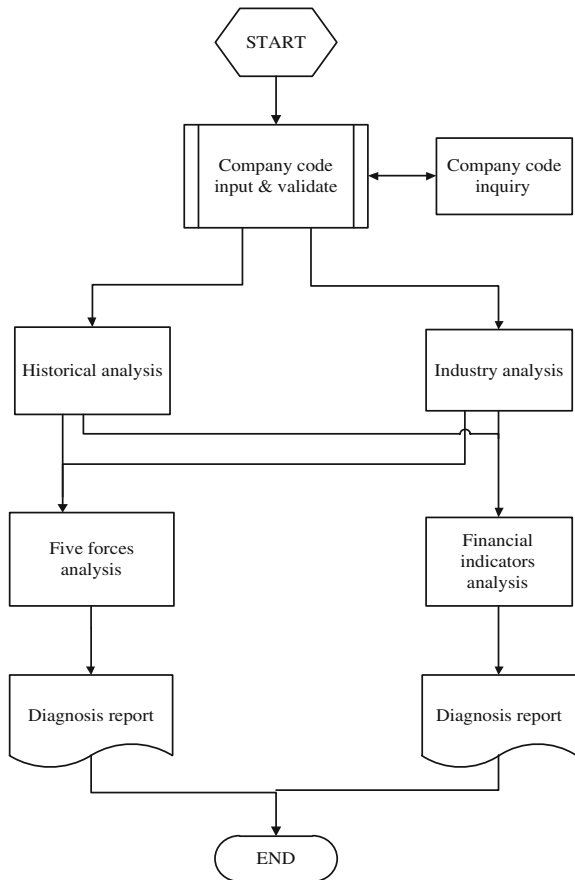
The companies of food industry listed in the Taiwan Stock Exchange (TWSE) were included. The empirical data included total number of employee, accounts of annual balance sheets and income statements of all sampled companies, collected for the period of 2007–2014.

The financial indicators and five forces include (Table 1).

3 System Implementation

See Fig. 1.

Fig. 1 System flowchart



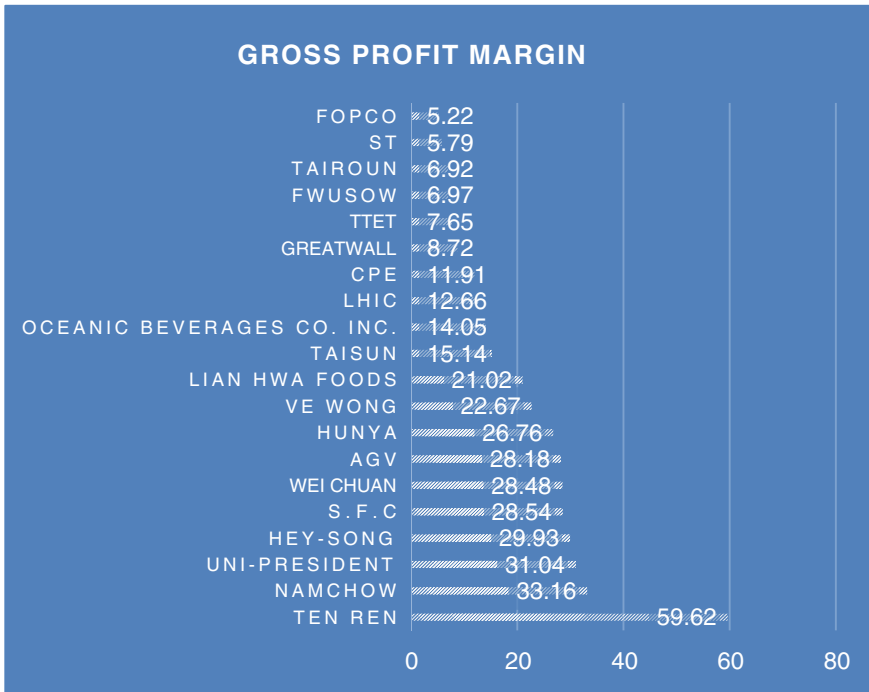


Fig. 2 Profitability comparison of all companies in the food industry

3.1 Industry Analysis

One of the five forces comparison of all companies in the food industry. Choose any ratio of the certain five forces (Fig. 2).

3.2 Historical Analysis

All ratios of the certain five forces historical comparison of a company for a period of years (Table 2).

3.3 Diagnosis Report for a Company

Rank the performance of a company comparative to others within food industry of each five forces into five levels, from one star to five stars, respectively represent the worst and the best (Table 3).

Table 2 Profitability comparison of a company for a period of years

UNI-PRESIDENT(1216)			
Year	Gross profit margin	Profit margin	Return on total assets
2007	27.28	6.06	7.35
2008	27.38	2.76	3.29
2009	30.92	5.17	5.07
2010	29.47	5.99	6.53
2011	28.43	4.68	5.37
2012	29.82	5.21	6.11
2013	30.93	5.98	6.35
2014	31.04	5.24	5.17

Table 3 Diagnosis report of five forces comparison for a company

UNI-PRESIDENT(1216)	
Five forces	Rank
Stability	*****
Profitability	***
Activity	****
Growth potential	**
Productivity	**

*each star represents up to 20 % of better performance rank among food industry

4 Conclusion

The FDS is proofed to be a faithful diagnosis system to evaluate the financial health for a food industry company. The future effort will put on the automatic data update programming. Hopefully, the data will be updated to the date automatically with the least human-effort involved. And I believed that this FDS will work for some industries and/or categories.

Classification Rule Discovery for Housing Purchase Life Cycle

Bo-Han Wu and Sun-Jen Huang

Abstract Housing purchase at different stages of life cycle of various people is a complicated but important decision. Classification rule mining is a common technology in the field of data mining. Classification Rules can be generalized to classify unknown samples or predict the future from the mining of historical data. This paper proposes a rule-based approach to housing purchase life cycle on demands through C4.5 classification method. An example, a real estate transaction data set in a government, is utilized in the paper to illustrate the utility of the proposed approach and further evaluated its prediction accuracy, precision and recall. The established prediction model with the aid of the derived classification rules helps various people make appropriate decisions following different housing purchase life cycle stages on demands.

Keywords Housing purchase life cycle · Housing demands · Classification rule · C4.5

1 Introduction

Housing purchase at different stages of life cycle of various people is a complicated but important decision. For young people, the public transportation, location, sense of safety, medical and health facility and education facility are top five factors affecting their housing purchase behavior [1]. For old people, they would like to choose their houses near the gardens, shops and public transportation [2]. In order to enhance working efficiency and qualities of services, Taiwanese government has been actively promoted computerization for many years. Therefore, there have been a large number of references which are valuable for people. Among the references, the ones relevant to real estate transaction are highly relative to life of people.

B.-H. Wu · S.-J. Huang (✉)

Department of Information Management, National Taiwan University of Science and Technology, No. 43, Sec. 4, Keelung Rd, Da'an District, Taipei 10607, Taiwan
e-mail: huangsj@mail.ntust.edu.tw

Most housing customers often pay a lot of emphasis on price and location. With the limited income, most of the housing customers have the goal of maximizing benefits to meet their satisfaction. The field of housing purchase research is mostly focused on external housing problems. For example, Malmberg [3] examined the housing market from low fertility. Hurst [4] established an energy efficiency rating systems from Australian real estate data. Luo and James [5] utilized multiple regression to analyze the influences of commercial housing advertisement on the housing buying behaviors.

With growing size of available data and databases, knowledge discovery from data became essential to determine a current condition or predict future behavior. Data mining techniques, used for achieving the above goals, can be classified into the following tasks: classification, estimation, prediction, clustering, association, and description. These techniques transform data to information to satisfy needs throughout an organization. Classification is probably the best understood of all data mining tasks. It seeks to build a model to find patterns that represent a description of a particular class. The task involves the process of finding some relations among the attributes and the predefined classes of a database. The process consists of two steps—learning and classification [6].

However, very little attention has been paid to analyze the real estate transactions data prior to predict housing purchase life cycle using rule-based approach. Hence, the aim of this study is to apply different data mining methods on demands reports for government to make appropriate decisions. This paper proposed the C4.5 algorithm to find the relations and rules about the housing purchase life cycle. We also used R language to implement data pre-processing. For the data analysis, WEKA software among other data-mining programs was chosen.

The structure of the paper is as follows: Sect. 2 briefly discusses the definition of housing purchase life cycle and rule-based approach; Sect. 3 describes the case studies of the real estate transactions data; Sect. 4 presents our empirical results and discusses the importance of classification rules. Finally, conclusions are made in Sect. 5.

2 Related Work

The objects of housing purchase are complicated. Over the life cycle, house purchase planning needs to be established. After the housing purchase planning is established, people can realize their demands for housing. Housing purchase life cycle means that people have different housing demands in different life stages. More specifically, house size can affect children's development [7]. Public transportation and community infrastructure are primarily considered by middle class residents [8]. The long-term care is primarily considered by the old people over 70 [9]. In a recent study, choosing good housing is good for children development, such as good cognition, active behavior, getting well-educated and getting nice jobs. In different stages, housing needs for children are different. When they are in their childhood, they need good neighborhood environment. In their studying stages, they need quiet

and independent learning space. Therefore, the young generations, middle-aged generations, or the old generations have different housing needs in different stages.

Data mining techniques have been successfully applied in similar fields. For example, Johnson [10] proposed spatial decision support system using data mining techniques for low-income families to select housing. Juan et al. [11] used data mining techniques to analyze the cost and quality of housing customization for non-professional customers to make appropriate decisions. However, most studies have focused either on the type of housing price or housing environment. This paper proposed the first study to discuss a housing purchase life cycle prediction based on classification rule.

2.1 Data Mining Techniques

Facing the age of data explosion, the amount of data is increasing very fast in databases. Those data generally involves hidden knowledge, and they can be used to improve the decision-making process of people. It is an analysis technique from statistics, machine learning method, mathematical algorithms. The techniques discovered unknown patterns and relationships in large data sets. It is also an integral part of cross industry standard process for data mining (CRISP-DM).

2.2 Classification Rules

Classification has been successfully applied to different fields [12]. Classification rule can be used to extract models and important data class or to predict categorical labels. When applied to housing purchase prediction, the rules become descriptive scenarios for life cycle of various people. The rule-based approach is extremely potential. This approach can classify the rules for people to meet their needs in housing purchase. A well-known application of classification rules is as follows: IF Conditions THEN Class. The conditions are a set of items in database, and each item includes three parts: "Attribute", "Operator" and "Value". "Attribute" of arguments is logical functions. They usually consist of "Operator" is "=". "Value" of arguments includes the numbers of "Attribute". For example, IF GENDER = male THEN age = 60. However, classification rules describe the relation between attribute conditions and class.

2.3 Classification by C4.5 Algorithm

Decision tree induction [13] and classification rule induction algorithms are often used to extract classification rules. C4.5 is one of the popular decision tree techniques, used for classification rule. It is easy to use C4.5 read and represent IF-THEN rules. With decision tree technique, a tree is constructed to model the classification

process. Once the tree is built, it is applied to each instance in the database. However, the C4.5 algorithm starts with the training data at root node of the tree. The attributes were selected through gain ratio to make these data into partition. Each decision of the classification rules can be traced on the path from the root node to leaf node.

3 Research Methodology

3.1 Description of the Data

In our study, the real estate transaction dataset used was from Department of Land, Taipei city government (DLT). In the case, five different version databases (Oracle SQL MS SQL) were integrated into real estate transaction dataset. These databases enable the users to input real estate transaction data, and display estate proprietors pertinent elements. More specifically, the databases were all obtained through the real-world system. We got the data from 2011 to 2013. Each instance of the datasets is a 'real-world' case. The dataset is composed of 214,839 transaction data. The attributes are shown in Table 1.

Table 1 List of the 22 attributes

Attributes	Description
CASETYPE	The transaction type of registration
CASEFLAG	The subjects of transaction of registration
BA	The housing addresses of transaction
ALLPRICE	The housing prices of transaction (NTD)
GENDER	The gender of estate proprietor
AGE	The age of estate proprietor
HR	The household register of estate proprietor
AREAS	The housing areas of transaction
ZONING	The land and housing zoning including (1) Business, and (2) housing
FLOOR	The housing floor of transaction
BUILD	The housing type of transaction
SIZE	The housing size of transaction (M ²)
BM	The housing materials of transaction
HC	The housing completion date (Year)
ROOM	The number of rooms
PARLOR	The number of parlor
BATHROOM	The number of bathroom
COMPARTMENT	To identify either compartment of housing or not
MANAGEMENT	To identify either management of housing or not
PARK	The type of parking including (1) plane parking space, and (2) mechanical parking space
PARKAREA	The size of parking
PARKPRICE	The price of parking

The aim of this study is to establish housing purchase life cycle model. Hence, we only selected 16 variables for our study: CASETYPE, CASEFLAG, ALLPRICE, GENDER, AGE, HR, AREAS, FLOOR, BUILD, SIZE, BM, ROOM, PARLOR, BATHROOM, COMPARTMENT, and MANAGEMENT. The description of attributes is shown in Table 2.

- CASETYPE: From these attributes, we only selected (1) purchase for our study.
- BA and AREAS: Both can be used to identify transaction areas and they are duplicated, BA is not considered in this experiment.
- PARK, PARKEA and PARKPRICE: In this dataset, parking data only have 12,537. Hence, three attributes are not considered in this experiment.
- ZONING: In this dataset, ZONING NULL data have 69,318. Hence, it is not considered in this experiment.
- AGE: The ages of the predicted target as defined by “Modeling the demand for housing over the life cycle” [14] and “Distinguishing household income according to the each household report in Taipei” [15]. The age is categorized in seven levels: under 30, 30–34, 35–39, 40–44, 45–54, 55–64, over 65. Therefore, these numerical data transform to categorical data.
- Finally, if there are any missing data and value data cleaning is considered to be taken.

As shown in Table 2, the processed dataset contains 159,627 data and 16 different attributes. And the dataset was considered appropriate for our study.

3.2 *Evaluation Methods*

For evaluating the different prediction models, the most widely accepted evaluation methods are Accuracy, Precision and Recall. These evaluation methods are all derived from the confusion matrix shown in Table 3.

The true positive (TP) and true negative (TN) are correct classifications prediction. The false positive (FP) means the output is incorrect classification. For example, the result is positive, but it is actually negative. The false negative (FN) means the output is incorrectly classification. For example, the result is negative but it is actually positive. Equation (1) shows the definition of accuracy. Recall and precision evaluation criterion are based on confusion matrix. The recall is the number of instance retrieved that relevant divided by total number of instances that are relevant [16]. Equation (2) shows the definition of recall. The precision is the number of instances retrieved that are relevant divided by total number of instances that are retrieved [16]. Equation (3) shows the definition of precision.

Table 2 Description of used attributes in this study

Attributes	Attributes type	Attributes definition
CASETYPE	Categorical	1. (Purchase) 2. (Pre-sale) 3. (Leasehold)
CASEFLAG	Categorical	1. (Land and housing) 2. (Land, housing and parking) 3. (Land) 4. (housing) 5. (Parking)
ALLPRICE	Numerical	The prices of transaction
AGE	Categorical	Include (under 30, 30–34, 35–39, 40–44, 45–54, 55–64, over 65)
GENDER	Categorical	1. (Male) 2. (Female)
HC	Date	The housing completion date
HR		1. (Taipei) 2. (New Taipei City) 3. (Other City)
AREAS	Categorical	1. (Songsshan) 2. (Sinyi) 3. (Da-an) 4. (Jhongsan) 5. (Jhongheng) 6. (Datong) 7. (Wanhua) 8. (Wunshan) 9. (Nangang) 10. (Neihu) 11. (Shihlin) 12. (Beitou)
FLOOR	Numerical	The number of transaction floor
BUILD	Categorical	1. (Flats) 2. (Town housing) 3. (Store) 4. (Commercial building) 5. (Apartment building) 6. (Height flats) 7. (Suite) 8. (Factory)
SIZE	Numerical	The number of transaction housing size
BM	Categorical	1. (Built of bricks) 2. (Building occupation permit) 3. (built of reinforced concrete)
ROOM	Numerical	The number of room
PARLOR	Numerical	The number of Parlor

(continued)

Table 2 (continued)

Attributes	Attributes type	Attributes definition
BATHROOM	Numerical	The number of Bathroom
COMPARTMENT	Boolean	Yes/no
MANAGEMENT	Boolean	Yes/no

Table 3 Confusion matrix

		Predicted Class	
		Yes	No
Accrual Class	Yes	TP	FN
	No	FP	TN

$$Accuracy = \frac{TP + TN}{TP + TN + TP + FN} \tag{1}$$

$$Recall = \frac{TP}{P} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

4 Experimental Results

Table 4 shows the results for C4.5 based on three evaluation methods (accuracy, precision, recall). As the results, a prediction model with a high value of accuracy gives better measure than a model with a higher value. The best model shows a value 100 % for accuracy evaluation methods.

4.1 The Classification Rules

These rules can be explored for housing purchase prediction in different stages. The C4.5 model contains more than 7000 rules. Table 5 shows the classification rules for C4.5 model.

Table 4 Evaluation results of model

Evaluation methods	Results
Accuracy (%)	87.5
Precision (%)	89.2
Recall (%)	87.7

Table 5 The classification rules of C4.5 models

Number	Rules
Rule-1	IF HR = New Taipei City AND Area = Neihu AND BUILD = Apartment Building AND SIZE <= 45.62 AND ALLPRICE <= 12,450,000 THEN Under30
Rule-2	IF HR = New Taipei City AND Area = Neihu AND BATHROOM <= 1 AND PARLOR > 1 AND SIZE > 76.48 AND SIZE <= 92.88 THEN 30to34
Rule-3	IF HR = New Taipei City AND Area = Neihu AND SIZE > 69.01 AND ROOM > 2 AND HC <= 74 AND ALLPRICE <= 16,850,000 AND GENDER = male THEN 35to39
Rule-4	IF HR = Taipei AND GENDER = female AND Area = Neihu AND SIZE > 33.69 AND SIZE <= 37.96 AND HC > 72 AND FLOOR > 8 THEN 40to44
Rule-5	IF HR = New Taipei City AND GENDER = male AND Area = Neihu AND SIZE > 97.67 THEN 45to54
Rule-6	IF HR = New Taipei City AND GENDER = female AND AREAS = Neihu AND HC <= 97 AND ALLPRICE > 19,100,000 AND PARLOR <= 1 AND FLOOR <= 11 AND ROOM <= 1 AND ALLPRICE <= 24,100,000 THEN 55to64
Rule-7	IF HR = Taipei AND GENDER = female AND AREAS = Neihu AND HC <= 97 AND ALLPRICE > 19,100,000 AND PARLOR <= 1 AND BATHROOM <= 1 AND ALLPRICE <= 51,500,000 THEN over65

From the point of view of single location transaction, decision of house at different stages of life cycle is different. The above seven classification rules are interpreted as below:

- The Rule-1 should be interpreted as: “if household register of estate proprietor is New Taipei City and housing location of transaction is Neihu and housing type of transaction is apartment building and housing size of transaction is less than or equal to 45.62 m² and housing price of transaction less than or equal to 12,450,000, then age will be under 30”.
- The Rule-2 should be interpreted as: “if household register of estate proprietor is New Taipei City and housing location of transaction is Neihu and housing less than or equal to 1 rooms and housing greater than to 1 parlor and housing sizes between greater than 76.48 and less than or equal to 92.88 m², then age will be 30–34”.
- The Rule-3 should be interpreted as: “if household register of estate proprietor is New Taipei City and housing location of transaction is Neihu and housing sizes greater than 69.01 and housing greater than to 2 rooms and housing completion date less than 74 year and housing price of transaction less than or equal to 16,850,000 and gender of estate proprietor is male, then age will be 35–39”.
- The Rule-4 should be interpreted as: “if household register of estate proprietor is Taipei and gender of estate proprietor is female and housing location of transaction is Neihu and housing sizes between greater than 33.69 and less than or equal to 37.96 m² and housing completion date greater than 72 year and housing floor of transaction greater than 8, then age will be 40–44”.

- The Rule-5 should be interpreted as: “if household register of estate proprietor is New Taipei City and gender of estate proprietor is male and housing location of transaction is Neihu and housing sizes greater than 97.67 m², then age will be 45–54”.
- The Rule-6 should be interpreted as: “if household register of estate proprietor is New Taipei City and gender of estate proprietor is female and housing location of transaction is Neihu and housing completion date less than 97 year and housing price of transaction between greater than 19,100,000 and less than 24,100,000 and housing less than or equal to 1 parlor and housing floor of transaction less than or equal 11 and housing less than or equal to 1 rooms, then age will be 55–64”.
- The Rule-7 should be interpreted as: “if household register of estate proprietor is Taipei and gender of estate proprietor is female and housing location of transaction is Neihu and housing completion date less than or equal 97 year and housing price of transaction between greater than 19,100,000 and less than 51,500,000 and housing less than or equal to 1 parlor and housing less than or equal to 1 rooms, then age will be over 65”.

Among these classification rules, the housing size of transaction is different between 30 and 30–34 age. The people under marriageable age tend to have small housing size of transaction. When getting older, they tend to purchase larger house. In addition, they would consider to have more rooms, parlors, and bathrooms. Interestingly, the people between 55–64 and the people over 65 prefer small houses. The results show that the people over 55 prefer to purchase small house for rental and investment.

Government is able to adopt the proposed C4.5 model to establish policy. However, the C4.5 model contains over than 7000 rules. It is difficult to show the overall rule sets of C4.5 models in this paper because the final results are very large. Hence, these rules can be input into the system for housing purchase prediction in different stages. Government may consider to apply the feasible rules of C4.5 model to establish real estate policy. For example, the government can establish preferential interest rate policy with bank for young people who are under 30 at special case.

5 Conclusion

A focus on rule-based approach to housing purchase life cycle modeling is presented in this paper. This is a special issue for researchers explore real estate transaction data about housing purchase. In each period of life, different people have different choices. Some of the people under 30 need small houses, some of the between 30 and 34 need large house for children education. Government and builders can investigate housing purchase over life cycle through this study.

C4.5 approach is taken for building prediction model from real estate transaction dataset. Three evaluation methods to measure prediction model are presented for

that purpose: accuracy, precision and recall. Real estate transaction datasets from five real-world system are used for demonstrating the modeling approach. However, in the case of housing purchase and as demonstrated by this study, it provides suggestions for the decision makers of the government to use prediction models to make appropriate decisions following different policies.

References

1. Wu F (2010) Housing environment preference of young consumers in Guangzhou China. *Property Manag* 28:174–192
2. Buckenberger C (2012) Meanings of housing qualities in suburbia: empirical evidence from Auckland, New Zealand. *J Housing Built Environ*, 27:69–88
3. Malmberg B (2010) Low fertility and the housing market: evidence from Swedish regional data. *Eur J Population*, 26:229–244
4. Hurst N (2012) Energy efficiency rating systems for housing: an Australian perspective. *Int J Hous Markets Anal*, 361–376
5. Luo Q, James PTJ (2013) Influences on the buying behavior of purchasing commercial housing in Nanning city of Guangxi province, China. *J Manag Mark Res*, 150
6. Han JW, Kamber M (2007) *Data mining: concepts and technique*, 2nd edn. Morgan Kaufmann
7. Li LH (2011) Impact of housing design factors on children's conduct at school: an empirical study of Hong Kong. *J Hous Built Environ*, 26:427–439
8. Li L (2011) Housing choice in an affluent Shanghai—decision process of middle class Shanghai Residents. *Mod Economy*, 2:427–439
9. Ball M, Nanda A (2013) Household attributes and the future demand for retirement housing. *Int J Hous Mark Anal*, 6:45–62
10. Johnson MP (2005) Spatial decision support for assisted housing mobility counseling. *Decis Support Syst*, 41:296–312
11. Juan YK, Shih SG, Perng YH (2006) Decision support for housing customization: a hybrid approach using case-based reasoning and genetic algorithm. *Expert Syst Appl*, 31:83–93
12. Fischa D, Kühbecka B, Sicka B, Ovaskab SJ (2010) So near and yet so far: new insight into properties of some well-known classifier paradigms. *Inf Sci*, 180:3381–3401
13. Quinlan JR (1993) *C4.5: programs for machine learning*. Springer, Berlin
14. Attanasio OP, Bottazzic R, Lowd HW, Nesheima L, Wakefieldc M (2012) Modeling the demand for housing over the life cycle. *Rev Econ Dyn*, 15:1–18
15. Department of Budget Accounting & Statistics, Taipei City Government, distinguishing household income according to the each household report in Taipei (2012)
16. Witten IH, Frank E, Hall MA (2008) *Data mining: practical machine learning tools and techniques*, 3rd edn. Morgan Kaufmann

Algorithms of AP+ Tree Operations for IoT System

Qianjin Tang, Zhizong Wu, Yixuan Wu and Jinfeng Ma

Abstract Internet of things (IoT) system have a large number of sensors, and each sensor will generate a large amount of real-time streaming data. So real-time database technology for IoT network is very important to achieve real-time data stream generated by data aggregation, query, analysis and data mining. The biggest problem of the real time data management is data overload and how to efficiently find the data, temporal data index is a good solution. The paper investigated AP+ tree design idea, operation algorithms including query, insertion, deletion and reconstruction. AP+ tree index structure is adopted for the real-time database in order to improve the efficiency of temporal queries. The result shows that the search efficiency of AP+ tree is 1.2 time of B+ tree under certain condition.

Keywords Temporal database · Massive sensors · AP+ tree · Iot

1 Introduction

The core technologies of Internet of things (IoT) are managing and processing data from sensors, including data storage, query, analysis, mining, understanding and perception data based decision-making and behavior theory and technology [1]. Sensor data stream is in the form of data transmission, how to store and manage the data flow of data has become a hot issue for IoT research [2–4]. Facts have proved

Q. Tang · Z. Wu (✉) · Y. Wu

The Third Research Institute of Ministry of Public Security, Shanghai, China
e-mail: Tangqj2008@163.com

Y. Wu

e-mail: Wyx876@16t3.com

J. Ma

Shanghai University, Shanghai, China
e-mail: Jinfengma886@shu.edu.cn

that [5], not all the data stream data can be stored in the database. From a theoretical view, a steady stream of sensor data can be treated as an infinite amount of data, so it is completely impossible that you want to store the entire data stream. For practical applications, the data streams per second may update to GB level. The traditional database solutions are also difficult to meet the IoT system requirements. In addition, from the perspective of practical application, it is not necessary to use all data streams. For applications of predictive traffic, network control and other aspects, they need not store all the stream data, but require detailed records of all data at a certain time period. However, even for these requirements, the amount of data stored is very great. So the biggest problem of the temporal data management is data overload, how to efficiently find the data, temporal data index is a good time to introduce.

Compared with traditional relational database, the index technique of real-time database to quickly find the conditions which is consistent with users from the huge data is different. Query of temporal data is related to time feature, and the traditional index structure and technology can not meet its requirements. In practical applications, the system stores large amount of data, and during a certain continuous period, the data increases incessantly. So the design of index structure needs to ensure that computation and memory cost of the various operations is as small as possible. In addition to large amounts of data, temporal data index also needs to solve a series of problems, including optimization of the time dimension of the data query and too many inequality predications [5]. At the same time, because the time has increment and one-way flow characteristics, and the index structure can be optimized using this feature.

In this paper, we investigated design idea of AP+ tree for real-time database. A detailed operation of queries AP+ tree were described, including query, insertion, deletion and reconstruction. Last, we made performance analysis.

2 AP+ Tree Structure Design

The design idea of AP tree is derived from traditional database technology, which is stored in permanent storage medium, and is the same to B+ and B+ tree-like. When we need query, all nodes which meet the conditions only read from permanent storage media. The layer number of AP and B+ tree is fewer, and they can accommodate large amount of information, so the number of nodes which is read into memory is not too large and performance is acceptable. But this is only limited to data storage appended in database. The real-time database uses a memory database technology and the entire AP trees cannot be placed in memory for space limitation. If AP tree dose not improve appropriately, the system wastes memory space largely and the performance will be affected. According to ways and means of storage of AP tree, we found that each node of the tree has a degree d , which means that the maximum number of AP tree node can store at a time point.

AP+ tree is designed in accordance with the above ideas. It is still a AP tree which meets the attribute of AP tree essentially, while it just changes the storage structure, which is lead to the changes of basic operation. The method “convert” makes the AP tree improvement to AP+ tree with less memory space. The core idea of method “convert” is that all leaf nodes at the first time point is spare, which are replaced by the pointer of non-leaf node except for the first leaf node. For the non-leaf node, the value of pointer is the index of temporal data set. The advantage of this method is saving storage space and improving query efficiency, while the disadvantage is consuming considerable computing resource when AP tree and AP+ tree transformation takes. So we propose that firstly system generates AP tree, and secondly the AP tree takes “convert” operations to generate a smaller space AP+ tree.

3 AP+ Tree Operations Algorithms

3.1 Query Operations

Two query search path algorithms for AP+ tree are described as follow.

1. Query algorithms at T_s (effective start time)

- Algorithms: APP_SEACHTS (T)
- Input: Query time T_s
- Output: If the query succeeds, it returns temporal data sets; else it returns failure code.
- Steps:
 - (1) If $V < T_s$, or $V > Lsr.END$, the search fails and the inquiry ends; otherwise, if $V = T_s$, it returns the result and the inquiry ends.
 - (2) If $V = Lsr.END$, it finds the rightmost leaf node through metadata information of the root node; if the rightmost leaf node is not found in V , the search fails and the inquiry ends; if it finds, it returns structure and the query ends.
 - (3) It searches from the root node and compares each $v+$ value until the leaf node. If V is not included in leaf node, the search fails and the inquiry ends; otherwise, it returns a result and the inquiry ends.

Where V represents query time, $Lsr.END$ represents the end point of a relation life cycle, T_s represents the minimum index value T_s of AP and AP+ tree.

2. Query algorithms at any point in time, and T_e (effective end time)

- Algorithms: APP_SEACHT (T)
- Input: Query time point T
- Output: If the query succeeds, it returns the temporal data sets of information; if query fails, it returns failure code.

- Steps:
 - (1) If $V < T_s$ or $V > Lsr.END$, the search fails, the inquiry ends; otherwise, if $V = Lsr.END$, it turn to step 3; otherwise it turns to step 2 and step 3.
 - (2) It searches from the root node and compares each $v +$ value until the leaf node. In the set of tuples $v +$, it performs a reverse lookup for each tuple. Find tuple until it satisfies that V belongs to $[x(T_s), x(T_e)]$ interval, it returns structure and ends the inquiry.
 - (3) It searches from the set of tuples each tuple until $T_s > T$ at the end. Find tuples x , which satisfies V belong to $[x(T_s), x(T_e)]$ interval, it returns structure and ends the inquiry.

3.2 Insertion Leaf Node Operations

The algorithms of AP+ tree insertion operation of one leaf node are shown as follow.

- Algorithm: A PP_INSERT_LEAF (*key*)
- Input: *key* value of insertion
- Output: If it inserts successfully, it returns success code, else it returns a failure code.
- Steps:
 - (1) If *RootPtr* = NULL, it creates a root node, and then inserts *NewKey* value, insertion operation ends; otherwise, it turns to step 2.
 - (2) It finds the rightmost leaf node based on metadata information, if *NewKey* value in the rightmost leaf node is found, the value of the pointer is updated and points to the newly inserted tuple, and the insertion operation ends; otherwise, it turns to step 3.
 - (3) If the rightmost leaf node is not full, the *NewKey* is inserted to the rightmost leaf node, insertion operation ends; otherwise, it turns to step 4.
 - (4) If the right-most leaf node is full, it establishes a new child leaf node *NewLeaf*. *NewKey* inserts this node and the parent node of *NewLeaf* is set to the parent node of the rightmost child leaf node.
 - (5) *rsib* (*RightMostLeaf*) = *NewLeaf*.
 - (6) *2rsib* (*NewLeaf*) = NULL.
 - (7) *lsib* (*NewLeaf*) = *RightMostLeaf*.

Where “*lowPtr*” represents the leftmost pointer of a node, value of the pointer is less than that of the minimum node; “*lsib*” and “*rsib*” represent the left and right child leaf nodes of the leaf nodes representatively, the value of two non-leaf nodes is set to NULL; “*RootPtr*” represents the pointer of the root node.

3.3 Deletion Operations

AP+ tree deletion operations have two cases. One is to delete AP+ tree index at time point T_s ; another is given at any time point.

1. Deletion time T_s

- Algorithms: APP_DELTS (T)
- Input: The deletion time T_s
- Output: If it deletes successfully, it returns success code; otherwise, it returns failure code.
- Steps:
 - (1) Get the leftmost leaf node pointer from the metadata information, assign *CurrentNode* value.
 - (2) In *CurrentNode* point nodes, it deletes all the eligibility key assignments which are less than *Cutoff*. If pointer of *CurrentNode* node is empty, the pointer of the right leaf brother node assigns *CurrentNode*. Then previous *CurrentNode* which was referred already empty nodes is deleted; otherwise, the keys of *CurrentNode* node are left-aligned.
 - (3) The value of T_s is assigned *Cutoff*.
 - (4) If the node *CurrentNode* has no brother leaf node, *CurrentNode* node is the root node, and the AP tree has only this one node; otherwise, Algorithm 3.5 is recalled to reconstruct AP tree.

2. Deletion operation at any time point

- Algorithms: APP_DELT (T)
- Input: The time T of deletion
- Output: If it deletes successfully and then returns success code; otherwise it returns failure code.
- Steps:
 - (1) Get the pointer of the leftmost leaf node from the metadata, it is assigned *CurrentNode*.
 - (2) From the leftmost leaf node, the query ends until the key is larger than *Cutoff*. All the tuples are deleted if the effective end time T_e of the valid key is less than *Cutoff*. If a node is empty, the node is removed.
 - (3) The non-empty leaf node is left-aligned; the minimum key of the leftmost leaf node is T_s .
 - (4) If the leftmost leaf node has no brother leaf node, the left-most leaf node is the root node, and the AP tree has only this node; otherwise, algorithm 3.5 is called to reconstruct AP tree.

3.4 Reconstruction Operations

- Algorithms: APP_REBUILDER (*AP_TREE*)
- Input: *AP_TREE* handle.
- Output: If it reconstructs successfully and returns success code; otherwise it returns failure code.
- Steps:
 - (1) A new node is created for d child nodes, and the appropriate key is inserted. If the rightmost node has only one child node, a new node need is created.
 - (2) If only a single node is created, it is set to root node; otherwise algorithm 3.5 is called recursively.

4 Performance Analysis

Only it analyses complexity of time and space of AP+ tree is not intuitive, so we compare B+ tree [5] index with AP+ tree index by the temporal data. The traditional B+ tree structure is transformed to index B+ tree. First, a bi-pointer is added among the leaf nodes. Secondly, the head node B+ tree stores the minimum effective start time T_s and the maximum effective end time T_e .

We assume that all the AP+ tree are loaded fully in addition to the right child tree, the value of each B+ tree node must meets the minimum requirements of 50 % load.

Under this assumption, AP+ and B+ tree heights are following respectively:

$$h_{AP}(T) = \lfloor \log_{d+1}(N_v + 1) \rfloor + 1 \quad (1)$$

$$h_B(T) = \lfloor \log_{\lceil d/2 \rceil + 1}(N_v + 1) \rfloor + 1 \quad (2)$$

We can get Eq. (3) by

$$h_B(T)/h_{AP}(T) = \frac{1}{1 - \log_d 2} \quad (3)$$

When degree d (each node of the tree) = 100, the height of B+ tree is about 1.2 times of the AP+ tree. With the increasing value of d , the height of B+ trees and AP tree will become increasingly close. The height of the tree determines the search performance. When the value of d is small, the search efficiency of B+ tree is not as good as AP+ tree.

5 Conclusions

Temporal data index is a good solution for IoT system. AP+ tree index structure is one of good methods for IoT database in order to improve the efficiency of temporal queries. The result shows that when the value of d is small, the performance of AP+ tree is better than that of B+ tree.

Acknowledgments This work was supported in part by the National Science and Technology Major Project under Grant 2011ZX03005-006, in part by the Program of Science and Technology Commission of Shanghai Municipality under Grant 14DZ2252900.

References

1. Tilak S, Abu-Ghazaleh NB, Heinzelman W (2002) A taxonomy of wireless micro-sensor network models. *Mob Comput Commun Rev* 1(2):1–8
2. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Top Comput* 2(3):376–387. doi:[10.1109/TETC.2014.2316525](https://doi.org/10.1109/TETC.2014.2316525)
3. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Generation Comput Syst* 43–44:40–50
4. Krithi R (1993) Real-time database. distributed and parallel. *Database* 21(1):199–266
5. Ming X, Krithi R, Jayan H, John S (1999) MIRROR: A state-conscious concurrency control protocol for replicated real-time databases. Computer science department, Faculty publication series

Dynamic Storage Method of Big Data Based on Layered and Configurable Technology

Wenjuan Liu, Shunxiang Zhang and Zheng Xu

Abstract In big data era, how to reasonably divide various types of data and efficiently store so large and complex data sets is an important challenge. This paper proposes a layered and configurable storage model to improve the storage capability of big data. First, three-layer hybrid row-column-store storage model is presented, which contains metadata definition layer, key-value model layer, and data physical storage layer. This model combines the advantages of row-store and column-store. Second, based on the three-layer storage model, the realization process is presented. According to the characteristics of each column of data, the suitable storage mode of row-store or column-store is chosen. The proposed model can provide a new technology support for the storage of big data.

Keywords Big data · Dynamic storage · Layered and configurable storage model · Hybrid row-column-store

1 Introduction

With the progress of society and the rapid development of information technology industry, the growth of data amount is explosive. According to the IDC report, the global data amount reach 1.8 ZB only in 2011, this trend is still accelerating. It is

W. Liu (✉) · S. Zhang

School of Computer Science and Engineering, Anhui University
of Science and Technology, Huainan 232001, China
e-mail: liuwj@aust.edu.cn

S. Zhang

e-mail: sxzhang@aust.edu.cn

Z. Xu

The Third Research Institute, The Ministry of Public Security,
Shanghai 200031, China
e-mail: xuzheng@shu.edu.cn

estimated that data growth rate will remain around 50 % per year in the future 10 years. So it is urgent to deal with big data storage [1].

Two kinds of data storage mode are mostly used in current data storage, row-store and column-store [2]. In row-store mode, data is directly stored based on the tuple [3]. The efficiency of write data is high, the integrity and reliability of the data is high. It is suitable for OLTP occasion. The defect is that there may be redundant columns at the time of read data. A record may include multiple data types, data compression is relatively lower [4]. In column-store mode, each column in the table is organized together to store, and different columns are independently stored [5]. At the time of read data, only need to read the attribute columns, the cost of read data can be greatly reduced. Data type in a column is same, so data compression ratio is high, and the treatment effect is more obvious for sparse data. It is suitable for OLAP, data warehouse, query intensive applications, etc. But the efficiency of write data is low. The integrity and reliability of data is less than row-store. It is not suitable for using in OLTP or update operation, especially in the frequent insert and delete occasions [6].

With the introduction of the Internet of Things, Internet and other new business data, the limitation of the two kinds of storage scheme have gradually appeared. These data not only include dense data, but also include massive sparse data. And the two kinds of data in the actual application often coexist, regardless of row-store or column-store can effectively solve the efficient storage problem of mixed data sets [7].

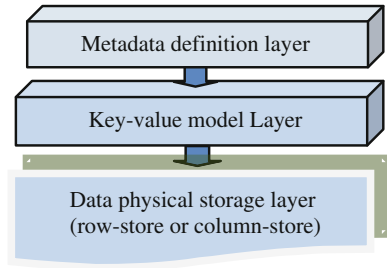
To solve the problem mentioned above, a layered and configurable storage scheme has been proposed which is based on hybrid row-column-store layout. The major contribution of our work is that we present a three-layer dynamic storage model and show how to realize data storage by an example. This layered and configurable model contains metadata definition layer, key-value model layer, and data physical storage layer. It combines the advantages of row-store and column-store. The dense data is stored as row-store to ensure high efficiency write capability. The sparse data is stored as column-store to improve the read capability and decrease the storage space.

The rest of this paper is organized as follows. Section 2 presents a layered and configurable storage model which is a three-layer hybrid row-column-store storage model. Section 3 uses an instance to show how to realize each layer. In the end, Sect. 4 gives the conclusions about this paper.

2 Layered and Configurable Model

In this section, three-layer storage architecture is proposed to meet the storage requirements of sparse data and dense data sets in data processing. Figure 1 shows the layer dividing of data storage architecture.

Fig. 1 Three-layer storage model



In Fig. 1, the metadata definition layer overall defines each attribute column storage strategy and data constraints, etc. The key-value model layer is a structured key-value model based on key-value pairs. Various types of mixed data sets reflect for a collection of key-value pairs, and each key-value pair corresponds to the attribute value of an entity. Through the key-value model, each attribute column is organized to form different subsets of the data. The third layer is the data physical storage layer. According to the definition of metadata and key-value model, physical storage format used row-store or column-store is defined for each subset of data.

2.1 Metadata Definition Layer

The traditional key-value model has the characteristics of simple model, fast search, large amount of data storage and high concurrency. But its key-value pairs only contain attribute values, lacking pattern definition and structured data. The metadata definition layer of the scheme integrates the pattern definition based on the traditional key-value pairs. From a visual view, the key-value pairs collection are defined as the elastic container of data storage, supporting model definition and extension. The information of metadata definition is listed in Table 1.

A table is composed of one or more column groups. Column group is a collection of columns, and different column groups do not overlap each other. Two kinds of column groups are given in the scheme: CG_R and CG_C. The data in column group CG_R are stored by row-store, and the data in column group CG_C are stored by column-store. Except the keyword column, other columns should belong to a particular column group. Usually different column groups in a table can share key range file. According to the characteristics of the column or column

Table 1 The information of metadata definition

ColumnGroup	CG_R			CG_C		
Column	Column1	Column2	Column3	Column4	Column5	Column6
KeyFile	KeyFile	KeyFile	KeyFile	KeyFile	KeyFile	KeyFile
DataFile	DataFile1	DataFile1	DataFile1	DataFile2	DataFile3	DataFile4

groups, their value range data can be stored in different data files. For example, one data file can be used to store column group CG_R , and corresponding to each attribute column in column group CG_C , one or more data files can be used.

2.2 Structured Key-Value Model Layer

In a structured key-value model definition, each key-value pair corresponds to the attribute value of an entity. An entity can contain multiple attributes, and a tuple is mapped to a collection of key-value pairs. Key range is a multidimensional mapping composed of multiple sub-keys. Where sub-key RK is the unique identifier of the tuple, sub-key CG is the attribute column array, including one or more attribute columns, and the address of attribute column array is saved in sub-key CA , which points to the data content in value range files. Key range is stored in a single physical file, equivalent to the data index file. The data content of key range are stored in value range. Key-value pair includes a linked list composed of a continuous multiple values, and each copy of the data has a timestamp to identify the different data versions. The value range are saved in multiple independent physical files, equivalent to the data file content. The key-value model is shown in Fig. 2.

Key range model in Fig. 2. Is saved in key range file, and value range model in Fig. 2 are saved in value range files. For a given column defined by sub-key CG in key-value model, sub-key CA points to the data unit of this column in the value range file. The first three sub-keys of key range constitute the query primary key, which can only determine a data unit. Considering the high concurrency of data processing, data unit must be effectively avoided read/write conflict. Transactional consistency control may be assisted with sub-key LT . At the same time, each data unit holds multiple versions of a data object. It is indexed by timestamp between versions, and arranged in inversed order according to the timestamp.

2.3 Data Physical Storage Layer

For the physical storage architecture of the entity key-value pair, the idea of horizontal fragmentation can be adopted to fit the data update. Each section is divided

Fig. 2 Structured key-value model

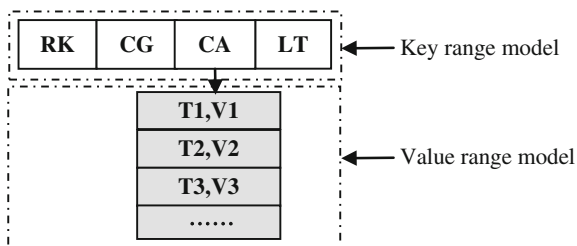


Table 2 Entity table *TableTest*

RowKey	Col1 (Number)	Col2 (Varchar)	Col3 (Number)	Col4 (Date)	Col5 (Varchar)
Row1	1	John	100	NULL	SH
Row2	2	White	NULL	2014/9/1	NULL
Row3	3	Li	NULL	NULL	NULL
Row4	4	Kitty	200	NULL	NULL

vertically into multiple partitions. For a given partition, row-store or column-store can be chosen. For dense data, row-store is used, and the whole partition is mapped to a single file, thus forming a row group, which may include one or more dense data columns. For sparse data, column-store is used, and according to the data characteristics of each attribute column, one column or several columns in the partition are stored in a physical file.

3 The Instances of Dynamic Storage Scheme

The entity table *TableTest* is used as an example, whose data is listed in Table 2. Due to the characteristics of sparse and dense data are not identical, the hybrid row-column-store method is chosen. In the table, column *Col1* and *Col2* are dense data columns, so row-store can be used, and they belong to column group *CG_R*. Column *Col3*, *Col4* and *Col5* are sparse data columns which have many null values, so column-store are used, and they belong to column group *CG_C*.

3.1 The Representation of Metadata Information

In Table 3, key range data is saved to the file “KeyFile”. Column *Col1* and *Col2* in value range data used row-store are saved to the file “DataFile1”. To enhance the query performance, each column is saved independently, so *Col3*, *Col4* and *Col5* are saved respectively to the file “DataFile2”, “DataFile3”, “DataFile4” by column-store.

Table 3 Metadata information definition

Table	TableTest				
ColumnGroup	CG_R		CG_C		
Column	Col1	Col2	Col3	Col4	Col5
KeyFile	KeyFile	KeyFile	KeyFile	KeyFile	KeyFile
DataFile	DataFile1	DataFile1	DataFile2	DataFile3	DataFile4
DataType	Number	Varchar	Number	Date	Varchar

Table 4 Key-value model definition

RK	CG	CA	LT
Row1	col1,col2,col3,col4,col5	f1_row1_col1,f1_row1_col2,f2_row1_col1, f3_row1_col1,f4_row1_col1	PID1
Row2	col1,col2,col3,col4,col5	f1_row2_col2,f1_row2_col2,f2_row2_col1, f3_row2_col1,f4_row2_col1	PID2
Row3	col1,col2,col3,col4,col5	f1_row3_col1,f1_row3_col2,f2_row3_col1, f3_row3_col1,f4_row3_col1	PID3
Row4	col1,col2,col3,col4,col5	f1_row4_col1,f1_row4_col2,f2_row4_col1, f3_row4_col1,f4_row3_col1	PID4

3.2 The Definition of Key-Value Model

In Table 4, the values of column *CA* are equivalent to the file index of data content in column *Col1* to *Col5*. Where *fn* corresponds to the specific data file. For example, *f1* corresponds to the file “DataFile1”, *f2* corresponds the file “DataFile2”, etc. *rown* corresponds the record at row *n*, and *coln* corresponds column *n* of the record, for example, *f1_row1_col1* corresponds the first column of first row in the file “DataFile1”. Comma is used as a separator between the columns.

3.3 The Implementation of Physical Storage

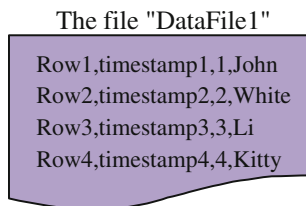
(1) Row-store

Figure 3 is a logical diagram of the file “DataFile1” which is stored as row-store. Keyword column, two dense columns and timestamp of each record are all saved in this file, to achieve better write data performance.

(2) Column-store

Figure 4 is a logical diagram which contains three value range files such as “DataFile2”, “DataFile3” and “DataFile4”. All these three files are stored as column-store. Keyword column, timestamp and one of sparse data column are saved in a file, to achieve better read data performance and higher compression ratio.

Fig. 3 Logic diagram of row-store file



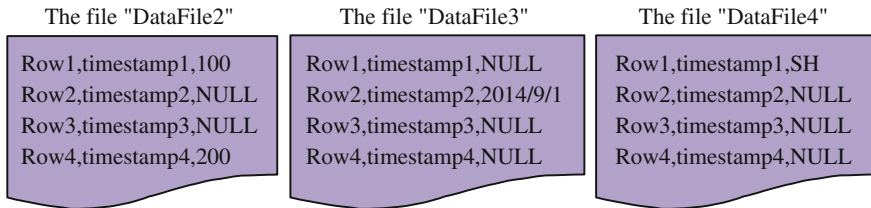


Fig. 4 Logic diagram of column-store files

4 Conclusions

A layered and configurable storage model has been proposed to improve the storage capability of big data. According to the data characteristics of different columns, the suitable storage mode are chosen to realize the dynamic data storage. When storing data, the data sets are mapped to a structured key-value pairs collection, and a hybrid row-column-store layout is provided. This model can simultaneously meet the storage requirements of sparse data and dense data sets in data processing, and can effectively solve the efficient storage problem of mixed data sets.

Acknowledgements This Research work was supported in part by the Youth Scientific Research Foundation of Anhui University of Science and Technology (Grant No. QN201321), and by the Natural Science Foundation of Anhui Province (Grant No. 1308085MF94).

References

1. Lai H, Xu M, Xu J, Ren Y, Zheng N (2013) Evaluating data storage structures of MapReduce. *Comput Sci Educ (ICCSE)*, pp 1041–1045
2. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
3. Halverson AJ, Beekmann JL, Naughton JF, Dewitt DJ (2006) A comparison of c-store and row-store in a common framework. Technical Report. University of Wisconsin-Madison, Department of Computer Sciences
4. Kanade AS, Gopal A (2013) Choosing right database system: row or column-store. In: *International conference on information communication and embedded systems (ICICES)*, 2013, pp 16–20
5. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Top Comput* 2(3):376–387
6. McKnight W (2014) *Information management: strategies for gaining a competitive advantage with data*. Morgan Kaufmann Publishers, pp 97–109
7. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50

MIC-Based Preconditioned Conjugate Gradient Method for Solving Large Sparse Linear Equations

Zhiwei Tang, Hailang Huang, Hong Jiang and Bin Li

Abstract High Performance computing (HPC) are becoming not only more complex but also challenging in terms of speedup and scalability. As the size of compute intensive problems increases, Intel MIC architecture comes true. In this paper, PCG method based on Intel MIC architecture is employed to solve large scale linear equations. The numerical results show that PCG method based on Intel MIC architecture has a considerable speedup and scalability.

Keywords Large sparse linear equations · PCG method · Intel MIC · Openmp

1 Introduction

Since announced in May of 2010, Intel Many Integrated Core (MIC) architecture drew more and more attention targeting to High Performance Computing field for the Peta Flops era [1]. The Intel MIC architecture is a multicore x86 architecture that combines many Intel architecture CPU cores on a single chip based on the streamlined x86 core and similar to the existing CPUs in which existing

Intel MIC architecture and OpenMP. PCG method is discussed in Sect. 3. In Sect. 4, storage scheme and detailed implementation of PCG method based on Intel MIC architecture is proposed. After showing the numerical results, we draw the conclusions.

Z. Tang · H. Huang (✉) · H. Jiang
The Third Research Institute of Ministry of Public Security, 339 Bisheng Road,
Shanghai, China
e-mail: huanghailangtao@126.com

B. Li
Shanghai University, 99 Shangda Road, Shanghai, China
e-mail: 99chaoyang@163.com

parallelization tools (MPI [2], OpenMP [3], etc.) can be used, and specialized versions of Intel's Fortran, C++ and math libraries [4]. Its SIMD instructions are further extends to very wide 512-bit and allow 512-bit numbers to be manipulated on a core simultaneously. MIC's 60 cores also greatly improve its parallel computing capabilities.

Krylov subspace solvers are widely employed for solving large sparse linear systems which are often derived from the discretization of partial differential equations (PDE) by finite element method, finite difference method and finite volume method, etc. The Preconditioned Conjugate Gradient (PCG) method is a well-known method iterative solver for large sparse linear systems that has been proven to be efficient and robust in a wide range of applications because their coefficient matrices are often symmetric, positive definite (SPD) [5, 6].

Our goal is to solve large sparse linear equations by applying PCG method on Intel MIC architecture. The paper is organized as follows. The next section introduces.

2 Intel MIC Architecture

The Intel MIC architecture combines many Intel CPU cores onto a single chip. Its key attribute of the microarchitecture is that it is built to provide a general-purpose programming environment similar to the Intel Xeon processor programming environment. Based on the Intel MIC architecture, the Intel Xeon Phi coprocessors are capable of running a full service Linux operating system, supporting x86 memory order model and IEEE 754 floating-point arithmetic, and are able to run applications written in industry-standard programming languages such as Fortran, C, and C++. The coprocessor is also supported by a rich development environment including compilers, numerous libraries, performance characterizing and tuning tools, and debuggers. The Intel Xeon Phi coprocessor is connected to an Intel Xeon processor (known as "host") through a PCI Express bus. The coprocessor supports heterogeneous applications wherein a part of the application executes on the host while a part executes on the coprocessor.

The Intel Xeon Phi coprocessor comprises processing cores, caches, memory controllers, PCIe client logic, and a very high bandwidth, bidirectional ring interconnect (Fig. 1) [7].

Most of the parallel programming options available on the host systems are available for the Intel[®] Xeon Phi[™] Coprocessor including Intel TBB, OpenMP, Intel Cilk Plus and pthread. In this paper, OpenMP is employed. There is no correspondence between OpenMP threads on the host CPU and on the Intel[®] Xeon Phi[™] Coprocessor. Because an OpenMP parallel region within an offload/pragma is offloaded as a unit, the offload compiler creates a team of threads based on the available resources on Intel[®] Xeon Phi[™] Coprocessor.

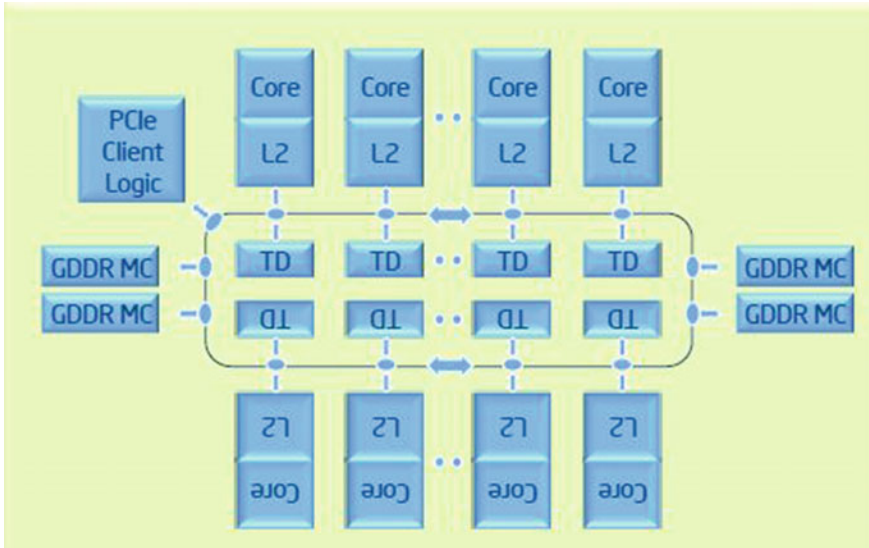


Fig. 1 Microarchitecture

3 PCG Method

Consider

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^n$, and A is a SPD matrix. On the basis of PCG method, Eq. (1) can be written as

$$M^{-1}Ax = M^{-1}b \tag{2}$$

where matrix M is a preconditioner.

Matrix transformations are employed by the preconditioner so that the new sparse linear equations are equal to the old sparse linear equations and the condition number of coefficient matrix decrease which improves the convergence [8, 9].

Given the inputs of A , b , a preconditioner M , a starting vector x_0 , a maximum number of iterations $iter_max$ and a error tolerance $error$, the PCG algorithm can be described in Algorithm 1. A set of α orthogonal search directions $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n$ are constructed by the conjugation of the residues $r_0, r_1, r_2, \dots, r_n$ respectively. x_k takes exactly one step of h_k along the direction α_k in the k th iteration step. When the convergence conditions $error < \epsilon$ and $k < iter_max$ are met, the iterative process is completed.

IC preconditioner is employed in our method to improve the convergence rate. An IC preconditioner can be obtained by factoring a matrix A into the form LL^T

where L is a lower Cholesky triangular matrix. As a consequence, M equals to LL^T and $d_k = M^{-1}r_k$ can be written as $(LL^T)d_k = r_k$. Therefore, d_k can be directly computed by forward and then backward substitutions.

Algorithm 1: PCG method

- 1: initialize x_0 ,
- 2: $r_0 = b - Ax_0$,
- 3: $d_0 = (LL^T)r_0$,
- 4: $h_0 = d_0$,
- 5: $\text{error}_{\text{old}} = \langle r_0, d_0 \rangle$,
- 6: $\text{error}_{\text{new}} = \text{error}_{\text{old}}$
- 7: while ($\text{error}_{\text{new}} < \varepsilon ||k < \text{iter_max}$)
- 8: $\alpha_k = \text{err}_{\text{new}} / \langle h_{k-1}, Ah_{k-1} \rangle$,
- 9: $x_k = x_{k-1} + \alpha_{k-1}h_{k-1}$,
- 10: $r_k = r_{k-1} - \alpha_{k-1}Ah_{k-1}$,
- 11: $d_k = (LL^T)r_k$
- 12: $\text{err}_{\text{old}} = \text{err}_{\text{new}}$,
- 13: $\text{err}_{\text{new}} = \langle r_k, d_k \rangle$,
- 14: $\beta = \text{err}_{\text{new}} / \text{err}_{\text{old}}$,
- 15: $h_k = d_k + \beta h_{k-1}$

4 PCG Method Based on Intel MIC

4.1 Storage Scheme

In order to take advantage of the large number of zero elements, special schemes are required to store sparse matrices. Compressed sparse row (CSR) format is adopted in this paper. CSR format has three arrays with the following functions [10].

A real array *val* contains the non-zero values stored row by row. The length of *val* is *nz*.

An integer array *colIdx* contains the column indices of the elements a_{ij} stored in the array *val*. The length of array *colIdx* is also *nz*.

An integer array *rowIdx* contains the pointers to the beginning of each row in the arrays *val* and *colIdx*.

Thus, the matrix

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 4 & 0 \\ 0 & 0 & 5 \\ 0 & 6 & 0 \end{pmatrix}$$

will be represented by $val[1\ 2\ 3\ 4\ 5\ 6]$, $colIdx[1\ 3\ 1\ 2\ 3\ 2]$ and $rowIdx[1\ 3\ 5\ 6]$, then we can obtain A_{ij} according to i in $rowIdx$ and j in $colIdx$.

4.2 PCG Method Based on Intel MIC

Intel MIC can be used in two ways: a platform for native execution, and an offload platform to execute a part of computation. In this paper, the latter one is adopted which is similar with GPGPU. Intel's compiler supports to offload arbitrary sections of code to the Intel MIC from a program running on the host by means of pragmas. The offload pragmas specify the data to be transferred between the device and the host. PCG method based on Intel MIC was described in Algorithm 2 in which no copy clause indicate that there is no data copy that can save copy time.

Algorithm 2: PCG method based on Intel MIC

```

1: initialize  $x_0$ ,
2: #pragma offload target(mic) nocopy(A, L, b)out(errorold)
3: #pragma omp parallel for
4: {
5:    $r_0 = b - Ax_0$ ,
6:    $h_0 = d_0 = (LL^T)r_0$ ,
7:    $error_{old} = \langle r_0, d_0 \rangle$ ,
8: }
9:  $error_{new} = error_{old}$ 
10: while ( $error_{new} < \epsilon ||k < iter\_max$ )
11:  $err_{old} = err_{new}$ ,
12: #pragma offload target(mic) out(x)
13: #pragma omp parallel for
14: {
15:    $\alpha_k = err_{new} / \langle h_{k-1}, Ah_{k-1} \rangle$ ,
16:    $x_k = x_{k-1} + \alpha_{k-1}h_{k-1}$ 
17:    $r_k = r_{k-1} - \alpha_{k-1}Ah_{k-1}$ ,
18:    $d_k = (LL^T)r_k$ 
19:    $\beta = \langle r_k, d_k \rangle / err_{old}$ ,
20:    $h_k = d_k + \beta h_{k-1}$ 
21: }
```

5 Numerical Results and Experimental Environment

The MIC device used in the experiments is Intel Xeon Phi coprocessor 5110p, 60 cores, which has the frequency of 1.053 GHz, and an 8 GB GDDR5 memory. The corresponding host is Intel Xeon CPU E5-2620*2, 2.4 GHz, 96 GB memory.

In order to assess the performance of our algorithm on Intel MIC, some matrices (Table 1) are taken from the University of Florida’s Sparse Matrix Collection (UFSPARSE) [11].

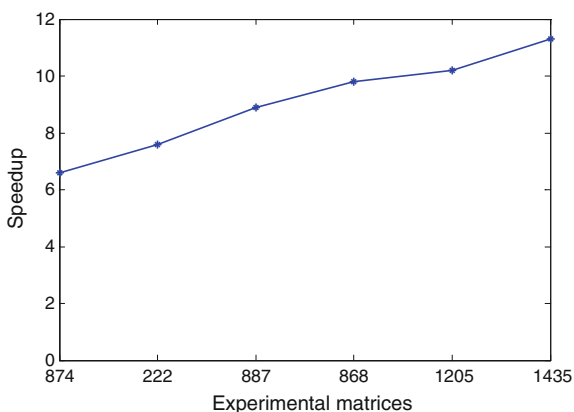
5.1 Performance Evaluation

As is depicted in Fig. 2, PCG method based on Intel MIC architecture has a considerable speedup. With the increase of non-zero elements, it has better speedup performance and less execution time.

Table 1 Experimental matrices

No.	ID	Group	Name	Rows	Non-zeros
1	874	Pothen	mesh1em6	48	306
2	222	HB	nos6	675	3255
3	887	Norris	fv1	9604	85,264
4	868	Pothen	bodyy4	17,546	1,21,938
5	1205	Oberwolfach	t2dah_e	11,445	1,76,117
6	1435	Oberwolfach	gyro	17,361	10,21,159

Fig. 2 Speedup of PCG method based on Intel MIC



6 Summary

In order to store large-scale sparse linear systems in a limited memory space, CSR format was adopted to significantly reduce the storage space, and PCG method based on Intel MIC architecture was presented and demonstrated to have a considerable speedup feature. Also, with the number of nonzero elements increased, the PCG method based on Intel MIC architecture obtained better speedup performance and less execution time.

References

1. Intel Corporation. <http://www.intel.com/content/www/us/en/architecture-and-technology/many-integrated-core/intel-many-integrated-core-architecture.html>
2. MPI. <http://www.mcs.anl.gov/research/projects/mpi/>
3. OpenMP. <http://openmp.org/wp/>
4. Intel® Math Kernel Library for Linux* OS User's Guide. https://software.intel.com/sites/products/documentation/doclib/mkl_sa/11/mkl_userguide_lnx/mkl_userguide_lnx.pdf
5. Intel Xeon Phi Coprocessor System Software Developers Guide (2013) Intel Corporation
6. Parallel Programming and Optimization with Intel Xeon Phi Coprocessors (2013) Handbook on the development and optimization of parallel applications for Intel Xeon processors and Intel Xeon Phi coprocessors. Colfax International
7. Intel® Xeon Phi™ Coprocessor—the Architecture. <https://software.intel.com/en-us/articles/intel-xeon-phi-coprocessor-codename-knights-corner>
8. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
9. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Top Comput* 2(3):376–387
10. Saad Y (2009) *Iterative methods for sparse linear systems*, 2nd edn. Science Press
11. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50
12. Intel® Xeon Phi™ Coprocessor Developer's Quick Start Guide. <https://software.intel.com/sites/default/files/managed/f5/60/intel-xeon-phi-coprocessor-quick-start-developers-guide-mpss-3.2.pdf>

Modeling and Assessing the Helpfulness of Chinese Online Reviews Based on Writing Behavior

Chenglei Qin, Xiao Wei, Li Xue and Hongbing Cao

Abstract At present, consumers are accustomed to judge the quality of goods according to online reviews. However, e-commerce sites are always filled with lots of less useful reviews, which is inconvenient for customers. This paper proposes a method for assessing the helpfulness of Chinese online reviews based on writing behavior. The proposed method recognizes the writing behavior, such as Tail-Insertion, Non-Tail-Insertion and Selected-Modification by monitoring the change of the comment input box on goods page, and then a linear weighted model is established on the writing behavior, writing speed and product features of the review and used to assess the helpfulness of reviews. Experimental results show that the model can accurately and efficiently recognize the useful reviews.

Keywords Assess the helpfulness of online reviews · Writing behavior · The quality of reviews · Product features

C. Qin (✉) · X. Wei · L. Xue
School of Computer Science and Information Engineering, Shanghai Institute
of Technology, Shanghai 201418, China
e-mail: qct2009@qq.com

X. Wei
e-mail: shawnwei@outlook.com

L. Xue
e-mail: xueli@sit.edu.cn

H. Cao
School of Computer and Information Engineering, Fuyang Teachers College,
Fuyang, Anhui 236000, China
e-mail: chb19811117@126.com

1 Introduction

In 2004, Godes et al. [1] found that nearly 50 % of consumers refer to the associated reviews of a product before making purchase decisions. Similar conclusions were obtained in “2011 Cone Online Influence Trend Tracker” released by Cone Communication: 64 % of consumers mainly verify the quality of the goods by the relevant product reviews [2]. However, in recent years, with the rapid development of e-commerce and the scale of online shoppers increasing, the quantity of a commodity’s comments has increased significantly [3], and besides, the characteristics of online reviews, such as updating fast, high subjectivity and low normalization, are becoming more and more obvious [4]. And existing methods are not able to assess the helpfulness of reviews in real-time actually [5–13]. Therefore, it is difficult for potential consumers to make purchase decisions accurately and efficiently in the face of the large quantity of online reviews which vary considerably in quality. In this paper, we propose that the writing behavior, writing speed and product features are important factors in assessing the helpfulness of online reviews, and the experiment has verified our hypothesis.

2 Observation and Analysis

In this section, we discuss three factors that influence the helpfulness of reviews: writing behavior, writing speed and product features.

2.1 Writing Behavior

A high quality review is always obtained by several times of modification, such as delete operation. It means that users make some “decoration” for the review, which shows with writing behavior and writing speed. Writing behavior can be recognized by monitoring the change of the comment text box, a monitor record is $dst(b, w, t)$, b represents writing behavior, w represents the corresponding characteristics, t represents the time of $b(ms)$ cost, for example, $dst(Tail - Insertion, beauty, 992ms)$. There are three states of the length of the comment text box value while the value changes in adjacent time: $L_{i+1} > L_i, L_{i+1} < L_i, L_{i+1} = L_i$ (L represents the length of the comment text box value) (Fig. 1).

The first case ($L_{i+1} > L_i$) represents one of the following writing behaviors: Tail-Insertion, Mid-Insertion, Head-Insertion, Tail-Selected-Modification, Mid-Selected-Modification, Head-Selected-Modification. Compare the comment text box value at t_i with the value at t_{i+1} , take the length of the comment text box value at time t_i as the object of reference, the same characteristics indicated by “1” and the differences indicated by “0”, finally we get the sequence of S which is shown in Fig. 2.

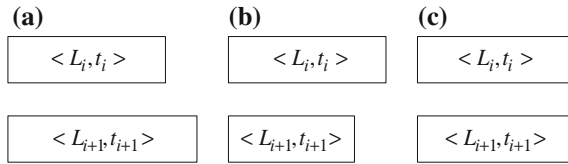


Fig. 1 The changes of the length of the comment text box value in adjacent time

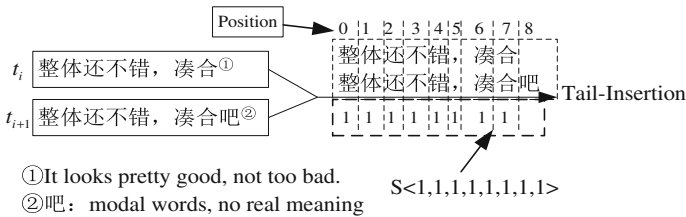


Fig. 2 A schematic diagram of Tail-insertion

We can find that the writing behavior shown in Fig. 2 is Tail-Insertion. The rest recognition methods of the writing behavior are similar to the above.

In order to reduce the computing scale of the model of assessing the helpfulness of online review, some adjustment as follows: merge Head-Insertion and Mid-Insertion to Non-Tail-Insert; Merge Head-Deletion and Mid-Deletion to Non-Tail-Deletion; Merge Head-selected-Modification, Mid-selected-Modification and Tail-selected-Modification to Selected-Modification. Therefore, there are five kinds of writing behavior in this paper, namely Tail-Insertion, Non-Tail-Insertion, Tail-Deletion, Non-Tail-Deletion, Selected-Modification, which are abbreviated as TI, NTI, TD, NTD and SM respectively.

2.2 Writing Speed

According to the analysis of the experimental data, we found that one of the important factors affecting the helpfulness of online reviews is writing speed. The writing speed mentioned in this paper can not narrow understanding of the speed of characteristics inputted by users, it points to the ratio of the length of the comment text box value and the time spent by each writing behavior. It means that the writing speed can reflect the degree of users' attention clearly while writing reviews.

Each change of the value of the comment text box corresponds to a writing behavior happen, and a record is created: $dst(b, w, t)$. After users submit comment, The server has a pretreatment process in calculating the writing speed: statistic the

length of the comment text box value. If the length is too short ($L \leq 5$), give up assessing the helpfulness of the review. Otherwise, use formula 1 to calculate the writing speed.

$$writing_speed = L / \sum_{i=1}^n t_i \quad (1)$$

2.3 Product Features

For understanding the main features of the goods better, such as size, color, weight, most of the potential consumers always view the reviews more carefully. And reviews usually have higher reference value to potential consumers which contain detailed description or evaluation for product features. Therefore, a review includes several product features will enhance the helpfulness.

The difficulty of identifying product features included in comments is lower in a known collection of product features. The collection of product features can be defined through manual acquisition and automatic acquisition. However, it is difficult for manual acquisition to guarantee the efficiency of the system, and besides, automatic acquisition requires amount of reviews while the number of online reviews is too small at the initial stage. According to the research of online reviews, we propose a feasible scheme. The product features recognition is divided into two stages. In the first stage, because of the small amount of reviews, we take the noun phrase and the noun morpheme as the product features while the number of the reviews reaches 200.

3 The Model of Assessing the Helpfulness of Online Reviews

We establish a linear weighted model for the three factors, namely writing behavior, writing speed and product features, to assess the helpfulness of online reviews in this section.

After users submit their reviews, the writing behavior of a review is known. Definition $H(w)$: the score of assessing the writing behavior. Statistic the frequency of the change of writing behavior in adjacent time, denoted by Q ; And statistic the kinds of writing behavior in a review, denoted by P . The change process of writing behavior is shown in Fig. 3.

In summary, the formula of calculating the score of writing behavior is as follows.

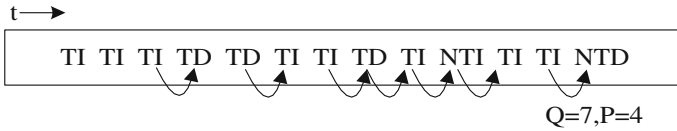


Fig. 3 The change process of writing behavior

$$H(w) = lg(Q * P) \tag{2}$$

While the value of $lg(Q * P)$ is bigger, it represents that users do some more modification for the reviews.

In Sect. 2.2, we have already put forward a formula to calculate the writing speed. Definition $H(v)$: the score of assessing the writing speed, the formula of calculating the score of writing speed is as follows.

$$H(v) = \sum_{i=1}^n t_i/L \tag{3}$$

In Sect. 2.3, we have illustrated the product features detailed. Definition $H(t)$: the score of assessing the semantic features; Definition F : the amount of product features of a review. The formula of calculating the score of semantic features is as follows.

$$H(t) = F/C \tag{4}$$

C is a constant ($C = 10$), usually the maximum number of the product features of a review is less than 10. The bigger of the value of F/C is, the more product features contained in a review that would improve the helpfulness of the review.

Definition H : the score of assessing the helpfulness of a review. Combine formula 2, 3 and 4. The model of assessing the helpfulness of reviews is as follows.

$$H = \alpha \cdot lg(Q * P) + \beta \cdot \sum_{i=1}^n t_i/L + \gamma \cdot F/C \tag{5}$$

The values of parameters α , β and γ are 0.213, 0.002 respectively, which can be obtained from experimental results.

4 Experiment

In order to verify the usefulness of the model, we develop a complete assessing system. Experiment is done with the help of three groups of volunteers who come from different departments in the university. The volunteers of A (25 people) are

using the same mobile phone (iphone 5s), the volunteers of B (28 people) are using another mobile phone (MI 3), and both of the volunteers of C (50 people) hope to have any one of the two (potential consumers).

We make a comparative experiment. The volunteers of A asked to write reviews for the mobile phone (*iphone 5s*) they are using on the assessing system, and the score of the review will be given according to the model at moment. After the volunteers of A finished the task, the volunteers of C begin to assess the helpfulness of the reviews by voting. Compare the score of the reviews with the votes, the usefulness of the model is verified. The results are shown in Table 1.

Table 1 The result of the comparative experiment

No.	Reviews	Votes	Score
1	外观很时尚, 机身修长有型, 系统很流畅没有卡顿现象, 材质也很给力, 造型很精致, 拿在手里感觉很轻薄, 单手操作无压力 (Appearance is fashion, love the slender, thinner body and smooth system, The design is delicate, operating by single hand is relaxed)	42	0.78389967
2	拍照很强, 系统很流畅, 机身轻薄, 土豪金 (Photo effects is pretty good, system is smooth, thinner body, and love the luxury gold color)	34	0.759067197
3	作为一款手机真的还不错, 硬件做的很好, 工艺也很好, 硬件与系统完美结合 (Everything is good)	39	0.688091428
4	手机外观不错, 屏幕相对比较大, 只是屏幕边缘部分不太灵敏, 新增指纹解锁, 安全性比较高 (Appearance is good, the screen is relatively large, the edge of the screen is not sensitive, New added Fingerprint unlock is improve the security)	41	0.686018252
5	系统很流畅, 指纹解锁很灵敏, 握在手里手感真好, 整体感觉还不错. (The system is smooth, Fingerprint unlock is highly sensitive, cool to the touch, and the overall feel is good)	29	0.608792917
6	感觉性价比不高, 玩游戏容易发烫, 外壳不经摔, 一摔一个坑, 还是支持国产吧 (The cost performance is low, it is easy to burn while playing games and easy to sunken while falling down, support domestic)	26	0.60790143
7	玩游戏很流畅, 电池不耐用, 其他还可以. (It is smooth while playing games, and the battery life is short, others is ok)	24	0.482812442
8	像素高, 拍照效果很好, 就是声音有点小, 其他还不错 (Photo pixels is high, photos look good, and volume too low, others are good)	30	0.45914675
9	内存不够用, 还容易发烫, 电池也不行, 亏了 (Not enough memory to use, also easy to burn, it is not worth buying)	14	0.45816215
10	手机还是不错的, 电池是硬伤 (Good phone, but the battery is flawed)	32	0.458051789
11	价格高, 容量小, 性价比不高 (High price, small capacity, and low cost performance)	21	0.451975231

(continued)

Table 1 (continued)

No.	Reviews	Votes	Score
12	前摄像头形同虚设, 经常卡顿, 不值得去买 (The front camera is useless, skip a bit often, it is not worth buying)	12	0.30887195
13	总体感觉还是比较好的, 机身轻薄, 就是发热量有点大 (The overall feel is good, thinner body, but big heat)	27	0.307405
14	屏幕不够大, 没什么新意 (The screen isn't big, and not much innovation)	14	0.302652
15	卡顿, 防抖功能不明显 (Skip a beat, anti shake function is not obvious)	7	0.302569107
16	电池续航不行, 外壳容易磨损 (Battery life is short, and easy to wear)	11	0.30208
17	无论外观还是性能, 都是一款不错的手机 (It's a good phone, regardless of appearance or performance)	16	0.302078947
18	要越狱, 不然各种不方便 (It need to break, or with lots of inconvenient)	12	0.157425481
19	拿在手里很能彰显个性, 价格再便宜点更好 (Great personality, it would be better if there is a discount)	8	0.152811368
20	价格有点贵, 其他都还好 (The price is a little expensive, others are good)	13	0.152424364
21	手机用来玩游戏不错, 玩游戏很少有闪退 (It's good for playing games, few flash back)	2	0.152317444
22	就是太贵了, 等降价再买吧 (It's too expensive, I'm waiting for the price to come down)	6	0.152316667
23	屏幕还是不够大 (The screen of the mobile phone is not big enough.)	9	0.151858571
24	没什么毛病 (There is no problems.)	0	0.1517785
25	很不错的手机, 大爱 (It's pretty good, I love it very much.)	4	0.001442222

5 Conclusions

It is usually difficult for people to get some helpful information among the huge amounts of reviews. Based on the analysis of online reviews, we propose a model to assess the helpfulness of Chinese online reviews from a new perspective. It has been testified that this method is practicable. And the method will make a better shopping experience for users.

Acknowledgements We would thank the volunteers for their active participation in the experiments.

References

1. Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Mark Sci* 23(4):545–560
2. Cone Company: cone online influence trend tracker (2011) <http://www.conecom.com/contentmgr/showdetails.php/id/4008>
3. CNNIC: Statistical report of the thirty-first Chinese internet development (2014)
4. Lin YM, Wang XL, Zhu T, Zhou AY (2014) Survey on quality evaluation and control of online reviews. *J Softw* 25(3):506–527 (in Chinese)
5. Yang M, Wei Q, Bin Y-X, Li Y (2012) Utility analysis for online product review. *J Manage Sci Chin* 15(5):55–73
6. Panasiuk P, Saeed K (2010) A modified algorithm for user identification by his typing on the keyboard. *Adv Intell Soft Comput* 84:113–120
7. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
8. Liu J, Cao Y, Lin CY (ed) (2007) Low-quality product review detection in opinion summarization. In: Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning. Association for Computational Linguistics, Stroudsburg, pp 334–342
9. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Top Comput* 2(3):376–387
10. Lu Y, Tsaparas P, Ntoulas A, Polanyi L (2010) Exploiting social context for review quality prediction. In: Proceedings of the 19th international conference on World Wide Web. New York, USA, pp 691–700
11. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Generation Comput Syst* 43–44:40–50
12. Wei J, Li Z, Yi D et al (2013) Analyzing Helpfulness of Online Reviews for User Requirements Elicitation. *Chin J Comput* 36(1):119–131
13. Jin L (2013) Research on the approach of automatically identifying usefulness of search product reviews based on SVM. Harbin Institute of Technology (2013)

The Average Path Length of Association Link Network

Shunxiang Zhang, Xiaosheng Wang and Zheng Xu

Abstract Association Link Network (ALN) which can effectively support many Web intelligent activities such as Web-based learning, Web knowledge discovery and semantic search. As one of the most robust measures of network topology, average path length plays an important role in the transport and communication within a network. This paper analyzes the average path length of Association Link Network (ALN). First, the network density of ALN is analyzed to get the functional relation between network density and its parameters. Then, based on the achieved result of network density function, the approximate solution of average path length for ALN is deduced. The experimental results show our approximate solution has high precision.

Keywords Association link network · Average path length · Network density · Approximate solution

1 Introduction

Many investigations have been done to recognize the structure of various networks and to analyze emerging complex properties [1–4]. Most of real web-based networks share three prominent structural features: small average path length (APL),

S. Zhang (✉) · X. Wang
School of Computer Science and Engineering, Anhui University of Science
and Technology, 232001 Huainan, China
e-mail: sxzhang@aust.edu.cn

X. Wang
e-mail: 825916595@qq.com

Z. Xu
The Third Research Institute, The Ministry of Public Security, Shanghai, China
e-mail: xuzheng@shu.edu.cn

Z. Xu
Tsinghua University, Beijing, China

high clustering and scale-free (SF) degree distribution [1–3]. Watts and Strogatz [1] have analyzed how the small APL arises, Barabási and Albert have introduced where the scale-free distributions in real networks comes from.

Association Link Network (ALN) is a kind of Semantic Link Network built by mining the association relations among Web resources for effectively supporting Web intelligent application such as Web semantic association search, Web knowledge discovery and recommendation [3, 4]. With the rapid development of information technology, human kinds are more likely to read and share information by similar intelligent applications. For example, the distributed and collaborative learning [5], semantic representation of scientific documents for supporting e-learning [6], discovering and searching of correlation between shared resources [7], and smart component technologies for human centric computing [8], etc.

The small APL of ALN has been discovered in our prior work. In this paper, we further explore the approximate solution of average path length for ALN. The major contributions of our work include two aspects: (1) The analysis of link density of ALN is given first. The function relation between the link density of ALN and the dynamic threshold was found according to regression analysis on the actual value of link density. (2) A theorem for adjustable APL of ALN was deduced, which presents the approximate solution for average path length.

The rest of this paper is organized as follows: Sect. 2 introduces the analysis of link density of ALN. Section 3 presents the theorem for adjustable APL of ALN. Section 4 is comparison experiments between the approximate solution and actual solution. Conclusions are given in Sect. 5.

2 The Analysis of Link Density of ALN

In this section, a definition of the density of ALN is presented first. Then the relation between the density of ALN and dynamic threshold is analyzed.

Definition 1: Link Density of ALN

The link density of Association Link Network (ALN) reflects the dense degree of directed links among all nodes in ALN. According to the graph theory, it can be defined as a value that the number of all existing edges/arcs/links divides the maximal possible number in ALN. So the density of ALN $D(\alpha)$ can be defined as.

$$D(\alpha) = \langle k(\alpha) \rangle * N/N(N - 1) = \langle k(\alpha) \rangle / (N - 1) \quad (1)$$

where $\langle k \rangle$ denotes the average degree of ALN, N is the network size of ALN, and α denotes the dynamic threshold which is adjustable parameter of density.

To find the possible function relation between the link density and the dynamic threshold, we select three fields Web resources (i.e. health, environment and internet) to build 6 ALNs according to the construction method [3, 4]. Then the filtering algorithm of ALN [3] is used to compute APLs of 6 ALN on the several

Table 1 The information of metadata definition

	Health	Environment	Internet
Network size	3509	3284	1615
	8168	6690	4158

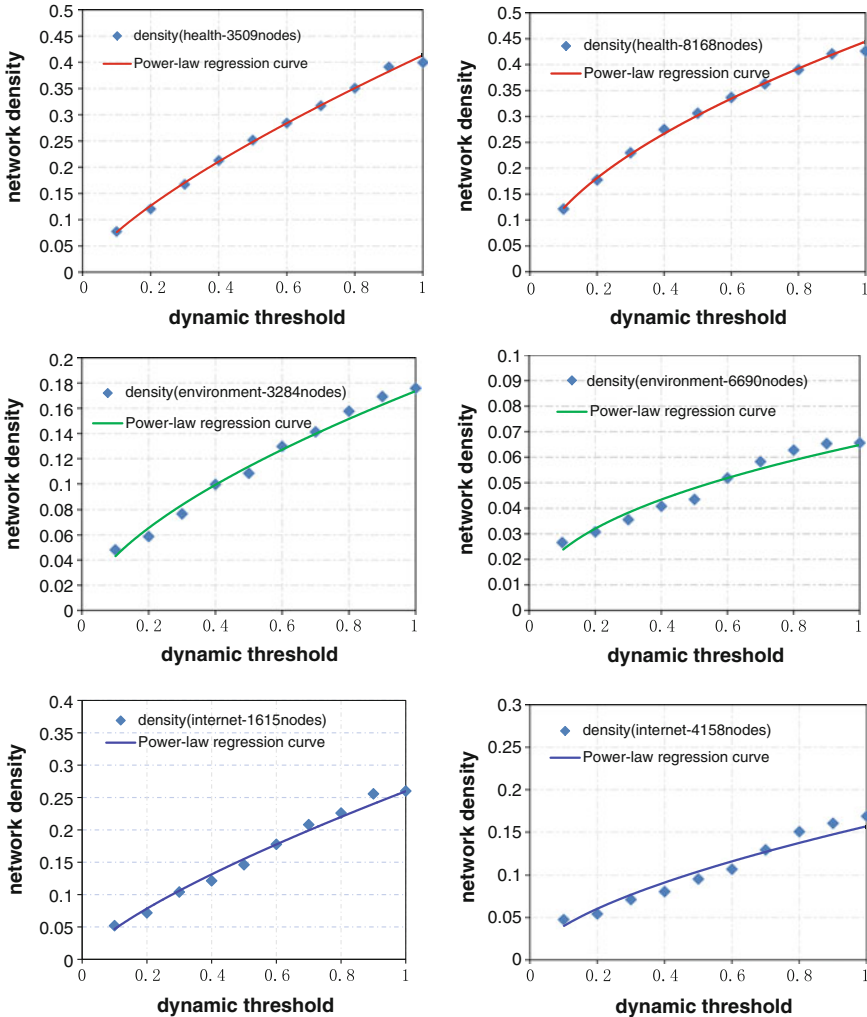


Fig. 1 The function relation analysis between density of ALN and dynamic threshold

network sizes and 10 adjustable dynamic threshold. After that, we analyze the possible function relation between the link density and the dynamic threshold by the method of regression analysis. The network sizes are shown in Table 1. The computing results of APL and regression analysis are plotted as Fig. 1.

Table 2 The information of metadata definition

	Health	Environment	Internet
<i>con</i>	1.05	1.01	0.98
<i>b</i>	0.55	0.54	0.51

From Fig. 1, we find that $D(\alpha)$ and α follow power-law function as formula (2). The high correlation coefficient (>0.9) indicates that the regression analysis has a high trustiness.

$$D(\alpha) = con * \alpha^b \tag{2}$$

where *con* is constant coefficient which is achieved by regression analysis, *b* is a power-law exponent. The rough values of *con* and *b* are shown in Table 2.

From the regression results listed in Table 2, we found that the value of the constant coefficient *con* is about 1, and the value of the power-law exponent *b* is about 0.5. In addition, the ALN in health field has the biggest power-law exponent, the ALN in environment field has the second bigger power-law exponent, and the ALN in internet has the smallest power-law exponent.

3 The Approximate Solution of Average Path Length of ALN

In this section, a theorem for adjustable APL of ALN was given first, then two corollaries about APL of ALN are presented. It can facilitate to rapidly estimate or compute the APL of ALN at a given network size and dynamic threshold.

Theorem 1 *The approximate solution for adjustable APL of ALN.*

*For a ALN of a given field, its network size is denoted as N . If the dynamic threshold $\alpha \in (0, 1]$, then there exists $L = \ln(N) / [\ln(N) + \ln(con) + b * \ln(\alpha)] + \Delta(\alpha)$. Where $\Delta(\alpha)$ is a increment function which is related to random graph on the same network size and average degree.*

Proof According to graph theory, the link density of ALN can be represented as,

$$D(\alpha) = \langle k(\alpha) \rangle / (N - 1) \tag{3}$$

In addition, according to the regression results [formula (2)] in Sect. 2, we have,

$$\langle k(\alpha) \rangle = (N - 1) * con * \alpha^b \tag{4}$$

From the experimental results in [3], we have,

$$L = \ln(N) / \ln(k(\alpha)) + \Delta(\alpha) \tag{5}$$

Substituting formula (4) into formula (5),

$$\begin{aligned}
 L &= \ln(N) / \ln[(N - 1) * con * \alpha^b] + \Delta(\alpha) \\
 &= \ln(N) / [\ln(N - 1) + \ln(con) + b \ln(\alpha)] + \Delta(\alpha)
 \end{aligned}
 \tag{6}$$

Thus Theorem 1 is proved.

Corollary 1 *If there is $con \rightarrow 1$, the approximate solution for adjustable APL of ALN can be reduced as $L = \ln(N) / [\ln(N) + b * \ln(\alpha)] + \Delta$.*

Corollary 1 is obvious result. Because if there is $con \rightarrow 1$, then there must be $\ln(con) \rightarrow 0$. So, $\ln(N) + \ln(con) + b * \ln(\alpha) \rightarrow \ln(N) + b * \ln(\alpha)$.

Corollary 2 *If there is $\alpha \rightarrow 0$, then we have $L \rightarrow \infty$. On the contrary, if there is $\alpha \rightarrow 1$, then we have $L \rightarrow \ln(N) / \ln(N - 1) + \Delta \rightarrow 1 + \Delta$.*

We understand the Corollary 2 from two aspects:

- (1) Theory view. This corollary is a truth. Because $\alpha \rightarrow 0$ means that there is no any links in an ALN. All the nodes in an ALN are isolated nodes. So there is $L \rightarrow \infty$. On the contrary, if there is $\alpha \rightarrow 1$, then $\ln(\alpha) \rightarrow 0$. So we have $L \rightarrow \ln(N) / \ln(N - 1) + \Delta \rightarrow 1 + \Delta$.
- (2) Application view. Theorem 1 has an important application value except these two extreme cases. When $\alpha \rightarrow 0$, there is no any links in an ALN. It is difficult to find the association among nodes. Accordingly, some knowledge service such as association knowledge flow, semantic search etc. will not be finished. On the other hand, when $\alpha \rightarrow 1$, ALN is closed to the original ALN. Too many links including many weak links lead to high complexity of ALN. This condition not only increases the computing complexity in finding knowledge service, but also decreases the precision of knowledge services.

4 Experiments

In this section, we compute the approximate solution of APLs of 6 ALNs listed in Table 1, and comparing them with the actual solution. The comparison experiments are shown as Fig. 2.

From Fig. 2, we find that the approximate solutions of APLs are all very closed to their actual solutions on the same network size and dynamic threshold based on 6 ALNs. The average error for approximate solution is smaller than 3 %. These experimental results show that the approximate solution of APL is reliable.

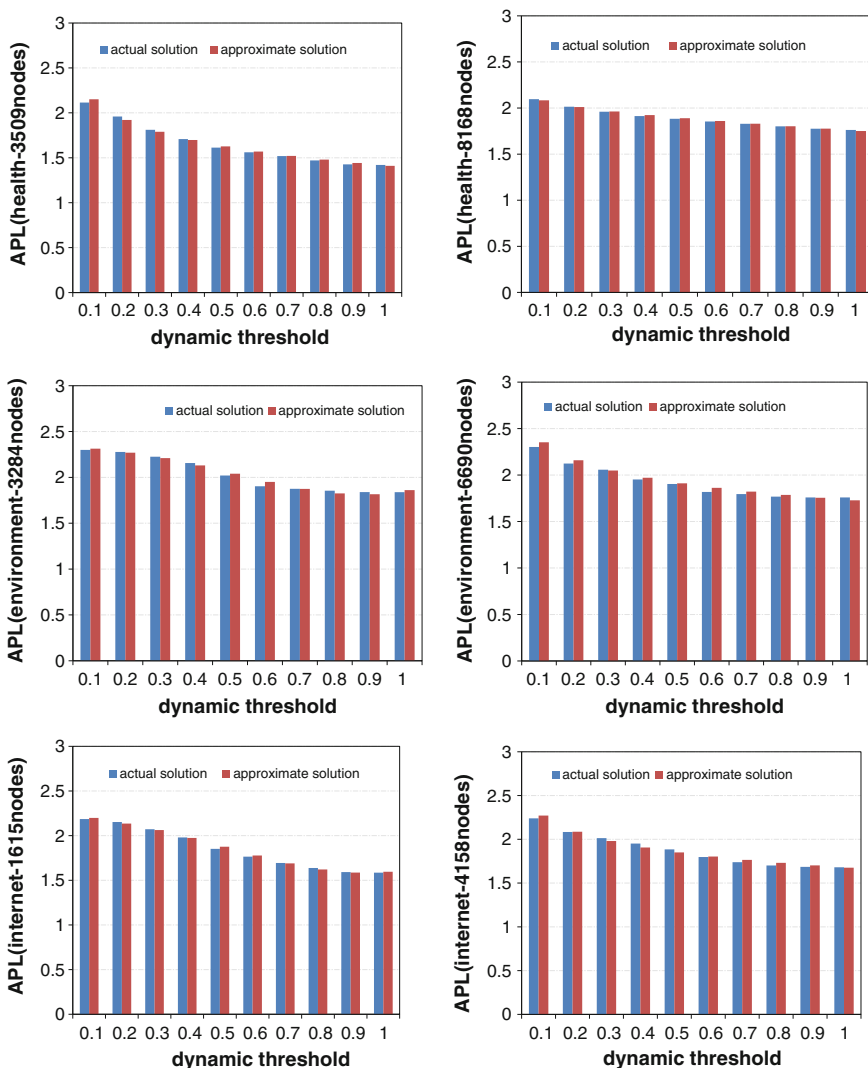


Fig. 2 The comparison experiments between approximate solution and actual solution for average path length

5 Conclusions

The approximate solution for adjustable APL can avoid the high time complexity in computing the shortest path between any two nodes in ALN. In this paper, we have explored the power-law function relation analysis between density of ALN and dynamic threshold by regression analysis. Further, we have presented a theorem

about the approximate solution for adjustable APL of ALN. Experimental results have demonstrated the validity of the approximate solution for adjustable APL. The theorem can provide the foundation on understanding the structure of ALN when we filter ALN to support Web intelligent application such as Web semantic association search, Web knowledge discovery and recommendation.

Acknowledgments This Research work was supported in part by the Natural Science Foundation of Anhui Province (No. 1308085MF94), by the National Science Foundation of China (Grant No. 61300202), and by the Opening Project of Shanghai Key Laboratory of Integrate Administration Technologies for Information Security (No. AGK2013002).

References

1. Watts DJ, Strogatz SH (1998) Collective dynamics of “small-world” networks. *Nature* 393:440–442
2. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45:167–256
3. Zhang SX, Luo XF, Xuan JY et al (2014) Discovering small-world in association link networks for association learning. *World Wide Web* 17(2):229–254
4. Luo XF, Xu ZH, Yu J et al (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
5. Li Q, Lau RWH, Shih TK et al (2008) Technology supports for distributed and collaborative learning over the internet. *ACM Trans Int Technol* 8(2):10:1–10:24
6. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Topics Comput* 2(3):376–387
7. Yen NY, Huang RH, Ma JH, Jin Q, Shih TK (2013) Intelligent route generation: discovery and search of correlation between shared resources. *Int J Commun Syst* 26:732–746
8. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50

The Intelligent Big Data Analytics Framework for Surveillance Video System

Zheng Xu, Yang Liu, Zhenyu Li and Lin Mei

Abstract Currently, with the explosion of multimedia data (image, video and audio) from remote sensors, mobile image captures, social sharing, the web, TV shows and movies, huge volume of images are being generated and consumed daily. The availability of massive images has created fundamental challenges to image processing and analysis. Big Data is a term used to refer to massive and complex datasets made up of a variety of data structures, including structured, semi-structured, and unstructured data. To address these challenges, we propose a model design methodology using collective intelligence for big data analytics. The data and data-transfer contracts then become the primary organizing constructs. With controlled data relations and timing, the system can then be built from independent agents with loosely coupled behaviors. This data-driven design technique is naturally supported by the Data Distribution Service (DDS) specification, which is a standard from the Object Management Group.

Keywords Big data · Surveillance video system · Hadoop

1 Introduction

Currently, with the explosion of multimedia data (image, video and audio) from remote sensors, mobile image captures, social sharing, the web, TV shows and movies, huge volume of images are being generated and consumed daily. The availability of massive images has created fundamental challenges to image processing and analysis. Big Data is a term used to refer to massive and complex

Z. Xu · Z. Li (✉) · L. Mei

The Third Research Institute of Ministry of Public Security, Shanghai, China

Z. Xu

Tsinghua University, Beijing, China

Y. Liu

Shanghai University, Shanghai, China

datasets made up of a variety of data structures, including structured, semi-structured, and unstructured data. Today, businesses are aware that big data can be used to generate new opportunities and process improvements through their processing and analysis. The emergence of big data has brought about a paradigm shift to many fields of computing. We have seen remarkable advances in computing power and storage capacity of big data management. But most big data systems currently in use handle data types of text or numbers. Novel and scalable data management and analytical frameworks are needed to meet the challenges posed by the big images. With the development of meteorological instrumentation and the network of surface weather stations, much more data was being collected, which leads to an extreme increase in the data's capacity. At the same time, a higher demand for easy access to all the data and new storage requirements are collected by end users [1, 2], especially for the grid data, the unstructured data. However, traditional data storage and management system (e.g., native file systems and SDBMS) [3, 4] cannot support massive data storage and processing. Cloud computing and distributed file system can give us a new resolution to these problems. Derived from Google's MapReduce and Google File System (GFS) papers, Apache Hadoop [5, 6] is an open source software framework that supports data-intensive distributed applications. Through cloud computing technology including distributed storage and computing framework, the problem brought by the huge amounts of data would be solved. Cloud computing can speed up the big data processing and storage efficiency. The design of a big data analytics system differs considerably from that of a traditional database-supported decision support system (DSS). Such a system involves more entities, data and participants; therefore, the system has special requirements in terms of data management, model design and quality of service (QoS). To address these challenges, we propose a model design methodology using collective intelligence for big data analytics. According to Wikipedia, collective intelligence (CI) is the shared or group intelligence that emerges from the collaboration, collective efforts, and competition of many individuals and appears in consensus decision making. CI systems, including multi-agent systems, complex adaptive systems, swarm intelligence and self-organizing systems, are complex by nature. The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 describes the basic framework. Section 4 proposes a surveillance video system. Finally, we draw the conclusions of this research.

2 Related Work

Big data analytics address large volumes and distributed aggregations of various types of data. The data may be from audio, video, social networks, or web forums. Big data no longer relies on databases or data warehouses. No SQL methods, such as data management and processes in memory, are incorporated into the system. Therefore, integrating different data management mechanisms is a considerable challenge. Big data analytics models are not typically predefined due to the

presence of dynamic environments. Such models typically require iterative solutions for testing and improvement. Moreover, business processes in big data analytics systems should be flexible [7]. Participants in such models include software systems, mobile devices, web services, and humans. Building dynamic business processes that allow for cooperation amongst various participants is another challenge. QoS is also vital in big data analytics systems [8]. Jacobs [9] states that “It is easier to get the data in than out”. Systems occasionally need to react to an event, such as a service outage or a change in a patient’s medical condition, in real time. Another problem is that data are often incomplete; therefore, they are inferred probabilistically, and the analysis results are fuzzy. Thus, obtaining an overview of QoS properties of the system during design is the third challenge work, improve the level of intelligent video surveillance system.

3 The Intelligent Big Data Analytics Framework

From different sources, meteorological data may include observation data, forecast and service product data, mete data. Some are collected from local observation network, some are received by CMA cast and others are shared by around provinces. Data of the same category, due to different sources, may be stored in different formats. Therefore, a wide variety of diverse formats, different forms of expression, a huge amount of data, complex category, are the characteristics of meteorological data. According to Miller and Mork, big data analytics means a value chain from data to decisions through a series of processes, including data discovery, data integration, and data exploitation. These data processes are not new to software systems. However, how to implement them to support the aforementioned features in a big data analytics environment is a new challenge. In this research, we choose the multi-agent paradigm in CI. The key to the design is to separate data from behavior. Each behavior is addressed by a group of agents. The data and data-transfer contracts then become the primary organizing constructs. With controlled data relations and timing, the system can then be built from independent agents with loosely coupled behaviors. This data-driven design technique is naturally supported by the Data Distribution Service (DDS) specification, which is a standard from the Object Management Group. Triangles represent intelligent agents. The solid triangles are the administrators in a group of agents. Administrator agents attempt to create and manage other agents involved in a certain task. A group of agents that share the same goal base is called an agent platform.

The Data Management Layer provides the basic process functions for various types of data. Different agent platforms perform different actions on SQL-like, non-SQL-like or memory-based data for storage, access and integration. Because big data systems typically require real-time functionality, such as in stream computing technology, we design several administrator agents to perform such no-storage data processes combined with traditional database operations in

distributed agent platforms. The processes in stream computing would be rapidly organized by the multi-agent mechanism as soon as the data are provided.

The Data Analysis Layer attempts to analyze data for data exploitation and decision making. The main actions on the data include query, model, analyze and visualize. Each action can be supported by a group of agents with their goal and knowledge base. These agents represent existing software systems, web services, cloud applications or any other participants in an organization. A composition of the four types of agents can be used in various business applications.

The QoS Layer provides information exchange contracts between agents. Traditional messaging designs focus on functional or operational interfaces. However, in the multi-agent system, the interface specifies the common, logically shared data model that are produced and used by an agent along with the QoS requirements as its goal, including timing, reliability, workload, and security. With such explicit QoS terms, responses to impedance mismatches can be automated, monitored, and governed.

4 Distributed Surveillance Video System Based on Hadoop

Huge small raw source files (several M or less than 1 M) with high update frequency (less than 6 min, some in seconds) should often be searched or queried in business systems, at the same time the results back to the ends users are requested to be finished in several seconds. The current data storage and download system don't find good solutions to it. In this paper, we try to first combine these small files to one big file and put it into HDFS, then begin some experiments.

Observation data. Observation data includes ground and upper-air sounding, agricultural data, radar, satellite, automatic weather station data, wind profile, GPS, microwave radiometer. Basically, Observation data is in text form, in the form of an Access database, in binary form or image format. The source data files in those categories such as automatic weather stations, microwave radiometer, road stations, atmosphere observation, radiation, in one month period, are combined into one file by using Sequence method. Objective field grid data Forecast and service product data includes four categories: numerical analysis and instructor products received from CMA cast and stored in binary format with GRIBS code, local numerical model instructor products in net cdf format, analysis products from MICAPS in micaps text format, and meteorological service product in text format (Doc, pdf, xml). Numerical model forecast products applied in the weather forecast business are from a wide variety system sources, including CMS, BMB, URUION, JAPAN and others. Therefore, there are many differences among file format, coding method, space range and resolution of source data files, which make business system burdened to parse these products to integrate application. The forecast of a single meteorology element (for examples, precipitation, temperature, wind, air

pressure, relative humidity, cloud amount and visibility) is extracted and recorded in net cdf format by interpreting them, as available objective field data for forecast operation.

According to the functions of an agent, the system contains administrator agents and executive agents. An administrator agent attempts to receive a task, make an executive plan, create executive agents, control the QoS, and send orders. The agent has a goal set and a group of calculation functions to make a plan. An executive agent attempts to realize a specific task according to orders from an administrator agent. Communication between an administrator agent and executive agent: An administrator agent sends various orders, such as creation, execution, and delete, to executive agents. This agent also sends a dataset for processing if necessary. The executive agent sends various statuses, such as finished, failed, and interrupted, to the administrator agent. This agent also sends the processed data back to the administrator agent in the data part. Communication between executive agents: If a task requires the collaboration of agents, executive agents can communicate with each other through statuses and data. Communication between administrator agents: The interactions between administrator agents aim at passing data and tasks between them.

5 Conclusions

Currently, with the explosion of multimedia data (image, video and audio) from remote sensors, mobile image captures, social sharing, the web, TV shows and movies, huge volume of images are being generated and consumed daily. The availability of massive images has created fundamental challenges to image processing and analysis. Big Data is a term used to refer to massive and complex datasets made up of a variety of data structures, including structured, semi-structured, and unstructured data. To address these challenges, we propose a model design methodology using collective intelligence for big data analytics. The data and data-transfer contracts then become the primary organizing constructs. With controlled data relations and timing, the system can then be built from independent agents with loosely coupled behaviors. This data-driven design technique is naturally supported by the Data Distribution Service (DDS) specification, which is a standard from the Object Management Group.

Acknowledgments This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, 2013AA014603, in part by National Key Technology Support Program under Grant 2012BAH07B01, in part by the National Science Foundation of China under Grant 61300202, 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009, in part by the China Postdoctoral Science Foundation under Grant 2014M560085, and in part by the Science Foundation of Shanghai under Grant 13ZR1452900.

References

1. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
2. Liu Y, Ni L, Hu C (2012) A generalized probabilistic topology control for wireless sensor networks. *IEEE J Sel Areas Commun* 30(9):1780–1788
3. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Top Comput* 2(3):376–387
4. Liu X, Yang Y, Yuan D, Chen J (2013) Do we need to handle every temporal violation in scientific workflow systems. *ACM Trans Softw Eng Methodol*
5. Wang L, Tao J et al (2013) G-Hadoop: MapReduce across distributed data centers for data-intensive computing. *Future Gener Comput Syst* 29(3):739–750
6. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50
7. Talia D (2013) Clouds for scalable big data analytics. *Computer* 46(5):98–101
8. Marx V (2013) Biology: the big challenges of big data. *Nature* 498(7453):255–260
9. Jacobs A (2009) The pathologies of big data. *Commun ACM* 52(8):36–44

The Intelligent Video Processing Platform Using Video Structural Description Technology for the Highway Traffic

Zheng Xu, Zhiguo Yan, Zhenyu Li and Lin Mei

Abstract The traffic surveillance system is used to quickly and accurately determine the traffic, release traffic information in time, and reduce the number of traffic accidents, traffic jams and road damage. The traffic surveillance system makes the highway fast, safe, comfortable and efficient. The existing surveillance system mainly relies on the surveillance personnel to monitor, but there are too many surveillance videos so that it is difficult to find the user's interested content. In this paper, we propose Video surveillance system which is based on the technology of video structural description. Then we apply this system to the high-speed service area. Video structural description (VSD) is according to the semantic relation and adopting spatiotemporal segmentation, feature extraction, object recognition and so on, putting video content into the text information which understood by the human and computer.

Keywords Intelligent video · Video structural description · Highway traffic

1 Introduction

At present, the transportation system also gradually becomes intelligent. There are three parts of the intelligent traffic [1, 2] which are traffic communication system, traffic surveillance system, charging system. The traffic surveillance system makes the highway fast, safe, comfortable and efficient. The High-speed service area is an important part of highway, it provides the safeguard for the highway closed, high-speed driving. Naturally the High-speed service area of efficient, safe management also constitute an important content of the highway operate efficiently. In order to do the management work of the service area, improve the service management level and reduce the service security hidden danger, we install surveillance

Z. Xu · Z. Yan · Z. Li (✉) · L. Mei

The Third Research Institute of Ministry of Public Security, Shanghai, China
e-mail: lizhenyu1959@yeah.net

Z. Xu

Tsinghua University, Beijing, China

system in the related position of the service area to achieve real-time surveillance of the scene [3]. The existing service area traffic surveillance system are based on the ground sense coil, it needs to be buried coil in advance and must be sealed when implementing service area. It is very inconvenient, and the cost is higher. Most of the video-based traffic surveillance devices place in the city, the light in the High-speed service area is insufficient at night, and the power supply in the High-speed service area is unstable. There are not corresponding improvement plan. With the help of cloud computing [4–7], internet of things [8–10], and Big Data [11], video structural description (VSD) is according to the semantic relation and adopting spatiotemporal segmentation, feature extraction, object recognition and so on, putting video content into the text information which understood by the human and computer. VSD includes two meanings: one is the video semantic content, namely under the standardized video content description standard organization, each interested in the video of the target and its characteristics and behavior is identified, in the form of text to describe the content of the video, this is a video of information extraction process; Second, correlation of video resources, set up single semantic interconnection (across) camera video resources, make use of data mining methods for effective analysis and semantic retrieval become possible, also makes the video resource semantic interconnection with other information systems resources possible, this is a video information organization, management and mining, and auxiliary process of business requirements. In this paper, the video surveillance system based on the technology of video structural description is applied to the High-speed service area. We have video structural description for the actual content of the video in each link of the collection, transmission, storage and use, of the actual content of to the video expressed in structured description. Characteristics of the system: The High-speed service area, based-video, statistics and analysis the vehicle flow, and provides alarm of parking Spaces, stranded and violation. This system includes: the GigE Vision standard camera which is used to output the original video stream; Collection and analysis equipment which is used for video collection, analysis, coding, storage and release, etc.; Real-time streaming media server which is used to publish and playback video; Web application server, which is used to provide analysis report, query and management applications to the police. Beside the camera set up lighting equipment; Collection and analysis equipment have the properties of temperature, voltage and stability etc.

2 Related Work

Intelligent Transportation System (ITS) is the development direction of the future traffic system. This study [12] proposed the bayonet speed measuring system can timely and accurately record the bayonet socket the information of passing vehicles, help to investigate and punish illegal behaviors such as speeding, for public security forensic provides important reference basis. AHD video technology [13] was proposed to establish as system based on hd, intelligent video image, the traffic video surveillance, event surveillance illegal surveillance and surveillance, traffic

surveillance, surveillance and so on many functions. The study [14] adopts B/S mode design a bayonet traffic management platform, relying on the bayonet system the existing data network for surveillance image network transmission, video surveillance technology was adopted to realize information sharing, remote surveillance and management. The study [15] proposed target detection based on the vehicle and the sequence of gray image compression coding, motion estimation bilinear interpolation method, form the final interface coding bit stream, the scheme can effectively solve the intelligent traffic surveillance system image sequence compression coding issues. Chang [4] through intelligent video analysis technology, the passive surveillance to active surveillance, the operation personnel from heavy surveillance work, improve the level of intelligent video surveillance system.

3 The Intelligent Video Processing Platform

The gateway program of Business Logic Layer is responsible for analyzing the results from the terminal of structural description, putting the information into a mysql database, solr, and file systems, and Sending the alarm data obtained from the analytical to the message queue. The gateway program can receive the information from the front-end configuration; Open API system will inheritance to remote invocation framework, is responsible for querying the database and statistics. Then it will put the statistical data which need to alarm in the message queue.

Presentation layer is mainly a Java web application. On the one hand, Presentation layer will access API system by means of remote method invocation (RMI), obtain the results of query and statistical. On the other hand, Presentation layer will get the alarm information from the message queue, and show the two results to the user.

The Video surveillance system in the high-speed service area is based on the video, to analysis the vehicle flow, and provides alarm of parking Spaces, stranded and violation. First we will install GigE Vision standard camera which is used to output the original video stream in the surveillance area, namely in road bayonet. Usually, the camera set up 8 m height above in order to make ultra high vehicles through; Second in the import and export of the high-speed service area respectively set up collection and analysis equipment, which is used for video capture, analysis, coding, storage, distribution, collection and release, etc. Collection and analysis equipment will count the number of vehicles, license plate, the speed of the vehicle, and color information of the vehicle by capturing moving objects in video and vehicles on the video structural description, according to the semantic relation of video content, using spatiotemporal segmentation, feature extraction, object recognition, organized into text information available for computers and human understanding of technology, offering illegal stay, retention service area, parking, early warning and alarm information. In order to reduce the influence of external environment on the cabinet of the equipment, the collection and analysis equipment cabinet is equipped with lightning protection, temperature, voltage regulation equipment. Then, the collection and analysis equipment will transmit collected

information to the service platform. Image information server (or streaming media server) is used to publish and video playback. Application management server (or Web application server) will provide analysis report to the police, query and management of application. Finally, application management server receives the alarm and statistical information, deposited it into the database, and show integration of information in the form of web pages to the user.

This system provides VOD functions. The realization of the function of VOD mainly depends on SQLite. Whenever Open a new file and write a video, needing to add a data in the database, the content including: file address, file name, video start time and Video over time. When the player receives the point in time on demand sent by the server, parsed into 0 points on January 1, 1970 since the start time of (UNIX) the number of seconds, and query the corresponding video files in the database, positioning to the nearest key frame, given the address to send data to the server. When finish sending a file, postpone request time, repeat the above steps, until the length is zero.

4 The Case Study on the Highway Traffic

In this section, the case study on the highway traffic processing platform using the proposed method is introduced. GigE Vision standard collection module is used to obtain raw video stream data of the camera output, and put the data into the collection buffer queue; Video decoding module take raw video frames in RAW format out from the collection buffer queue, decode into YUV420P format, and put them into the encoding queue for encoding module for encoding. While the original video frame is decoded into RGB24 with real timestamp, and output video data for visual analysis through the pipeline.

Visual analysis module analyze the obtained RGB frames to get traffic, license plate, vehicle speed, vehicle color and other information according to the pre-established rules, and sent warning information and statistics information to the Web server. Video encoding module will take the video frame in YUV420P format out from encoding queue, achieve H.263 and H.264 encoding by calling Ffmpeg and x264 encoder. And the package formats are all FLV. Publish video storage module will publish encoded video to the streaming server real-timely and Segmented cycle storage.

This system provides VOD functions. The realization of the function of VOD mainly depends on SQLite. Whenever Open a new file and write a video, needing to add a data in the database, the content including: file address, file name, video start time and Video over time. When the player receives the point in time on demand sent by the server, parsed into 0 points on January 1, 1970 since the start time of (UNIX) the number of seconds, and query the corresponding video files in the database, positioning to the nearest key frame, given the address to send data to the server. When finish sending a file, postpone request time, repeat the above steps, until the length is zero.

Web server receives the alarm and statistical information, deposited it into the database, and show integration of information in the form of web pages to the user.

According to the practical application of video surveillance needs, in the “video structural description key technology research,” the study process, we developed a “semantic knowledge modeling tools video surveillance software.” We use this software to build video surveillance scene semantic knowledge description model, and the model is successfully applied to structural description of intelligent video analysis, to achieve the typical video surveillance structured description of the scene, including the vehicle, license plate, vehicle color, road traffic, pedestrian lane into behavioral descriptions, detection and other functions.

5 Conclusions

In this paper, the video surveillance system based on the technology of video structural description is applied to the High-speed service area. We have video structural description for the actual content of the video in each link of the collection, transmission, storage and use, of the actual content of to the video expressed in structured description. Characteristics of the system: The High-speed service area, based-video, statistics and analysis the vehicle flow, and provides alarm of parking Spaces, stranded and violation. This system includes: the GigE Vision standard camera which is used to output the original video stream; Collection and analysis equipment which is used for video collection, analysis, coding, storage and release, etc.; Real-time streaming media server which is used to publish and playback video; Web application server, which is used to provide analysis report, query and management applications to the police. Beside the camera set up lighting equipment; Collection and analysis equipment have the properties of temperature, voltage and stability etc.

Acknowledgments This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, 2013AA014603, in part by National Key Technology Support Program under Grant 2012BAH07B01, in part by the National Science Foundation of China under Grant 61300202, 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009, in part by the China Postdoctoral Science Foundation under Grant 2014M560085, and in part by the Science Foundation of Shanghai under Grant 13ZR1452900.

References

1. Liu L, Li Z, Delp E (2009) Efficient and low-complexity surveillance video compression using backward-channel aware Wyner-Ziv video coding. *IEEE Trans Circ Syst Video Technol* 19 (4):453–465
2. Baskar L, De Schutter B, Hellendoorn J, Papp Z (2011) Traffic control and intelligent vehicle highway systems: a survey. *IET Intell Transp Syst* 5(1):38–52

3. Haritaoglu I, Harwood D, Davis L (2000) W4: real time surveillance of people and their activities. *IEEE Trans Pattern Anal Mach Intell* 22(8):809–830
4. Liu Y, Ni LM, Hu C (2012) A generalized probabilistic topology control for wireless sensor networks. *IEEE J Sel Areas Commun* 30(9):1780–1788
5. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Topics Comput* 2(3):376–387
6. Liu Y, Zhu Y, Ni LM, Xue G (2011) A reliability-oriented transmission service in wireless sensor networks. *IEEE Trans Parallel Distrib Syst* 22(12):2100–2107
7. Liu Y, Zhang Q, Ni LM (2010) Opportunity-based topology control in wireless sensor networks. *IEEE Trans Parallel Distrib Syst* 21(3):405–416
8. Liu X, Yang Y, Yuan D, Chen J (2013) Do we need to handle every temporal violation in scientific workflow systems. *ACM Trans Soft Eng Methodol*
9. Wang L, Tao J et al (2013) G-Hadoop: MapReduce across distributed data centers for data-intensive computing. *Future Gener Comput Syst* 29(3):739–750
10. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50
11. Xu Z, Luo X, Zhang S, Wei X, Mei L, Hu C (2014) Mining temporal explicit and implicit semantic relations between entities using web search engines. *Future Gener Comput Syst* 37:468–477
12. Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. *Sci Am* 284(5):34–43
13. Zhuge H (2011) Semantic linking through spaces for cyber-physical-socio intelligence: a methodology. *Artif Intell* 175:988–1019
14. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
15. Xu Z, Luo X, Wang L (2011) Incremental building association link network. *Comput Syst Sci Eng* 26(3):153–162

The Scheme of the Cooperative Gun-Dome Face Image Acquisition in Surveillance Sensors

Zhiguo Yan, Zheng Xu, Huan Du and Lin Mei

Abstract How to automatically realize acquisition, refining and fast retrieval of the pedestrian face image in surveillance video is of great importance in public security visual surveillance field. This paper proposes a new gun-dome camera cooperative system which solves the above problem partly. The system adopts static panorama-variable view dual-camera cooperative video-monitoring system. As respect to the face detection, the deep learning architecture is exploited and proves it effectiveness. The experimental results show the effectiveness and efficiency of the dual-camera system in close-up face image acquisition.

Keywords Gun-dome camera cooperation · Calibration · Deformable part model · Pedestrian detection · LBP

1 Introduction

At present, there are two common solutions to trace and capture image of concerned targets in surveillance video. One is on the basis of continuous tracking of moving target with single dome camera, and the other bases on the concerned target collaborative tracing using omnidirectional camera and active camera. In the former solution, when the object of attention appears in the dome camera wide scene, it will be focused, zoomed, and traced continuously. Nevertheless, the other target information in the wide scene will be neglected. Because once one concerned target appears, the only one dome camera fully zooms its view to the target area and start to track the target continuously. For the latter method, the omnidirectional camera transmits position information of the concerned object to the active camera, and the

Z. Yan · Z. Xu (✉) · H. Du · L. Mei

The Third Research Institute of Ministry of Public Security, Beijing, China
e-mail: xuzheng@shu.edu.cn

Z. Xu

Tsinghua University, Beijing, China

object was further consistently traced by the active camera. However, this kind of application mode currently does not have a broad application prospect. The reason is that the narrow application of omnidirectional camera in visual surveillance field and the much high requirement of dual-camera calibration technique, the complicated operation and the high requirement of operators' professional, which be accounted by the inconsistency of the resolution from the centre to the periphery and nonlinear mapping of moving trajectory of the omnidirectional camera.

Combining the advantages of the static-panorama camera and the camera with variable field of view (FOV), we can get the close shot of specific objects in the long shot. Meanwhile, we can also keep the attention to others objects in the distant scenery. By using this mechanism, we can expand the breadth and depth of video surveillance system. In the dome-gun cooperative system, the concerned target can be observed carefully by PTZ mechanism and the rest targets still be observed in the gun-camera view. Moving target detection method can be used to detect the target, since the wide angle camera is static. And if the target motion is not so rapid in the FOV of the panorama camera, it is easy to trace.

As to the gun-dome camera system layout, there are two typical installation methods, see Fig. 1. In Fig. 1a, the dome camera and gun camera are mounted on the trestle side-by-side at the same height. In Fig. 1b, the gun camera is fixed over the dome camera. It must be pointed out that at the initial status, Regardless of what kind of layouts, the gun camera and dome camera has a widely common view and the optical axis are parallel.

Figure 2 shows the passenger snap while they pass the security gateway. In this figure, the left column is the full view of the spot acquired by the gun camera, the right column are the face images acquired by the dome camera under the guidance of the gun camera.

Deformable Part Model (DPM) is used as target detection method in this system. In this paper, the look-up table method is proved feasible when the depth of field of

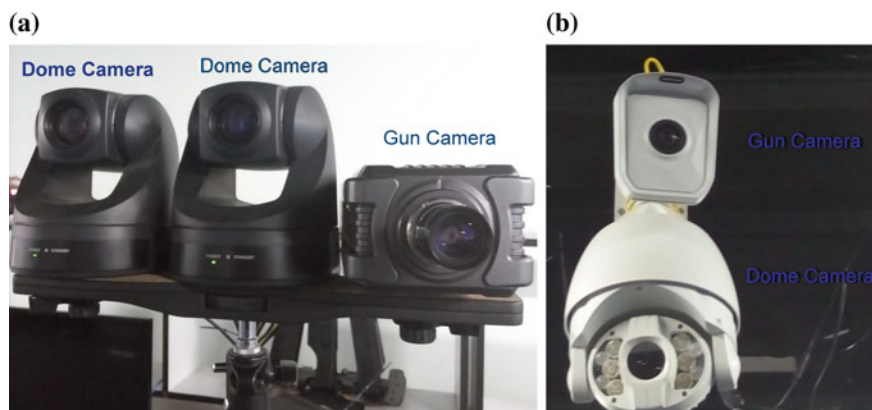


Fig. 1 a Parallel mount. b Vertical mount



Fig. 2 Face image acquisition in surveillance scene

moving target changes small. It is used to calibrate dual-camera system, and it can obtain the angle of rotation which the moving camera rotates to aim at the arbitrary position of the static camera.

2 Methodology

CNN is multi-layer feed-forward architected, which uses the supervised learning to extract invariant multi-stage features from image data. In CNN architecture, the individual neurons are tiled in the way that they respond to overlapping regions. Ideally, the “deep” representation would learn hierarchies of feature detectors and combine the top-bottom and bottom-up processing of an image [1–5]. For instance, lower layers could support object detection. Conversely, information about objects in the higher layers could resolve the lower-level ambiguities. The structure of the CNN is illustrated in Fig. 3.

Normally, a CNN is composed of several stages, as the example showed in Fig. 3, there are two stages [6]. Each stage has a convolutional layer following with a non-linearity operation and a spatial feature pooling layer. The convolutional layer consists of several trainable filter banks and additive bias. All the filters in filter banks can be trained. The filters with a specifically sized window are used to process the small local parts of the input image. The pooling layer lowers the spatial resolution by using a window of specific size, which leads to a strong robustness against geometric distortions. Normally the window size is smaller in lower layers. Because in higher layer, in order to deal with the more complex part of the input image, a window with bigger size is needed to get lower resolution. Before the classification procedure, the features from all positions are combined and fully connected [7]. CNN has the ability to learn not only the low-level features but also the mid-level features.

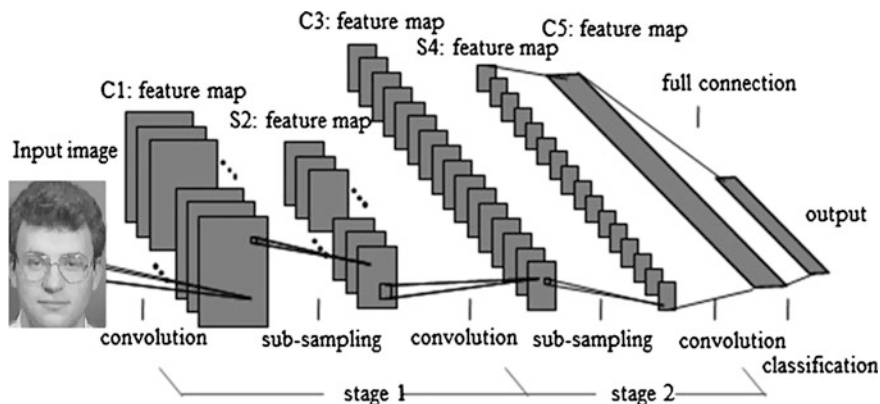


Fig. 3 CNN structure used in face detection

While execute the CNN to detect the face, Simple convolutional and sub-sampling operations were performed instead of the costly pre-processing and filtering operations. The efficiency of detection process can be improved by using two strategies: decrease the checking time for each window or decrease the number of checked windows. To reduce the number of checked windows, a pre-filtering approach is a good choice. However, it is time consuming for the pre-processing procedure. To remedy the problem, an Haar-like feature-based detector proposed by Viola and Jones [8] is a cascaded face verifier for real time frontal face detection (24 frames per second).

The key advantage of a Haar-like feature over most other features is its calculation speed. Due to the use of integral images, a Haar-like feature of any size can be calculated in constant time.

With the development of cloud computing [9–12], internet of things [13–15], and Big Data [16, 17], the scheme for the cooperative gun-dome system composed of one gun camera and one dome camera can extend to the multi-camera network.

3 Experiments

The image dataset used for the face detection experiment was composed of two types of images, one image set including 151 facial images and the other including 151 non-face images. The image set was composed of the old, young, women and men face images and exists slightly illumination and orientation change. In the non-face image set, there are various kinds of furniture, buildings and natural scenes. 70 facial images and non-facial images are randomly selected from the two image sets for training respectively, and the left facial images and non-facial images were utilized to verify the performances of the CNN-based face detection strategy. The CNN that is used in present experiment consists of 7 layers, which is similar



Fig. 4 The face image features extracted by the CNN

with the CNN. Layer C1 extracted 16 features from input images by using a 5×5 kernel in each feature map. After subsampling, by using a 4×4 kernel the layer C3 produces 16 features, layer C5 produces 120 features.

The training procedure of CNN is similar to the procedure of training BP neural network, which consists of two steps [18].

1. Step 1: Feed Forward
“Feed forward” describes the process of putting the input samples into the network, and then calculating the output.
2. Step 2: Back Propagation

The adopted approach is based on CNN-based deep learning mechanism. The experiment used 8 kernels in layer C1, 16 kernels in layer C3, and 120 kernels in layer C5 to extract image features. Figure 4 shows the features extracted by the CNN on the face image dataset. Different from other feature extraction methods, the features extracted by CNN do not have the intuitive meanings, they characterize the object with the intrinsic feature elaborately and almost impossible can be described by certain regulation.

The CNN converged after the 7th iteration, the final accuracy rate at the 10th iteration over the test dataset is 100 %.

To verify the effective of the proposed method for face detection, we utilized it on real-time video stream and face image database respectively. For the simplicity, we mount the static analog cameras on the tripod to test the effectiveness of the proposed methods. Further study shows the proposed technique also has good performance on the IP camera and other types of surveillance cameras.

4 Conclusions

In the proposed scheme of face acquisition in surveillance scene, there are one wide angle camera and one Pan Tilt Zoom (PTZ) dome machine. The wide angle camera is responsible for the target detection in wide field of view, and PTZ dome machine (also known as active camera) for focusing and amplifying and tracking continuously for the target of attention. It provides much better performance than the single

dom camera. DPM and Look-up Table are used in this system. Furthermore, the CNN is utilized to execute the face detection and orientation detection. The experiments prove the effectiveness and efficiency of the proposed scheme.

Acknowledgments This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, 2013AA014603, in part by National Key Technology Support Program under Grant 2012BAH07B01, in part by the National Science Foundation of China under Grant 61300202, 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009, in part by the China Postdoctoral Science Foundation under Grant 2014M560085, and in part by the Science Foundation of Shanghai under Grant 13ZR1452900.

References

1. Felzenszwalb P, McAllester D, Ramanan D (2008) A discriminatively trained, multiscale, deformable part model. In: Computer vision and pattern recognition. IEEE, Anchorage, AK, pp 1–8
2. Cho H et al (2012) Real-time Pedestrian detection with deformable part models. In: Intelligent vehicles symposium. IEEE, Alcalá de Henares, Spain, pp 1035–1042
3. Beriault S (2008) Multi-camera system design, calibration and 3D reconstruction for markerless motion capture. In: School of information technology and engineering, Engineering University of Ottawa, p 146
4. Dong R, Li B, Chen Q-M (2009) An automatic calibration method for PTZ camera in expressway monitoring system. In: WRI world congress on computer science and information engineering, pp 636–640
5. Li H, Shen C (2006) An LMI approach for reliable PTZ camera self-calibration. In: International conference on advanced video and signal based surveillance (AVSS'06), IEEE
6. Wang Y (2013) Distributed multi-object tracking with multi-camera systems composed of overlapping and non-overlapping cameras. In: Graduate College, University of Nebraska, p 183
7. Hao Z, Zhang X, Yu P (2010) Video object tracing based on particle filter with ant colony optimization. In: The 2nd IEEE international conference on advanced computer control, Shenyang, China, pp 232–236
8. Kim J-M, Kim K-H, Song M-K (2009) Real time face detection and recognition using rectangular feature based classifier and modified matching algorithm. In: Fifth international conference on natural computation, Korea, pp 171–175
9. Liu Y, Ni Lionel M, Hu C (2012) A generalized probabilistic topology control for wireless sensor networks. *IEEE J Sel Areas Commun* 30(9):1780–1788
10. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
11. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Topics Comput* 2(3):376–387
12. Liu Y, Zhu Y, Ni LM, Xue G (2011) A reliability-oriented transmission service in wireless sensor networks. *IEEE Trans Parallel Distrib Syst* 22(12):2100–2107
13. Liu Y, Zhang Q, Ni LM (2010) Opportunity-based topology control in wireless sensor networks. *IEEE Trans Parallel Distrib Syst* 21(3):405–416
14. Liu X, Yang Y, Yuan D, Chen J (2013) Do we need to handle every temporal violation in scientific workflow systems. *ACM Trans Softw Eng Methodol*

15. Wang L, Tao J et al (2013) G-Hadoop: Mapreduce across distributed data centers for data-intensive computing. *Future Gener Comput Syst* 29(3):739–750
16. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50
17. Xu Z, Luo X, Zhang S, Wei X, Mei L, Hu C (2014) Mining temporal explicit and implicit semantic relations between entities using web search engines. *Future Gener Comput Syst* 37:468–477
18. Xie D, Dang L, Tong R (2012) Video based head detection and tracking surveillance system. In: International conference on fuzzy systems and knowledge discovery (FSKD). IEEE, Sichuan, pp 2832–2836

Vehicle Color Recognition Based on CUDA Acceleration

Zhiwei Tang, Yong Chen, Bin Li and Liangyi Li

Abstract The processing efficiency of Intelligent Transportation System has become an issue which is attracting more and more attention in order to combat vehicle crimes. The vehicle color recognition plays an important role in identifying, searching, improving and enhancing vehicle Intelligent Transportation System. However, area identification or location and surface high light detection are needed before vehicle color recognition. Convolution neural network training algorithm is taken as the best choice to conduct vehicle color recognition for its parameters are less and computation speed is faster. Neural network algorithm is of high parallelism, and its algorithm can even be further optimized if CUDA acceleration is selected. This research results have reference value for improving the processing efficiency of Intelligent Transportation System.

Keywords Vehiclecolor recognition · Convolutional neural network · GPU acceleration

1 Introduction

The color recognition is characteristic with its little dependent on object size, direction and angle of observation, robustness, and etc. It can effectively make up for the information needed in license plate recognition, which can be easily influenced if drivers intentionally keep out their plates, use fake plates or different plates at a single ride. It also can further improve the reliability of the Intelligent Transportation System.

Z. Tang · Y. Chen (✉)

The Third Research Institute of Ministry of Public Security, 339 Bisheng Road, Shanghai, China

e-mail: yangxuesansuo@126.com

B. Li · L. Li

Shanghai University, 99 Shangda Road, Shanghai, China

e-mail: 99chaoyang@163.com

Some of the present vehicle color recognition algorithms use color based method. Literature [1] is a comparative research on the color recognition with color difference formula based on multiple color spaces. The literature has found color space model and its corresponding color difference formula on color recognition. However, the truth is that there is no ideal uniform color space and color based methods still cannot meet the visual characteristics of human eyes, so it is necessary to develop vehicle color recognition method based on neural network algorithm.

This paper first introduces area identification or location and highlight detection needed before vehicle color recognition, and then it focuses on the convolution neural network algorithm and the ways to accelerate CUDA speed. Convolution neural network is an advanced version of BP neural network, and it is averagely superior than any other neural networks; its input image keeps high logical topology with the network; and feature extraction and pattern classification can be performed at the same time, even during the training process. At last, the algorithm accelerated is applied to the vehicle color recognition system. It has found that this algorithm can improve the efficiency of color recognition and bring more abundant outcomes. So it has great significance for the improvement of the efficiency of vehicle detection.

2 Positioning of Identifiable Area and Highlight Inspection

Before doing identification on vehicle color, positioning of identifiable area and highlight inspection on vehicle surface are the necessary steps. Because of factor influence such as vehicle window and license plate etc., there will be trouble to take the whole car as identifiable area for color identification; the partial area that can completely represent vehicle color is the best choice. At present, the methods adopted by literatures on positioning of identifiable area are mainly difference method, self-adaptive background cancellation method [2], differentiation [3] etc., due to existence of impulse noise in the image, it will affect positioning of color identifiable area, so we need to firstly make filtration on color image and then eliminate disturbance by noise. This paper adopts part in front of vehicle adjacent to exhaust fan as the identifiable area on vehicle color, and then use a series of technologies procession to position the identifiable area.

Vehicle surface will be subjected to disturbance by house, tree and especially for highlight of characterization source color in the environment to eliminate the highlight on surface of vehicle is one important problem on identification of vehicle color. Literature 9 pointed out that in Best Fit Window, diffuse pixels cluster is obtained by using K-means algorithm in the five dimensions feature vector space: CIE L^*a^*b information and pixel location information. And then the highlight cluster is detected. Literature 10 puts forward an algorithm which calculates the

surface highlight of vehicles according to one single colored picture. This algorithm effectively neglects the influence of high light and provides better real time performance and high accuracy of vehicle color recognition.

3 Convolutional Neural Networks Algorithm

The existing identification methods on vehicle color are k nearest neighbor algorithm, difference method, artificial neural network and support vector machine etc. This paper mainly introduces convolutional neural network algorithm. Artificial neural network is also regarded as neural network or connection model; it is one kind of algorithm mathematic model of simulating neural network of animal as characteristic and making distributed parallel processing. The parallel coordinated processing of neural network and distributed storage principle of plenty of nerve centers, the simulation ability of human cognitive system and high-efficient learning algorithm make it suitable to solve model identification problem. Let us recall BP neural network, each layer node is the linear and one-dimensional array state, the network n ode between layers are completely connected. If the middle payer of BP network and the nods between layers are changed as partial connection, it will become the simplest one-dimensional convolutional network. Expand this idea to two-dimensional, which will be convolutional neural network, which reduces the complication of network model. It will become much more obvious when inputting multiple-dimensional image, it can enable image to be directly as network input and has become the research focus on current speech analysis and image identification area. The general network structure is indicated by Fig. 1.

In the structure diagram of neural network, layer C is the network layer composed of nerve center in convolutional layer, layer S is the network layer composed of nerve center in sub-sampling. In the convolutional layer, fold the characteristic

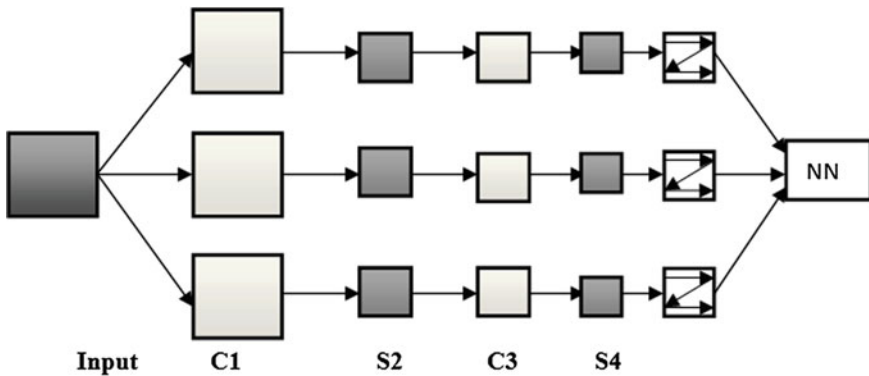


Fig. 1 Structure diagram of convolutional neural network

diagram of the front layer with one core that can be learned and then forms the characteristic diagram of this layer by activation function output. The general expression is as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (1)$$

Of which, $Sa(\cdot)$ indicates sub-sampling function, it is the area summation on input image in this layer of $n \times n$ size. Each output characteristic diagram has its own β and b .

There will be one sampling layer behind the convolutional layer to reduce calculation time and establish invariance between space and structure. The characteristic diagram after sampling will become smaller, when inputting the n pieces of characteristic diagram, the size of output image is $1/n$ of that on input; its general expression is indicated as follows:

$$x_j^l = f(\beta_j^l Sa(x_j^{l-1}) + b_j^l) \quad (2)$$

Of which, $Sa(\cdot)$ indicates sub-sampling function, it is the area summation on input image in this layer of $n \times n$ size. Each output characteristic diagram has its own β and b .

The input image together with 3 trainable filters and additive bias to make convolution, it will produce 3 characteristic mapped picture, and then makes summation, weighted value and biased for 4 elements of one group in the characteristic mapped picture, it gets characteristic picture of 3 S2 layer by Sigmoid function. These mapped pictures get C3 layer by filtration. This layer structure and S2 produce S4 likewise. Finally, these elements value will be rasterized and connect into one vector to the traditional neural network and gets output.

The traditional neural network can adopt BP neural network model, it mainly makes characteristic classification, and the number of nerve center in the final output layer represents classification number. Its learning regulation is to adopt steepest descent method and gradually adjust weight and threshold value by counter-propagation.

4 Identification Systems on Vehicle Color

Identification on vehicle color is the typical model identification problem, the traditional statistics model classification can classification capability have theoretical guarantee when the sample quantity is enough. On collecting plenty of color data of vehicle it establishes data set used for training and test, through the noise filtration, positioning of identifiable area, highlight inspection and convolutional neural networks algorithm after CUDA acceleration introduced by the above paper, the identification system of the whole vehicle can be presented in front of us.

Table 1 The compared result of the papers and the proposal algorithm

Algorithm	Accuracy (%)	Error rate (%)	FRR (%)
K nearest neighbor algorithm	70.52	25.18	4.30
HIS color difference identification method	75.37	21.51	3.12
CIE lab color difference identification method	83.64	12.19	4.17
The identification algorithm of this paper	98.68	1.32	0.0

Convolutional neural network algorithm has one learning process and it can be tested afterwards. So identification system of vehicle color is divided into 2 modules: training module and test module. The training module obtains vehicle image by sensor and extracts RGB color value of vehicle in the image, and then makes training on training sample storage. The expansion structure of convolutional neural network is very simple, but it needs plenty of calculation amounts, so the training process uses CUDA to make acceleration. The test process obtains vehicle image through sensor after convolutional neural network training is completed, make positioning on color identifiable area and test highlight element, extract RGB color value in vehicle, finally it uses result of training module to guide this kind of color classification to the designated classification.

If we have composition part of vehicle color identification system, we can make test on the overall capability of system, and use a series of time spent on operation and accuracy of color identification as judgment. We respectively use the identification method and color difference identification method of this paper to make test on vehicle image, the test result is indicated by Table 1.

We can see that the identification accuracy of this paper is higher than that of color difference method, and its FRR is 0.0 %.

5 Conclusion

This paper discusses and studies on convolution neural network based on CUDA acceleration used for vehicle color recognition. Convolution neural network algorithm has the advantages of simple structure, stronger adaptability, and faster recognition speed and other characteristics, so it can be widely used in the field of Intelligent Transportation System. At the same time, the results show that the computation speed of convolution neural network algorithm based on CUDA acceleration is about 6 times faster than CPU computation speed in the aspect of training. This also helps further improve the efficiency of color recognition. The implementation of convolution neural network algorithm based on accelerated CUDA has a very significant research meaning, especially when the traffic is heavy, and a higher detection rate of vehicle color recognition is needed.

References

1. Li G, Liu Z, You Z, Zhuang Y (2004) Car-body color recognition algorithm based on color difference and color normalization. *Comput Appl* (In Chinese)
2. Jin-Feng L, Lei G (2011) A forward propagation implementation of neural network on GPU. *Microcomput ITS Appl* 30(18) (In Chinese)
3. Szarvas M, Yoshizawa A, Yamamoto M, Ogata J (2005) Pedestrian detection with convolutional neural networks. In: *IEEE intelligent vehicles symposium proceedings, USA*, pp 224–229

Video Retargeting for Intelligent Sensing of Surveillance Devices

Huan Du, Zheng Xu and Zhiguo Yan

Abstract This paper proposes a video retargeting method for intelligent sensing of surveillance devices. The spatiotemporal saliency model is firstly built by evaluating the contrasts between the global histograms and each regional histogram. Based on the spatiotemporal saliency map, a salient object detection method is used to locate salient object regions in the video. Then the size of cropping window is evaluated based on the moving objects. Finally cropping and scaling operations are performed on the basis of salient object regions to generate the retargeted video.

Keywords Spatiotemporal salient map · Salient object detection · Cropping · Uniform scaling

1 Introduction

Video surveillance is an integrated system with strong prevention capabilities and widely used in military, customs, police, fire fighting, airports, railways, urban transport and many other public places. It's an important part of security system because of its visualized, accurate, timely and rich information content. Recently, content-aware video retargeting methods can be classified into four classes: cropping [1–4], warping [5–8], seam carving [9–14] and multi-operator approach [15–23]. The latter three are usually extended from image to video by constraining temporal relativity in video sequence. However, computational load and geometric distortion usually hinder their widely applications. As a distortion-free method, cropping method finds a cropping window with minimum information loss in the original frame by a variety of techniques, e.g. scanning [1], back-tracing [2], max-flow min-cut [3] and so on, and then cuts out the window region as the

H. Du · Z. Xu (✉) · Z. Yan

The Third Research Institute of Ministry of Public Security, Shanghai, China
e-mail: xuzheng@shu.edu.cn

Z. Xu

Tsinghua University, Beijing, China

retargeted frame. Nevertheless, these methods neither do maintain visual consistency along temporal axis, nor can ensure the integrity of important objects. Take the back-tracing method presented by Deselaers et al. [2] for example. It dynamically determines the trace of crop pane and generates a retargeted video with frame consistency. But this method seriously depends on the initial parameter of crop pane estimated from the initial frame, and can not crop salient objects whose positions have large displacement from the first frame.

In this paper, we propose a video retargeting method based on a spatiotemporal saliency model. The main contribution of our approach is threefold. Firstly, the contrasts between the global histograms and each regional histogram are evaluated, and then the spatiotemporal saliency model is built. Secondly, the size of cropping window is evaluated based on the moving objects. Thirdly, cropping and scaling operations are performed on the basis of salient object regions to generate the retargeted video. Due to the above mentioned three characteristics, our approach can preserve important objects well with a high retargeting performance.

2 Basic Model

To retarget a video Vo , a spatial histogram and a temporal histogram of the video frame are calculated respectively. Subsequently, through measuring contrast of histograms, a spatiotemporal model is built to highlight salient objects in a saliency map [22].

The spatial saliency map for the frame image is:

$$\begin{cases} S_{map}(p) = RD(H_k^c), & p \in R_k \\ RD(H_k^c) = \sum_{j=1}^M \left[H_k^c(j) \sum_{i=1}^M \|c_j - c_i\| \cdot H_0^c(i) \right] \end{cases} \quad (1)$$

where H_k^c denotes the regional color histogram of the region R_k , pixel p belongs to region R_k . $H_k^c(j)$ denotes the probability of color c_j in the regional color histogram H_k^c , $H_0^c(i)$ is the probability of color c_i in the global color histogram H_0^c .

The temporal saliency map for the frame image is:

$$\begin{cases} T_{map}(p) = RD(H_k^m), & p \in R_k \\ RD(H_k^m) = \sum_{j=1}^N \left[H_k^m(j) \sum_{i=1}^N \left| \vec{m}_j - \vec{m}_i \right| \cdot H_0^m(i) \right] \end{cases} \quad (2)$$

where H_k^m denotes the regional motion histogram of the region R_k , pixel p belongs to region R_k . $H_k^m(j)$ denotes the probability of motion vector \vec{m}_j in the regional motion histogram H_k^m , $H_0^m(i)$ denotes the probability of motion vector \vec{m}_i in the global motion histogram H_0^m .

Based on the spatial saliency map and temporal saliency map, the spatiotemporal saliency map is defined as their product to highlight both salient regions:

$$ST_{map}(p) = S_{map}(p) \cdot T_{map}(p) \quad (3)$$

3 The Proposed Method

After obtain the smoothed spatiotemporal saliency map, the salient object detection method [24] is employed to locate salient objects in smoothed spatiotemporal saliency map and record center of the box. Center of the box in the n th frame is assumed as (x_n, y_n) . In order to ensure coherence between frames, central positions of frames need to be smoothed. The smoothed center of the box in the n th frame is defined as:

$$(\tilde{x}_n, \tilde{y}_n) = \left(\frac{x_{n-\lambda} + x_{n-\lambda+1} + \dots + x_n + \dots + x_{n+\lambda}}{2\lambda + 1}, \frac{y_{n-\lambda} + y_{n-\lambda+1} + \dots + y_n + \dots + y_{n+\lambda}}{2\lambda + 1} \right) \quad (4)$$

where λ denotes the smooth step.

In our past work [22], in order to retain information as much as possible we make the cropping window maximize. Actually, most people only pay their attention on moving objects in video. In order to highlight moving objects as far as possible, the size of the cropping window $W_c \times H_c$ is given as:

$$\begin{cases} W_c = \min\{\tau_t \cdot h_m, w_p\}, H_c = h_m, & \text{if } \frac{W_m}{h_m} \leq \tau_t \\ W_c = w_m, H_c = \min\{w_m/\tau_t, h_p\}, & \text{else} \end{cases} \quad (5)$$

$$\begin{cases} w_p = W_o, h_p = W_o/\tau_t, & \text{if } \frac{W_o}{H_o} \leq \tau_t \\ w_p = \tau_t \cdot H_o, h_p = H_o, & \text{else} \end{cases} \quad (6)$$

where $\tau_t = W_t/H_t$ denotes the aspect ratio of the target video size. (w_p, h_p) denotes the intersection point P that the target aspect ratio intersects with the height/width of original video. w_m and h_m respectively denotes the longest width and the longest height in all detection boxes. We mark (w_m, h_m) as the longest width & height point M .

As shown in Fig. 1, the blue point O denotes the size of original video. The solid line denotes the aspect ratio of the target video. The purple point M and the orange point M' denote two cases of the positional relation between the longest width & height point and the target aspect ratio line. The green point C and the red point C' respectively denotes the size of the cropping window in these two cases.

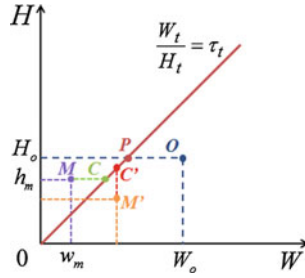


Fig. 1 Schematic of confirming the size of the cropping window

After confirming the size of the cropping window, we use the smoothed center of the box as the center of the cropping window and cut out salient object regions. Since the cropping video has the same aspect ratio with the target video, the uniform scaling is finally performed to generate the target video V_t (See Fig. 2).

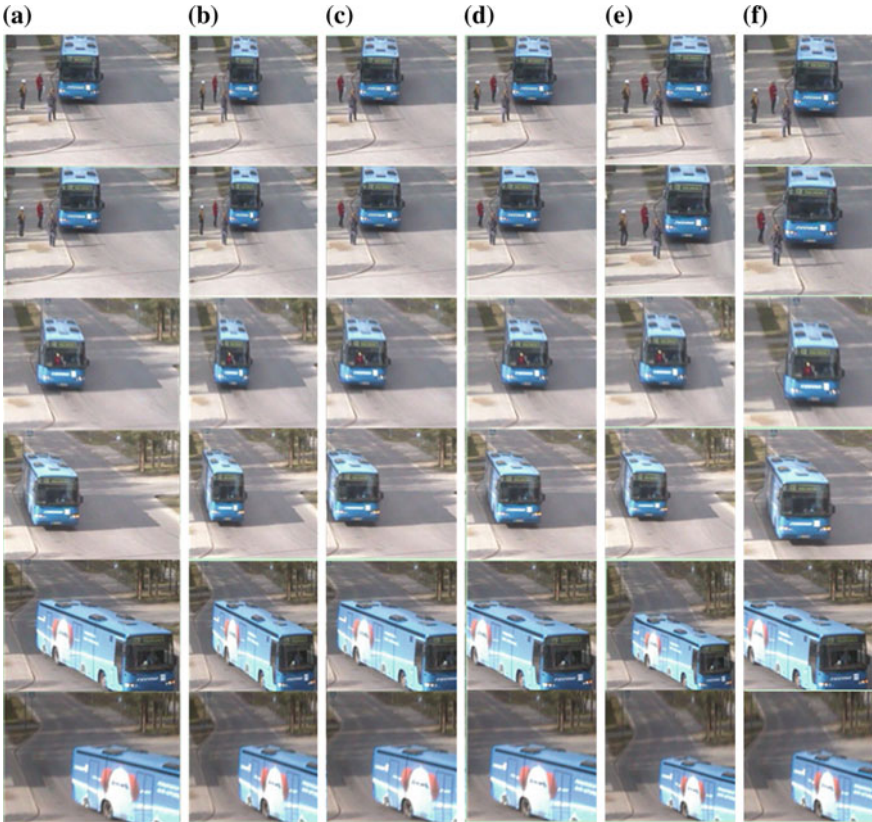


Fig. 2 Retargeting an outdoor surveillance video from 320×240 to 240×240 . a Original video thumbnails, b Scaling, c Cropping, d VSR, e VSRB, f Our approach

4 Experimental Results

We implement our video retargeting approach using C++, and evaluate the performance using a variety of videos including animations, news, movies, etc. We compare our approach with scaling, cropping, VRS [22] and VRSB [23]. As a simple operation, scaling can retain the most complete information, but salient objects may have deformation due to the original video and target video has different aspect ratio. Cropping can prevent salient objects from noticeable distortions, but can not guarantee there are complete salient objects in target video when salient objects are in the boundary of the video scene. Compared with the above two approaches, results demonstrate the better retargeting performance of VRS. But some objects in VSR are not highlighted enough because of its max cropping box. VSRB not only can highlight the important objects of video but also can preserve the completeness of the video content. However, some objects may have slight distortion such as cattle, due to its block partition sometimes have some deviation. In contrast, our approach preserves the aspect ratio and highlights objects in the optimal cropping window, which shift smoothly without jitter effects over the whole sequence.

5 Conclusions

The spatiotemporal saliency model is firstly built by evaluating the contrasts between the global histograms and each regional histogram. Based on the spatiotemporal saliency map, a salient object detection method is used to locate salient object regions in the video. Then the size of cropping window is evaluated based on the moving objects. Finally cropping and scaling operations are performed on the basis of salient object regions to generate the retargeted video.

Acknowledgments This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, 2013AA014603, in part by National Key Technology Support Program under Grant 2012BAH07B01, in part by the National Science Foundation of China under Grant 61300202, 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009, in part by the China Postdoctoral Science Foundation under Grant 2014M560085, and in part by the Science Foundation of Shanghai under Grant 13ZR1452900.

References

1. Liu F, Gleicher M (2006) Automating pan and scan. In: Proceedings of ACM multimedia, Santa Barbara, California, USA, pp 241–250
2. Deselaers T, Dreuw P, Ney H (2008) Pan, zoom, scan-time-coherent, trained automatic video cropping. In: Proceedings of IEEE CVPR, Anchorage, Alaska, USA, pp 1–8
3. Li Y, Tian Y, Yang J et al. (2010) Video retargeting with multi-scale trajectory optimization. In: Proceedings of multimedia information retrieval, Philadelphia, Pennsylvania, USA, pp 45–54
4. Xue YZ, Du H, Li Z et al (2011) Video retargeting using optimized crop-and-scale. *J Shanghai Univ* 15(4):331–334
5. Wolf L, Guttman M, Cohenor D (2007) Non-homogeneous content-driven video-retargeting. In: Proceedings of IEEE ICCV, Rio de Janeiro, Brazil, pp 4409010
6. Kim JS, Kim JH, Kim CS (2009) Adaptive image and video retargeting technique based on Fourier analysis. In: Proceedings of IEEE CVPR, Miami, Florida, USA, pp 1730–1737
7. Hu Y, Rajan D (2010) Hybrid shift map for video retargeting. In: Proceedings of IEEE CVPR, San Francisco, California, USA, pp 577–584
8. Wang YS, Fu HB, Sorkine O et al (2009) Motion-aware temporal coherence for video resizing. *ACM Trans Graph* 28(5):127
9. Rubinstein M, Shamir A, Avidan S (2008) Improved seam carving for video retargeting. *ACM Trans Graph* 27:16
10. Chao WL, Su HH, Chien SY et al. (2011) Coarse-to-fine temporal optimization for video retargeting based on seam carving. In: Proceedings of ICME, Barcelona, Spain, pp 1–6
11. Grundmann M, Kvatra V, Essa I (2010) Discontinuous seam-carving for video retargeting. In: Proceedings of IEEE CVPR, San Francisco, California, USA, pp 569–576
12. Du H, Liu Z, Xue YZ et al. (2011) Fast seam carving based on direction map. In: Proceedings of IET international communication conference on wireless mobile & computing, Shanghai, China, pp 70–75
13. Chiang CK, Wang SF, Chen YL et al (2009) Fast JND-based video carving with GPU acceleration for real-time video retargeting. *IEEE Trans Circuits Syst Video Technol* 19:1588–1597
14. Kopf S, Kiess J, Lemelson H et al. (2009) FSCAV: fast seam carving for size adaptation of videos. In: Proceedings of ACM multimedia, Beijing, China, pp 321–330
15. Rubinstein M, Shamir A, Avidan S (2009) Multi-operator media retargeting. *ACM Trans Graph* 28:23
16. Du H, Liu Z, Jiang JL et al (2013) Stretchability-aware block scaling for image retargeting. *J Vis Commun Image Represent* 24(4):499–508
17. Wang YS, Lin HC, Sorkine O et al (2010) Motion-based video retargeting with optimized crop-and-warp. *ACM Trans Graph* 29(4):90
18. Wang SF, Lai SH (2011) Compressibility-aware media retargeting with structure preserving. *IEEE Trans Image Process* 20(3):855–865
19. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 8(3):482–494
20. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Top Comput* 2(3):376–387
21. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50
22. Du H, Liu Z, Wang J, Mei L, He Y (2014) Video retargeting based on spatiotemporal saliency. In: Proceedings of FTRA, Zhangjiajie, China, pp 397–402
23. Du H, Liu Z, Yan Z, Jiang Y (2014) Video retargeting based on stretchability-aware blocks scaling. In: Proceedings of SKG, Beijing, China, pp 171–176
24. Shi R, Liu Z, Du H et al (2012) Region diversity maximization for salient object detection. *IEEE Signal Process Lett* 19(4):215–218

Web Knowledge Acquisition Model Based on Human Cognitive Process

Xiaobo Yin and Xiangfeng Luo

Abstract Web oriented knowledge acquisition has become the important way for people to acquire knowledge. How to help user accurately to obtain personalized knowledge in network, is an important problem for Web service. In this paper, we present a web knowledge acquisition model based on human cognitive process that can improve the precision of knowledge acquisition and meet the personalized requirements of users. The user cognitive model (UCM) designed in this paper can improve the interaction with web, and then promote the web services development of personalized, accurate and intelligent.

Keywords Knowledge acquisition · Cognitive process · User cognitive model · Personalized knowledge

1 Introduction

Knowledge acquisition is the process of extracting, structuring and organizing knowledge from expert, expert system, text, database, Internet or other knowledge source [1]. Currently, web information search model has become the most important way for people to acquire knowledge from internet; many researchers have contributed on method, model of knowledge acquisition. Many related research works have been done, such as personalized search model [2–4], cognitive information retrieval model [5, 6] and interactive information retrieval model [7, 8] etc. However, above web knowledge acquisition models have many disadvantages:

X. Yin (✉) · X. Luo

School of Computer Engineering and Science, Shanghai University, Shanghai, China
e-mail: xbyin@aust.edu.cn

X. Luo

e-mail: xfluo@shu.edu.cn

X. Yin

College of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, China

(1) without use model, some important context may be lost in the process of interaction. (2) When facing more complex search query tasks, search engine can not provide user enough help. In another words, the navigation ability of search engine is too weak.

In order to help user obtain personalized knowledge through web service, we presents a web knowledge acquisition model based on human cognitive process. First, we build a cognitive process model of web knowledge acquisition, by imitating human cognitive process. Second, we design user cognitive model (UCM), according to the interactive computing theory. Third, we present a knowledge discovery method based on clue. The method we present in this paper, can improve the precision of knowledge acquisition and provide user the personalized help. UCM can improve the interaction with web, and then promote the web services development of personalized, accurate and intelligent.

2 Web Knowledge Acquisition Cognitive Process Model

Based on human cognitive process, we map human learning process to web service. Then, we model the whole web knowledge acquisition cognitive process framework as shown in Fig. 1.

Human cognitive process and UCM cognitive process have several similar cognitive activities, such as perception, interpretation, understanding, evaluation,

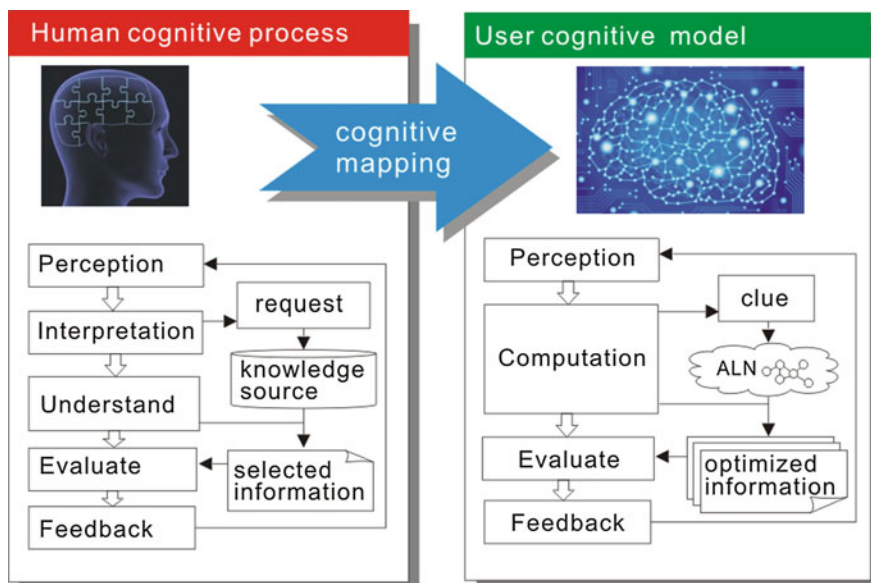


Fig. 1 Web knowledge acquisition cognitive process model

feedback [9, 10]. In UCM cognitive process, we use clue correspond to user request, and use Associated Link Network (ALN) [11] correspond to knowledge source. ALN is developed to re-organize the web resources based on association relations, it would let us be able to extract results exceeding just using keyword matching method.

3 UCM

According to the interactive computing theory, we design the UCM cognitive process between user and web. As shown in Fig. 2, UCM is an intermediate link of interactive process between user and web. On the one hand, UCM can help the user to obtain personalized knowledge, on the other hand, UCM can help ALN network recombine knowledge structure.

UCM cognitive process includes seven steps as follow: (1) Enter clue. (2) Obtain topic from ALN. (3) Calculate clue evolution. (4) Calculate interestingness. (5) Optimize topic ranking. (6) Knowledge extraction. (7) Feedback.

A Web knowledge acquisition process begins in the user’s query. A clue which is the content of query, should be sent from UCM to ALN. Then, UCM gets a set of ranked topics form ALN. Next, UCM will calculate the degree of clue evolution and the user interestingness to obtained topic. Next, UCM should return optimized topics to user. Next, user will extract knowledge according to optimized topics and his own priori knowledge. Last, UCM will feedback new knowledge structure both to user and ALN.

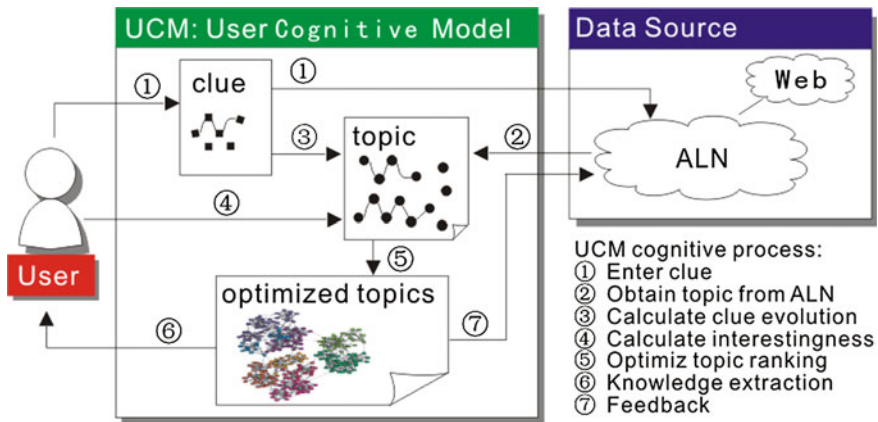


Fig. 2 UCM: user cognitive model

4 Knowledge Discovery Method Based on Clue

Along the UCM cognitive process, we proposed a knowledge discovery method based on clue. The detail step of this **method** will be described as follows.

(1) **Enter Clue**

UCM should receive several clues in this stage. Clues can be the keywords, sentences or the combination of keywords and sentences. The set of clues C will be defined as

$$C = \{c_1, c_2, \dots, c_i, \dots, c_m\} \quad (1)$$

In Eq. (1), c_i is a clue and m is the number of clues.

(2) **Obtain Topic from ALN**

In the stage of obtaining topic, UCM enter a clue and obtain a set of ranked topics from the results returned by Association Link Network (ALN), which can be defined as

$$T = Q(C) = \{t_1, t_2, \dots, t_j, \dots, t_n\} \quad (2)$$

In Eq. (2), $Q(C)$ is an abstract function denoting the topic obtaining, T is a set of topic t_j , and n is the number of topics.

(3) **Calculate Clue Evolution**

We can calculate the degree of clue evolution. First, we extract the keywords of clues (W_c) and the keywords of obtained topics (W_t). Next, we use web searching engine to search W_c and W_t in ALN. Then, the degree of clue evolution (E_c) can be measured as follows:

$$E_c = \log_2 \frac{P(W_c, W_t)^2}{P(W_c) \cdot P(W_t)} \quad (3)$$

In formula (3), $P(W_c)$ is the total number of searched pages for W_c , $P(W_t)$ is the total number of searched pages for W_t , $P(W_c, W_t)$ is the total number of searched pages for both W_c and W_t .

(4) **Calculate Interestingness**

The interestingness of obtained topic (I_t) can be calculated by user behavior and choice. This process is described as follows:

$$I_t = INTEREST(T) \quad (4)$$

(5) **Optimize Topic Ranking**

Which topic will be confirmed as optimized topic? According to E_c and I_t , we measure the weight of the obtained topics (WT) as follows:

$$WT = WEIGHT(Ec, It, \beta) \tag{5}$$

In formula (5), β is the weighting coefficient which can adjust the weight of Ec and It .

Then, topics need to be re-ranked according to the WT . The process of this step is described as follows:

$$T' = OPITIMIZE(T, WT) = \{t'_1, t'_2, \dots, t'_i, \dots, t'_n\} \tag{6}$$

In Eq. (6), T' is a set of optimized topic t'_i ordered by the topic's weight WT , n is the number of topics.

(6) **Knowledge Extraction**

Knowledge extraction cannot do without user's priori knowledge Kp . Knowledge extraction process *EXTRACTION* can be expressed as follows:

$$Ks = EXTRACTION(T', Kp) = \{k_1, k_2, \dots, k_j, \dots, k_m\} \tag{7}$$

In Eq. (7), Ks is the set of knowledge k_j which can be extracted form optimized topic set T' .

(7) **Feedback**

Feedback can finally help the user to obtain personalized knowledge, the process. can be expressed as

$$K[S] + Ks \rightarrow K[S + K_L] \tag{8}$$

In Eq. (8), $K[S]$ is user knowledge structure, Ks is the extracted knowledge which denote the knowledge of short-term memory, K_L is the knowledge of long-term memory.

In the meanwhile, feedback can also help ALN network recombine knowledge structure.

5 Conclusions and Future Work

In this paper, we present a web knowledge acquisition model based on human cognitive process that can improve the precision of knowledge acquisition and meet the personalized requirements of users. The major contributions of our work include three aspects: (1) We build a cognitive process model of web knowledge acquisition, according to human cognitive process. (2) We design UCM, according to the interactive computing theory. (3) We present a knowledge discovery method based on clue. The research work can improve the interaction between user and web, and then promote the web services development of personalized, accurate and intelligent.

In future work, we will further study the clue evolution, topic optimization algorithm, and knowledge acquisition precision evaluation.

Acknowledgements This Research work reported in this paper is supported by the Anhui province science and Technology Agency Natural Science Foundation key project (project no. 1308085MF94) and by the China National Natural Science Found project (project no. 69010815). We thank some teachers and students for their precious proposal.

References

1. Knowledge acquisition (2015) Wikipedia. http://en.wikipedia.org/wiki/Knowledge_acquisition
2. Gauch S, Chaffee J, Pretschner A (2003) Ontology-based personalized search and browsing. *Web Intell Agent Syst*
3. Ravindran D, Gauch S (2004) Exploiting hierarchical relationships in conceptual search. In: *Proceedings of the 13th international conference on information and knowledge management*, Washington, D.C.
4. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. *IEEE Trans Emerg Topics Comput* 2(3):376–387
5. Wang P, Soergel D (1998) A cognitive model of document use during a research project: study I: document selection. *J Am Soc Inf Sci* 49(2):115–133
6. Spink A, Cole C (2005) *New directions in cognitive information retrieval*. Springer, Berlin
7. Belkin NJ (1996) Intelligent information retrieval: whose intelligence? In: *ISI '96: proceedings of the fifth international symposium for information science*. Konstanz: Universtaetsverlag Konstanz, pp 25–31
8. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. *Future Gener Comput Syst* 43–44:40–50
9. Sutcliffe A, Ennis M (1998) Towards a cognitive theory of information retrieval. *Interact Comput* 321–351
10. Wechsler K, Baier J, Nussbaum M, Baeza-yates R (2004) Semantic search in the WWW supported by a cognitive model. conference: web-age information management—WAIM, 2004, pp 315–324
11. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. *IEEE Trans Autom Sci Eng* 482–494

An Investigation on the Relationship Among Employees' Job Stress, Satisfaction and Performance

Che-Chang Chang and Fang-Tzu Chen

Abstract With rapid economic development, enterprises start to much concern about employees' stress while working. The higher the job stress, the lower the job performances. The present research investigated the relationship among job stress, satisfaction, and performance by examining the employees' attitudes through collecting data from the industries of medical care, plastic spare parts, and automobile spare parts. The results indicated that employees' performance in working places was somehow related to their work stress and the extent of job satisfaction. Based on the results, suggestions were provided for managers for better employees' management.

Keywords Job stress · Job performance · Job satisfaction · Employee

1 Introduction

Recently, the issue of job stress has drawn lots of researchers' attention in their research area of business management. Organizations are aware of the fact that lots of human potentials are drained away due to job stress. Besides, numerous employees express that they experience extreme pressure and heavy workload caused by their jobs. That is, workers generally and extensively experience job-related stresses. Job stress might threaten working efficiency and cause serious problems in business development. Because the human resource managers in some organizations have noticed that stress is a great barrier in regard to the employees' working efficiency and task performances. Job stress—derived from overload, overtime, job insecurity, role conflicts, and so on—has become a serious and

C.-C. Chang (✉)

School of Management, Fujian University of Technology, Fuzhou, Fujian, China
e-mail: 1794298328@qq.com

F.-T. Chen

Department of International Business Management, Tainan University of Technology,
Tainan, Taiwan
e-mail: t20012@mail.tut.edu.tw

widespread aftermath, and negatively impacts employees' task performances. On the other hand, if the employees feel satisfied and enjoy in their job, it might counterbalance the negative impacts brought by job stresses. The present research will investigate employees' attitudes by examining the relationship among employees' job stress, job satisfaction, and working efficiency. The paper finally concludes with research implications and suggestions for managers in enterprises.

2 Literature Review

2.1 Job Stress

As showing on National Institute of Occupational Safety and Health (1999), job stress is defined as harmful physical and emotional responses that occur when job requirements do not match the worker's capabilities, resources, and needs. Job stress is generally recognized as a major challenge to threaten not only individual mental and physical health, but also organizational health. Stressed workers are more likely to be mentally or physically unhealthy, de-motivated, unproductive, and insecure while they are on duty. As a consequence, the organizations will be highly unlikely to be struggling in fiercely competitive markets. According to some analysis, job-related stress costs the national economy a staggering amount in sick pay, lost productivity, health care and litigation charges.

Cooper believes that stress is caused by a misfit between individuals and their environment. Stress is a dynamic state whereby people are confronted with anxious, uncertainty, or undesirable demands. As a result, it would lead to unstable, negative, and terrifying [1].

Recently researchers reported that stress is results of extensive work pressure, a lack of control over working situations, and unsatisfactory interpersonal relationships. It was found that one out of three people voiced that the more competitive the working places, they more stressful they feel, especially those in the high-technology companies.

There are varied factor would caused job pressure. These can be classified into organization and workers factors. Moreover, in this study, we classified the stresses into the following: job loading, an organization's atmosphere, worker's ability, job controlling, and role conflict.

2.2 Relationship Between Job Stress and Performance

In general, job stress is what the job has brought to individuals, and it threatens their health physically and psychologically. Therefore, what the researchers concern is not only employees' physical and psychological health, but also what human resource management should do to solve these problems. Nowadays more and more

managers tend to use task performance criteria to evaluate the employees' productivity. They urge workers by all means to work hard in order to reach the expected goals of the enterprises. What the managers do to their employees, undoubtedly, might positively increase employees' productivity; however, it might also make them feel extremely stressful. Many researchers have demonstrated that job stress is a critical factor against job performance. Job-related psychosocial stress is a multi-faceted phenomenon which may affect employees' attitudes and behaviors. Same as stresses, also have negative impacts on task performances [2].

2.3 Relationship Between Job Satisfaction and Performance

There have been quite a few definitions on job satisfaction in previous studies. For example, McNamara defined job satisfaction as "one's feelings or state of mind regarding the nature of their work". While Ostroff defined job satisfaction as employees' job attitude that directly tied to individual needs, which means that higher satisfaction leads to more efficient organizations.

Job performance can be regarded as an activity in which an individual can successfully accomplish the given task which is subjected to the normal constraints of reasonable utilization of the available resources [3]. Again, research shows that job satisfaction is a significantly positive factor to influence job performance.

3 Methodology

3.1 Research Framework

The questionnaire used in the present research investigated the background information of the participants by probing their job stress, satisfaction, and performance. Job stress and satisfaction were independent variables, and job performance was a dependent variable. The research framework is shown as Fig. 1.

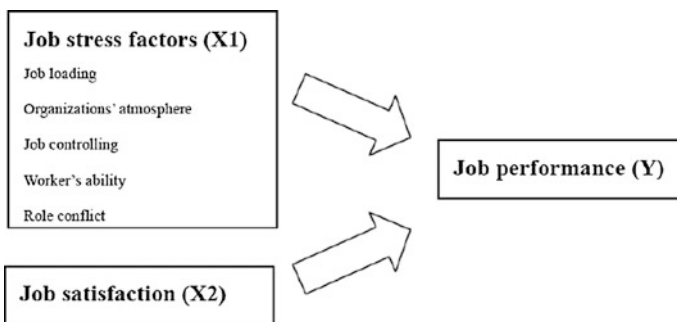


Fig. 1 Research design

3.2 Research Hypotheses

H₁: job stress will have a significant effect on job performance.

H₂: job satisfaction will have a significant effect on job performance.

3.3 Questionnaire

A questionnaire comprising 18 test items (Table 1) was designed with an attempt to examine the participants' background information and verify the research hypotheses. The questionnaire included four sections: job stress, satisfaction, performance, and personal information for each participant. SPSS (Statistical Package for the Social Sciences) was employed to examine the statistically significant effects between dependent and independent variables.

3.3.1 Factor Analysis

In the questionnaire, there were 18 test items in the category of job stress (Table 1). Based on factor analysis, all questions were divided into five components: job

Table 1 Questionnaire

Main category	Subcategory	Number of test items
Job stress	Job loading	4
	Organization's atmosphere	4
	Worker's ability	4
	Job controlling	4
	Role conflict	2
Job satisfaction		4
Job performance		14
Personal information		5

Table 2 Factor analysis of the questionnaire

Major factor	Sub-factor	Cronbach's α	Total variance explained (%)
Job stress	Job loading	0.752	65.52
	Organization's atmosphere	0.766	
	Worker's ability	0.743	
	Job controlling	0.690	
	Role conflict	0.728	
Job satisfaction		0.532	66.54
Job performance		0.844	60.61

loading, an organization’s atmosphere, worker’s ability, job controlling, and role conflict. Besides, four test items explored job satisfaction, and fourteen test items queried job performance.

According to Table 2, the total variance explained for each factor was respectively 65.52, 66.54, and 60.61 %. The Cronbach’s α for each factor exceeded 0.70. This indicates that the questionnaire has good construct validity and reliability.

4 Results Analysis

The results of the collected data were shown as Table 3. A total of 120 questionnaires were distributed and 95 were valid; the remaining invalid ones were excluded from data analysis. The distribution of the participants was as follows:

- X group: medical industries, out of 35 questionnaires, 33 were valid.
- Y group: plastic spare parts industries, out of 40 questionnaires, 38 were valid.
- Z group: automobile spare parts industries, out of 27 questionnaires, 24 were valid.

Table 3 Descriptive statistics of the questionnaire

Item		X	Y	Z	%
Gender	Man	12	22	12	48.4
	Woman	21	16	12	51.6
Age	21–25	11	11	10	33.7
	26–30	6	14	11	32.6
	31–35	8	8	1	17.9
	36–40	2	0	2	4.2
	41–45	5	4	0	9.5
	46 and above	1	1	0	2.1
Education	Junior high school	7	1	0	8.4
	Senior high school	10	3	3	16.8
	College	7	21	13	43.2
	University and above	9	13	8	31.6
Working experience	Below 1	2	3	2	7.4
	1–3 years	7	10	11	29.4
	3–5 years	7	9	7	24.2
	5–7 years	6	9	4	20.0
	7–9 years	5	4	0	9.5
	More than 9 years	6	3	0	9.5
Position	General workers	15	15	9	41.1
	Staff	5	6	10	22.1
	Technical people	2	11	2	15.7
	Managing people	6	6	3	15.8
	Sales person	5	0	0	5.3

To explain cause-effect relationships in the research model, a multiple linear regression was used. The multiple regression analysis was used to examine and predicate the linear relationship between two independent constructs and one dependent construct (i.e., job stress and job satisfaction to determine their job performance). According to Table 4, VIF = 1.116 < 10. This indicates that multi-collinearity doesn't exist between job stress and job satisfaction. The table also shows that the constant = 3.491 (t = 9.614); job stress = -0.192 (t = -2.735, *p* value = 0.007); job satisfaction = 0.255 (t = 3.598, *p* value = 0.001). This explains that job stress and satisfaction have significant effects on job performance. Its relationship can be shown as below:

$$\text{Job Performance}(Y) = 3.49 - 10.192 * \text{Job Stress}(X1) + 0.255 * \text{Job Satisfaction}(X2)$$

Besides, Table 5 shows the regression analysis among the sub-factors of job stress (job load, an organization's atmosphere, worker's ability, job controlling, and role conflict) and job satisfaction on the dependent construct. The results indicate that both factors have significant explanations (34.1 %) for job performance, *f* = 7.588, *p* = 0.000, and Adj . R2 = 0.296.

The results show that the job satisfaction, (*p* = 0.002, standardized coefficients = 0.303) and an organization's atmosphere (*p* = 0.006, standardized coefficients = -0.332) have significant effects on job performance. This means that the more satisfied the employees feel, and the better job performance they have. Besides, the lower the stress the employees experience from the organization's atmosphere, the better their job performance.

Moreover, employees' ability also plays a role to predict job performance (*p* = 0.012, standardized coefficient = -0.255). This shows that the more stress the employees experience from the job, the less productive they are. Similarly, in terms of job loading, the stress the employees experience is positively correlated with their job performance (*p* = 0.042, standardized coefficient = 0.232). In this research,

Table 4 Coefficients of job stress and job satisfaction towards Job performance

	未標準化係數		標準化係數	t	P值	共線性統計量	
	B 之估計值	標準誤差	Beta 分配			允差	VIF
(常數)	3.491	.363		9.614	.000***		
工作壓力(X1)	-.192	.070	-.262	-2.735	.007**	.896	1.116
工作滿意度 (X2)	.255	.071	.344	3.598	.001**	.896	1.116
R ² /AdjR ²				0.245/0.229			
F/p				14.948/0.000			

Table 5 Results of regression analysis among independent and dependent constructs

		Un-standardized coefficients		Standardized coefficients	t	Sig.	VIF
		B	Std. error	Beta			
(Constant)		3.577	0.354		10.102	0.000	
Job stress	Job loading	0.115	0.056	0.232	2.060	0.042*	1.698
	Organization atmosphere	-0.171	0.060	-0.332	-2.825	0.006**	1.844
	Worker's ability	-0.136	0.053	-0.255	-2.551	0.012*	1.335
	Job controlling	0.046	0.073	0.073	0.635	0.527	1.750
	Role conflict	-0.056	0.049	-0.123	-1.161	0.249	1.511
Job satisfaction		0.225	0.069	0.303	3.268	0.002**	1.146
R ² /Adj. R ²		0.341/0.296					
F/p		7.588/0.000***					

* $p < 0.05$, ** $p > 0.01$, *** $p > 0.001$

the factors of job satisfaction ($p = 0.002$), an organization's atmosphere ($p = 0.006$), workers' ability ($p = 0.012$) and job loading ($p = 0.042$) have significant effects on job performance; however, the factors of job controlling ($p = 0.527$) and role conflict ($p = 0.249$) showed non-significant effects on job performance.

5 Conclusions and Suggestions

5.1 Build up an Ideal Communication System

Based on the results of the investigation, an organization's atmosphere has significant effects on job performance. Thus, it is important for enterprises to build up an ideal communication system through formal or informal frameworks. This mechanism might encourage their employees to communicate with each other in order to reduce their pressure; and it can also create a friendly atmosphere in working places, increase employees' motivation, productivity, and task performances.

5.2 Encourage Cooperation and Empower Employees

Job loading is an important factor to affect job performance. The best way to solve loading problems is to empower the employees, encourage them to establish teams to conduct team works, in which team members can share tasks and exchange ideas.

Encouraging employees to build up their own teams and cooperate with each other can enrich their job and reduce their stress.

5.3 Set up Self-training Mechanisms

Finally, workers' inabilities will cause job stress and negatively influence their task performance. It is necessary for enterprises to set up a training mechanism. Particularly, to catch up with the rapid development of competitive markets, incorporating innovative technologies with employees-training might improve employees' abilities and facilitate their working efficiency.

References

1. Ruyter K, Wetzel M (2001) Role stress in call centers its effect on performance and satisfaction. *J Interact Mark* 15:23–30
2. Gilboa S, Shirom A, Fried Y, Cooper CA (2005) Meta-analysis of stress and performance at work: examining the moderating effect of gender, and and tenure. Working paper 6/2005, Henry Crown Institute of Business Research in Israel
3. Salami AO, Ojokuku RM, Ilesanmi OA (2010) Impact of job stress on managers' performance. *Eur J Sci Res* 45(2):249–260

Research on Influence Factors of the Formation of Virtual Innovation Clusters

Dong Qiu and Qiu-Ming Wu

Abstract With the development of science and technology and the increasingly fierce competition of innovation, the virtual innovation cluster (VIC) has become an important way of innovation. This paper studies the influence factors of the formation of VIC, in order that more VICs can be built well. A new model of factors is proposed, in which some new variables such as searching for partners, negotiation, internal recognition, external recognition and structure level are used to describe the situation of VIC's formation, and other new variables such as members' quality and relationship are increased into the influence factors. The empirical research shows that the factors of the government's promotion, members' quality, members' relationship and innovative spirit play an important role in the formation of VIC.

Keywords Innovation cluster · Virtual cluster · Formation · Influence factors

1 Introduction

The virtual innovation cluster (VIC) is a joint body composed of innovation subjects that are collaborating to achieve the common goals of innovation by modern communication technologies. Compared with the traditional innovation cluster, the most salient features of the VIC are “virtual”:

D. Qiu (✉)

School of Management, Fujian University of Technology,
Fuzhou, Fujian, China
e-mail: Qd613@126.com

D. Qiu · Q.-M. Wu

School of Economics and Management, Fuzhou University,
Fuzhou, Fujian, China
e-mail: Qiuming30@sina.com

1. Virtual organization. Innovation subjects maintained independence and freedom. They choose partners and sign contracts, according to their own free will.
2. Virtual innovation environment. With the help of the Internet and large-scale data processing technology, innovation subjects establish the virtual innovation environment, in which the efficiency of information sharing, collaborative innovation and interdisciplinary is greatly improved.
3. The inter regional cooperation. It is because of modern information and communication technology, the space limitation of innovation activities is dramatically reduced, and regional and global collaborations become easy.

In a word, the biggest advantage of the VIC is to efficiently use the distributed innovation resources. Therefore, in today's background of the continuous development of the science and technology and the increasingly fierce competition of innovation, VIC has become popular. In recent years, In China, a kind of VIC called the Collaborative Innovation Center (CIC) has become the most important model of innovation, which is developed from the model of the Engineering Research Center (ERC) in the USA. Because of its obvious advantages, the government hopes to develop more VICs.

The VIC's formation is very different from the traditional cluster:

1. The initial size of the VIC can be very small.
2. The VIC's formation process is faster than the traditional cluster, which needs a long time of development and accumulation.
3. It may not be easily found in the early stage of formation.

Therefore, this paper focuses on the initial forming stage of VIC. The paper theoretically and empirically researches on the factors influencing the formation of VIC, so that we can master the rules of VIC's formation and construct and develop VIC better.

2 Literature Review

2.1 Virtual Innovation Cluster

The concept of cluster dates back to 1890 as agglomeration economies put forward by Alfred Marshall. Then Porter [1] introduced and popularized the industrial clusters in *The Competitive Advantage of Nations*. Porter [2] claims that clusters have the potential to affect competition in three ways: by increasing the productivity of the companies in the cluster, by driving innovation in the field, and by stimulating new businesses in the field.

Geographic concentration is an important feature of traditional clusters. But With the development of economic globalization, enterprises are no longer limited to

local, but to find suitable partners in the global scope of product development and marketing. Innovation network between enterprises has become increasingly expanded and internationalized [3]. Passionate and Secundo [4] studied the nature and features of virtual cluster, which is considered as a combination of the participants whom having different core competitiveness, and sharing innovation cost and risk. Therefore, the virtual cluster is a new way to promote the traditional innovation system to expand to the global scope.

2.2 Formation of Cluster

There are two kinds of view of the forming process of cluster.

The first one is the theory of life cycle, which is the most popular. The life cycle of the industrial cluster is usually divided into four phases, including formation, development, maturity and decline [5, 6]. The key factor is different in each phase of the formation of the cluster. And in different ways of forming, the characteristics of the life cycle are also different.

The second one is the theory of the direction of the forming process. Jiang [7] divided the cluster formation path into two basic categories: top-down formation path, namely the government guides cluster formation; bottom-up formation path, namely the enterprises are main force to guide the cluster formation.

The two kinds of view above is of the system, logic and experience, and has positive significance for people to understand the formation mechanism of cluster. However, the disadvantage of both views is to focus on the macro level. Therefore, this paper will explore the formation of clusters further from the cognitive and behavioral perspective.

2.3 Influence Factors of Cluster's Formation

Previous studies show that the factors mainly include two aspects: one is an external factor, such as the market factor, the technical factor, the government, the fund, the pressure of survival, the competition effect, innovation culture and so on; the other is the internal factor, such as interest driven, internal incentive, imitation effect [8].

At the same time, some scholars pay attention to, in addition to the external and internal factors, some interaction factors between innovation subjects that will drive the formation of cooperation, union or cluster, such as the strategic synergy, complementary technology and technology difference [9].

3 Methodology

3.1 Analysis of VIC's Characteristics

The forming stages of VIC are defined as follows:

1. The initial cluster should have a certain scale, which means the cluster should include a certain number of members and certain qualities. The quantity and quality of members is an important index for the development of evaluation of traditional clusters, and is also suitable for the platform cluster.
2. Cluster members should be internally recognized by team members, and also externally recognized by the outsiders. It can be said, those who have not yet received internal or external recognitions are not qualified to be the cluster.
3. The cluster should have a formal contract to stipulate the provisions of the organizational structure, membership, decision-making mechanism, collaboration rules etc.

In short, those who possess the above characteristics are qualified to form clusters. The formation of clusters does not mean the maturity of clusters. After forming clusters, they also have to develop, and experience evolution process. The above features for clusters form a clear boundary between the forming stage and the collaborative development stage.

3.2 Factor Analysis

1. Internal factors

Schumpeter believed that innovation impetus derive from the pursuit of excess profits and entrepreneurship; and the entrepreneur spirit contains pioneering spirit, the desire to be successful, dedication and adventure spirit. From here, Vic formation can be attributed to two aspects: one is the pursuit of innovation itself, and the other is the pursuit of interest.

2. External factors

Starting from the external environment framework of PEST, the external factors for the formation of VIC mainly from four aspects: promotion of science and technology, market drives, government drives and social drives.

3. Interaction factors

Wu puts forward the internal force model of social system elements. (1) The integration forces and comprehensive quality of two elements should be positively correlated. That is the higher the comprehensive quality, the better the integration between the elements. (2) The relationship between integration forces and elements should be positively correlated. The better the relationship, the easier the integration.

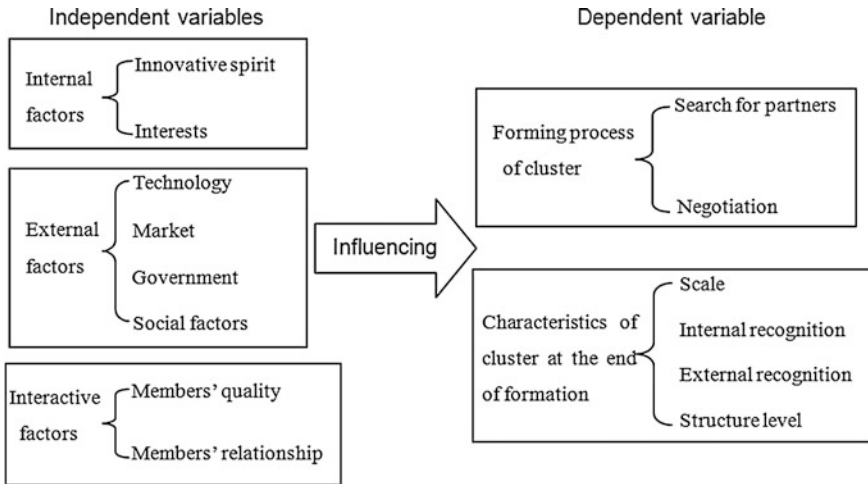


Fig. 1 Model of the influence factors of formation of VIC

3.3 Model

The whole model of the influence factors of formation of VIC is shown as Fig. 1.

4 Empirical Study

4.1 The Survey

To implement the empirical study conveniently, we didn't collect the numerical value of the indications, but had interviews with people who manage the relevant organizations. Through the managers' expression, we determined the degree of the indications. This method is easy to implement, and can reflect the real situation on the whole.

The preliminary interview include 15 Vic. To ensure the accuracy, validity and reliability of the survey, the researchers interview the core members by a questionnaire and oral interview, consult the interviewees, and finally conclude the results of survey.

4.2 The Questionnaire

The questionnaire is designed as Table 1, and the Likert scale form is used to answer the questions.

Table 1 Framework of questionnaire

Main indication	Sub indication	Question
Internal factors	Spirit of innovation	The main purpose of joining the cluster is to produce high level achievements
	Interests	Funds are easy to get after joining the cluster
External factors	Technology	It's difficult to get high level achievements without the cluster
	Market	A favorable position can be obtained in the region or industry by joining the cluster
	Government	The government has given great support for the establishment of the cluster
	Social factors	Collaborative innovation is popular
Interactive factors	Total quality of members	The main members have a strong influence in the industry
	Relationship among members	The relationship among the members of the cluster is very close
Forming process of cluster	Searching for partners	It quickly became clear backup. The member list was determined very soon
	Negotiation	Alternate members agreed soon to participate in the cluster
Characteristics of cluster at the end of formation	Scale	The number of members
	Structure Level	The rules of cluster is perfect, clear and being implemented
	External recognition	The cluster has a strong influence in the industry
	Internal recognition	The cluster has a very strong appeal and the members will participate in joint action actively

4.3 Results Analysis

We use the grey correlation analysis to reveal the relationships between variables as Table 2.

1. To the scale of the VIC, the relational grade of the factor of government is the highest. It can be interpreted as: the more attention the government paid and the more support the government gave, the more members would be attracted to the innovation alliance.
On the other hand, the relational grade of market to the scale is the lowest. It can be interpreted as: If an innovation subject hoped to occupy a place in the market, the scale of the alliance should be small. If the scale is too large, the alliance will have no significance.
2. To the internal recognition of the VIC, the relational grade of the members' total quality is the highest. It can be interpreted as: the comprehensive quality

Table 2 Grey relational grade between variables

Grev relational grade	Scale	Internal recognition	External recognition	Structuring	Searching for partners	Negotiation
Spirit of innovation	0.68	0.73	0.70	0.82	0.82	0.73
Interests	0.73	0.66	0.66	0.67	0.66	0.61
Science and technology	0.65	0.6	0.64	0.57	0.66	0.58
Market	0.57	0.77	0.65	0.7	0.78	0.65
Government	0.84	0.73	0.74	0.63	0.69	0.62
Social factors	0.63	0.65	0.68	0.7	0.72	0.71
Total quality of members	0.62	0.83	0.71	0.74	0.78	0.74
Relationship among members	0.72	0.70	0.61	0.87	0.74	0.80

of members is the key factor to make the cluster and its members recognized. So it's important for the VIC to have high-quality members.

3. To the external recognition of the VIC, the relational grade of the government is the highest. It can be interpreted as: if the VIC is recognized and improved by the government, it would also be recognized by others In the whole industry easily.
4. To the structuring of the VIC, the relational grade of the relationship among members and the spirit of innovation are the highest. It can be interpreted as: if the members are eager for innovation, and the member relationship is close, it will be easy to form the rules of the cluster. But it is difficult for the factors of the science and technology and the government to promote the members to be of one mind.
5. To the forming process of the VIC, including searching for partners and negotiation, the relational grade shows that the spirit of innovation, relationship among members and quality of members are conducive to the formation of the VIC. But the science and technology and interests, which may make the conflict between members, will make the forming process cost more.

5 Conclusions and Suggestions

Based on the above analysis, government's promotion, members' quality, members' relationship and innovative spirit play a relatively important role in the formation process of the VIC. So the suggestions of cultivating and developing the VIC are as follows:

1. The government should play an active role to cultivate and develop the VIC, including functions of direction guide, financial aid, relationships coordination, supervision and appraisal.

2. The formation of VIC should be led by an authority in the industry. And at the same time, core members of VIC should have higher comprehensive quality.
3. The cultivation of VIC should be on the basis of the existing cooperation, so as to establish a solid relationship network.
4. To create a good atmosphere for innovation in the cluster. Through the cluster culture to cultivate and stimulate the innovative spirit of the members.

Acknowledgments This research is supported by the National Natural Science Foundation of China-‘Mechanism of Formation and Collaboration of Scientific and Technological Innovation Platforms Cluster’ (Grant No.: 71350014).

References

1. Porter ME (1990) *The competitive advantage of nations*. The Free Press, New York
2. Porter ME (1998) Clusters and the new economics of competition. *Harvard Bus Rev*, 76(6):77 (Nov/Dec98)
3. Tracey P, Clark GL (2003) Alliances, networks and competitive strategy: rethinking cluster of innovation. *Growth Change* 34(1):1
4. Passiante G, Secundo G (2002) From geographical innovation clusters towards virtual innovation clusters: the innovation virtual system. 42th ERSACongress, University of Dortmund, Germany
5. Tan ZD, Wang WP (2005) Study on the evolution process of industrial clusters in different ways of formation. *Mod Manage Sci* 12:30–31
6. Wu YJ (2010) Formation and evolution mechanism of industrial cluster. *Gansu Soc Sci* 5:181–184
7. Jiang CL (2008) Study on the connotation and the formation path of innovation Clusters. *Mod Manage Sci* 11:10–11
8. Zhou Z (2013) The study of dynamic mechanism of collaborative innovation. *Soft Sci* 7:52–56
9. Xu J (2012) Research on the dynamic mechanism of cooperative innovation. *China Sci Technol Forum* 7:74–80

Research on the Development Path of New-Type R&D Organization in Guangdong Province, China

Li Huang

Abstract In China, new-type R&D organization have become a new policy way to explore and deepen the reform of system of science and technology. After reviewing the developing history of new-type R&D organization in Guangdong, summing up its development path, analysing the reasons for its success, this paper put three suggestions for the development of chinese new research institutions. Firstly, the new-type R&D organization is an effective way to promote the sustainable development of local economy. Secondly, scientific research institutions and development cannot do without government's support and guidance. Thirdly, the planning of the development path of scientific research institutions in different province should be combined with the situation of itself.

Keywords New-type R&D organization · Development path · Science and technology system

1 Introduction

In recent years, a number of new-type R&D organizations emerged in Guangdong province, China. Generally they are builded by enterprises, colleges and universities under the guidance of the government. The Council is the decision-making body of these organizations, formed by members from government departments, universities, institutes, experts and so on. Directors come from China or abroad and are responsible for the daily management. These R&D organizations set science and technology innovation and industrialization in one, which are market oriented and enterprise operated, with flexible operating mechanism, pursuing profit through the research and development of technology and innovation. Although now they are still in the initial stage, their effectiveness and development pattern has caused great concern. This paper studies the development path of new-type R&D organization in

L. Huang (✉)

School of Management, Fujian University of Technology, Fuzhou 350118, China
e-mail: 27938770@qq.com

Guangdong, China, hoping for improving chinese Industry-University-Government cooperation and technology and innovation development.

2 The Development History of New-Type R&D Organization in Guangdong

According to the degree of development, the development of new-type R&D organization in Guangdong have gone through three periods.

2.1 The First Period: 1994–2004

During this stage, the government and research institutions of Guangdong province tried to reform around the combination of scientific research and market orientation, and accumulated a lot of experience for the new-type R&D organization sprouting out later. In December 1996, the Shenzhen government and Tsinghua University built the Research Institute of Tsinghua University, Shenzhen in the form of enterprise. Guided by the spirit of strengthening technological innovation and industrializing of central government, Guangdong government launched a programme which deepened the reform of system of science and technology in 1999, which classified the 69 provincial scientific research institutions into technology development institutions, consultancy service institutions and public institutions, and required technology development institutions changing from public entity into business entity. Later some of these technology development institutions gradually developed into new-type R&D organizations.

2.2 The Second Period: 2005–2010

In 2005, Guangdong Province identified independent innovation as the main melody of science and education strategy. In September of the same year, Guangdong government and the Ministry of education of P.R.C. signed a cooperation agreement on improving the capability of independent innovation and hastening the social and economic development of Guangdong Province. This agreement helped Guangdong Province to establish its Industry-University-Research system, which was assisted by Ministry of education, Ministry of science and technology, Ministry of industry and technology, Chinese Academy of engineering, Chinese Academy of Sciences. Under this background, several representative state-run new-type R&D organizations has set up, such as Guangzhou Institute of Industrial Technology of Guangzhou and Chinese Academy of Sciences, BGI and so on, beginning the exploration of Guangdong Construction

of new-type R&D organization [1]. During this period, the new-type R&D organizations owned by enterprise were developed too. In 2005, East Sunshine Pharmaceutical Group established Dongguan East Sunshine Pharmaceutical Research Institute. In July 2006, Guangzhou Automobile Group Co., Ltd. established Automotive Engineering Research Institute. In 2008, the financial crisis outbreak. Guangdong Province attracted many overseas research and development people to return home and start their business, because of its relatively mature market conditions, good policy treatment and location advantages. Kuang-Chi Institute of Advanced Technology was one representative of the success. New-type R&D organization of Guangdong province intensively constructed and developed at this stage.

2.3 The Third Period: Since 2011

After the second stage, Guangdong government changed their support way from direct support to individual new-type R&D organization to improve the policy environment and strengthen the industrial guide. At this stage, the effect that new research institutions serve the regional economy began releasing. For example, by the end of 2011, the Research Institute of Tsinghua University, Shenzhen has invested or established more than 180 high-tech enterprises, with an annual output value of more than 26 billion yuan. Kuang-Chi Institute have applied 1229 meta-material patents at home and abroad and achieved a coverage of 80 % of the underlying patent in this field [2]. Now, there are more than 100 new-type R&D organizations in Guangdong. The scale of individual institutions is generally small. Most of them are builded by social strength and distributed in four areas of Shenzhen, Guangzhou, Dongguan and Foshan [2].

3 The Development Path of New-Type R&D Organization in Guangdong

Development path emphasizes the direction and trend of things, which is dynamic. The development time of new-type R&D organization in Guangdong is not long. But the result is remarkable and shows the local characteristics. Now let us do a simple summary to the development path of new-type R&D organization in Guangdong. In the early stage, both local governments and research institutes explored the effective methods of scientific research combined with the market. After gaining some experience, the local government began to create conditions benefit to the construction and development of new-type R&D organization, and led the establishment of the typical new-type R&D organization. These favorable support conditions and the demonstration effect attracted a lot of social resource to participate in the construction of local new-type R&D organizations. After the

construction task was initially completed, the government changed its support methods to the industry guidance and support policy making, and consolidated the preliminary results and gradually solved the influence of obstacles at the same time. Social power has gradually become the main power of new-type R&D organization building.

3.1 The Characteristics of the Development Path of New-Type R&D Organization in Guangdong

Compared with Jiangsu, Fujian, Beijing, Hebei and other provinces and cities, the development path of new-type R&D organizations in Guangdong Province has its own characteristics.

3.1.1 A Lot of Support from the Local Government Involved in ...

Guangdong government has taken strong support measures for the development of new-type R&D organization. Government officers and the technology department invested a lot of energy and did a lot of detailed construction work. For example, Shenzhen issued a series of important documents attracting resource such as talent, capital to strengthen the independent innovation and promote high technology industry. According to “peacock plan” launched by Shenzhen government, peacock team can get the maximum 80 million yuan of special fund to start a business or R&D projects. It can be said, without these great effort of Guangdong government, it is impossible for new-type R&D organizations developing so rapidly in Guangdong.

3.1.2 The Daring Spirit of Guangdong Government

Until now, the domestic support of other provinces for the development of new-type R&D organizations is much less, because of the “new” features. Because the new-type R&D institutions are much different from traditional R&D institutions, the understanding to it is very fuzzy. This cognitive confusion greatly hinders support and planning work of the government in various provinces and cities. However, the Guangdong government took a daring spirit to pay a lot of effort to explore experience for new-type R&D organization construction despite of all these obstacles. So far, Guangdong Province has not yet issued any special support policy for new-type R&D organizations, but they have been in a booming development.

3.1.3 The Effective Accumulation of High Levels of Resources

By virtue of its geographical advantage and policy advantages, Guangdong Province seeks out and introduces talent and technology resources in the country and even the world's top. According to statistics, Guangdong Province in recent years has attracted a total of more than 110 academicians, more than 330 scientific research institutions and 310 universities to serve local economy, including 23 the overseas universities and 187 domestic colleges and universities outside Guangdong. Most of these universities are the top ranked universities in China [3]. Guangdong Province also took actions like this for the construction of new-type R&D organizations. No matter the introduction of research teams or investment amount, the support of Guangdong government was great. Strategically advantageous position not only improves the Guangdong R&D strength, but also enhances the possibility of successful construction of new-type R&D organization a lot.

4 Reasons for the Success of Development Path of New-Type R&D Organization in Guangdong

There are two main reasons why the development path of new-type R&D organization in Guangdong achieved positive results in such a short period of time.

4.1 The Development of the New-Type R&D Organization Complies with the Local Economic Development Needs

As for the experiment of chinese reform and opening, Guangdong Province achieved rapid economic development by TFPs, changing from an economically backward agricultural province to the first economic province in China. But the export-oriented development mode brings problems more and more obviously, such as the extensive mode of economic growth, unreasonable industrial structure, rapidly rising costs, environment difficult to sustain and so on. Industrial upgrading and changing development mode has become the need of the sustainable development of the local economy. According to the successful experience of United States, Japan and other developed countries, high technology is an effective way to solve the dilemma in the development of Guangdong's economy. However the research and development in China could not support economic development effectively. The technological achievements conversion rate is only about 10 %, far below the 40 % in developed countries [4]. With the upgrading of industrial structure, the need for technological innovation of the industries in Guangdong becomes more and more urgent, especially for the private economy. The development of new-type R&D organization

conform to this economic demand. Through the system and mechanism innovation, new-type R&D organization combined research and development with local industries closely, which not only been recognized by the market, but also obtained sustainable development power. Conforming to the needs of local economic development is the fundamental reason for the success of new-type R&D organizations in Guangdong.

4.2 Guangdong Government has the Entrepreneurial Spirit

Starting early and the support from local government are two reasons why the development of new-type R&D organizations in Guangdong can be leading in the country. In the sense that the traditional development model was difficult to meet the need of local economy, Guangdong provincial party committee and the provincial government had proposed building an innovation oriented Guangdong and accelerated the transformation and upgrading by improving the capability of independent innovation. New-type R&D organizations conformed to the need of industrial upgrading and technology innovation of enterprise, which got the attention of local government. Because there was no mature precedents in the country, Guangdong government explored the support actively. This actions was not blind but based on the correct understanding to solve the problem of economic development, which reflected the spirit of innovative practice. In his classic “innovation and entrepreneurship”, Peter Drucker points out that only those organizations including public service departments have entrepreneurial spirit that be happy to embrace change and practice innovation. These organizations understand practicing innovation is optimal choice of the lowest risk when opportunities for innovation emerges. Due to the entrepreneurial spirit of Guangdong government, reasonable judgment and the courage to practice, new-type R&D organizations in Guangdong could develop so well. From the development process, we can see Guangdong government judged the development need of new-type R&D organizations accurately and focused on solving these problem at each stage, in compliance with the development law, and promoted the development of new-type R&D organizations greatly.

5 Enlightenments of the Development Path of New-Type R&D Organization in Guangdong

The development path of new-type R&D organization in Guangdong could bring the following enlightenments for chinese innovation development.

5.1 New-Type R&D Organization Is an Effective Way to Promote the Sustainable Development of Local Economy

At present, the phenomenon of irrational economy structure and the extensive mode of economic growth is still prevalent in most provinces of China. Shortage of resources, environmental pollution and ecological imbalance become serious constraints to the modernization of local economy. The experience of Guangdong show that new-type R&D organization could be an effective way to solve these problems. Now some new-type R&D organizations have emerged in other provinces of China. But many of them are still in the initial stage and their effectiveness of the servicing the local economy is not obvious, which make the construction work difficult. After keep on developing for almost ten years, the positive effect of new-type R&D organization in Guangdong began to release, and there were some failed cases too. Therefore, we should not question the positive effect of new-type R&D organization but guide and support them to develop and serve local economy.

5.2 The Construction and Development of New-Type R&D Organization Needs the Support and Guidance of Local Government

From the development path of new-type R&D organization in Guangdong, we can clearly see the local government plays a positive role in promoting the development of new-type R&D organization. When economy develop to a certain level, New-type R&D organization emerges and takes the market as the main service object and demands of talent, capital and other external resource. As it is still in the embryonic stage, it will take a relatively long time for new-type R&D organization to achieve a certain level of development under chinese current social and economic conditions. The support of government could help them go through the start-up period successfully and speed up the process of construction and promote the positive effects releasing. In addition, if new-type R&D organization wants to serve the local economy and upgrades the area traditional industries effectively, it is necessary to comply to the technical requirements of the development of local industries and is inseparable from the planning and guidance of local government.

5.3 The Development Path of New-Type R&D Organization Should Meet the Need of Area Development

As for the experiment of chinese reform and opening, the economy of Guangdong developed quickly, with perfect market economic system, business environment

and a lot of R&D investment. It was reported by the Chinese report of regional innovation capability that the 2013 proportion of government expenditure for science and technology in GDP in Guangdong was 2.25 %, ranking second in the country. These conditions make Guangdong could select a development path of new-type R&D with a lot of support resource, which may not be able to be copied by other provinces. Compared to Guangdong, the economic development level and management experience of many provinces in China is much worse. They do not have the ability to follow the same way completely. We should learn the successful experience of the development path of new-type R&D organization from Guangdong Province. At the same time, we must adjust the development path according to our own situation. For example, in the area whose economy is not developed, introducing some branches of those successful new-type R&D institutes could be considered as a way to increase the likelihood of success. Only to find a development path suited to the actual situation in the province, the development of new-type R&D organization could be sustainable.

References

1. Li D, Chen Y (2013) The status quo and countermeasures of the development of new-type R&D Organizations in Guangdong. *Sci Technol Manag Res* 2013(3):99–101
2. Xu D (2012) The development of new-type R&D Organizations in Guangdong stimulates social creativity. *People's Dly*, 10 May 2012
3. Industry-University-Research in Guangdong (2014) An important starting point for innovation driving the development. *Sci Technol J* 12, 13 Mar 2014
4. Dong G (2013) The National Development and Reform Commission Officials: the Chinese Transformation Rate of Scientific and Technological Achievements is only 10 %. <http://finance.chinanews.com/cj/2013/12-21/5647840.shtml>, 21 Dec 2013

Analysis of Technology Diffusion Among Agricultural Industry Clusters by Game Theory

Chun-Hua Zheng and He-Liang Huang

Abstract With Chinese agricultural developments, agricultural industry clusters emerge as a new medium to promote agricultural technology. The extent of government intervention, the completeness of market, the internal competition in clusters, and the extent of technology demands show great differences among the government- and market-led industry clusters. Therefore, each entity's agricultural technology diffusion mechanisms and dominating force are different as well. Government policies and leading enterprises play important roles to promote agricultural technology diffusion in agricultural industry clusters.

Keywords Game theory · Agricultural industry cluster · Technology diffusion

1 Introduction

Industry clusters are effective to facilitate regional economic development, and make individuals to be competitive in order to adapt to a competitive market while moving from individual competition towards network competition. Chinese agricultural industrialization makes agricultural industry become more versatile. The agricultural productive organizations—as realized in regional clusters—mainly deal with production, processing, and storage. Many Chinese towns and cities have developed into agricultural industry clusters, and their productivity has great impacts on economic development. Based on the extent of government intervention and the

C.-H. Zheng · H.-L. Huang (✉)

School of Economy and Management, Fujian Agriculture and Forestry University,
Fujian, People's Republic of China
e-mail: hhh370@163.com

C.-H. Zheng
e-mail: zchua1972@sohu.com

C.-H. Zheng
School of Management, Fujian University of Technology,
Fujian, People's Republic of China

interaction between market mechanisms and government, agricultural industry clusters can be classified into two types: (1) Market-led agricultural industry clusters, such as Shouguang Vegetables, Anxi Tea Industry, and Guangdong Bamboo Industry. (2) Government-led agricultural industry clusters, such as agricultural technology parks and modern agricultural demonstration districts. According to results in previous research, industry clusters are actually a unique innovative technology system. Whether clusters can successfully promote regional agricultural competitiveness depends on the efficiency of technology innovation diffusion.

2 Framework of Analysis

2.1 The Main Bodies of Agricultural Technology Diffusion

In agricultural industry clusters, the process of technology diffusion is complex. The interactions take place not only between government and agricultural enterprises, or between different kinds of agricultural enterprises, but also between agricultural enterprises and farmers. These entities regarding technology diffusion have different interest demands. This paper selects three subjects: the government (on behalf of the public interest), agricultural enterprises and farmers (on behalf of private interests). It is assumed that each subject is a rational “economist”, and each subject wants to maximize his own interest under specific conditions. Government, enterprises and farmers interact and adapt with each other in the process of technology diffusion. Adopting game theory to analyze their relationship will be more accurate and realistic. The following will introduce each subject in the model.

2.1.1 Agricultural Enterprises

Agricultural enterprises, especially leading enterprises, play an important role in clusters; because they are facilitators and administrators of technology diffusion. Gaining profits is the reason for agricultural enterprises to decide to choose active or passive strategies. Agricultural enterprises are autonomous, rather than waiting passively, to adopt strategies.

2.1.2 Farmers

A large number of farmers, who generally cooperate with agricultural leading enterprises, mainly engage in producing agricultural products. With the establishment of market economy, farmers’ intentions tend to be similar as the industries. That is, farmers also want profits. Gaining profits or not determines farmers’ decision in adopting active or passive strategies.

2.1.3 Government

The government is generally a policy maker. The government is always influenced by other entities while making or modifying policies. The government is also accumulating experiences through making or amending policies, and hopes policies can maximize public interest.

2.2 The Establishment and Analysis of the Game Model

In order to analyze the technology diffusion in clusters and find out the best policies, the researchers use game theory to explain the interactions and strategies-adoption issue that take place between government and agricultural enterprises and between enterprises and farmers. According to the above two, we choose different participants in the game and establish different patterns.

2.2.1 Government-Led Agricultural Industry Clusters

The agricultural industry clusters in agriculture parks and modern agriculture demonstration zones are primarily established by the government. The government invests a lot in the construction of the parks, and hopes to facilitate regional and farmer economy, social or ecological efficiency. The goal of agricultural enterprises is to obtain excess profits while engaging in researching and developing innovative technology and encouraging farmers to adopt new technology. In agricultural technology diffusion, the government will indispensably support enterprises with policies and financial aids. Whether or not the enterprises will adopt new technology will have direct impacts on government’s policy. Therefore, the interaction between the government and enterprises concerning strategy-adoption and profits-gaining is like a game.

Pattern One—The Static Condition of Complete Information

It is assumed that government holds two policies (passive support and active support), and agricultural enterprises also hold two strategies (passive diffusion and active diffusion). The government and the enterprises do not get along with each other, and their relationship will be like the following matrix (Table 1):

Table 1 Static condition of complete information

		Government	
		Passive support	Active support
Agricultural enterprises	Active diffusion	(2, 2)	(2, 4)
	Negative diffusion	(4, 2)	(4, 4)

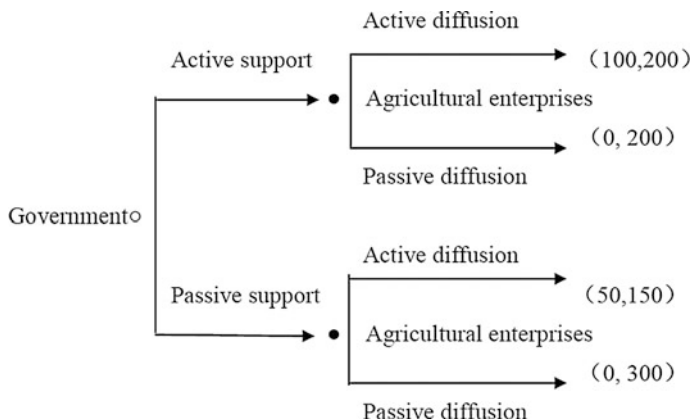


Fig. 1 The dynamic condition of complete information

From the perspective of a static game, the relationship between a rational government and agricultural enterprises will show the Nash equilibrium (active support, active diffusion). Its total profits will be 8 (4 from the industries + 4 from the government). Obviously, this is the maximum profit. Therefore, under the static condition of complete information, the government will make the most supportive policies or provide subsidies for the industries, and then the enterprises will actively administer the policies made by the government.

Pattern Two—The Dynamic Condition of Complete Information

In terms of technology diffusion, the government will initially adopt realistic policies. It is assumed that the government will first adopt technology diffusion policies, namely positive or negative support. Enterprises will have two options (passive diffusion, active diffusion). In the first stage, the government selects and issues clear policies. Secondly, the enterprises make decisions. Figure 1 shows the relationship between enterprises and the government. The hollow dot refers to those who make the choice at first in the game. Then, the solid dots refer to those who already have the background information of the game and it is their turn to make choices.

In the distribution of profit shown above, if the government makes active supporting policies for technology diffusion, then agricultural enterprises will, of course, choose corresponding active diffusion strategies. Then the profit is 100, total profit = government’s profit + enterprise’s profit = possibly 300 the maximum. If the government formulates passive policies, but agricultural enterprises pursue maximum profit and choose active diffusion strategies. Then if the total profit did not reach the maximum (50 + 150 = 200), and the industries do not choose passive strategies either. Then the overall profit will still be the best (0 + 300 = 300), even if the profit of the enterprises is 0. Based on the above analysis, under the dynamic

condition of complete information, no matter whether the government adopts active or passive policies, the clusters will generally adopt positive strategies.

2.2.2 Market-Led Agricultural Industry Clusters

Among the clusters, leading enterprises will base on market demands to produce regionally advantageous products; gather a number of collaborative industries, farmers and organizations; apply technologies to R&D; and manufacture, process and market products. Then government intervention has no place for this type of clusters. For the convenience of analysis, it is assumed that the farmers and small-and-medium-sized enterprises share similar natures. Then leading and collaborative enterprises are treated as the subjects in this pattern.

Pattern One—The Static Condition of Complete Information

Leading and collaborative enterprises will adopt two strategies (passive diffusion strategy, active diffusion strategy), which are uncoordinated in technology diffusion. When the information is complete and symmetric, then their relationship will be as Table 2.

From the perspective of a static game, both leading and cooperative enterprises are rational. There is only Nash equilibrium (active diffusion strategy, active diffusion strategy) in the game, and the total profit is 8 (4 from leading enterprises + 4 from small-to-medium-sized enterprises). This is obviously the maximum profit. Therefore, under the complete information static condition, the market-led leading enterprises, cooperative enterprises, and farmers will simultaneously adopt a positive strategy.

Pattern Two—The Static Condition of Incomplete Information

Under this condition, the enterprises only have two strategies (passive diffusion strategy, active diffusion strategy) and they are uncoordinated. The information is incomplete and asymmetric. While not knowing what their opponents will choose,

Table 2 Static condition of complete information

		Collaborative enterprises	
		Passive diffusion strategies	Active diffusion strategies
Leading enterprises	Passive diffusion strategies	(2, 2)	(2, 4)
	Active diffusion strategies	(4, 2)	(4, 4)

Table 3 Static condition of incomplete information

		Collaborative enterprises	
		Passive diffusion strategies	Active diffusion strategies
Leading enterprises	Passive diffusion strategies	(2, 2)	(1, 4)
	Active diffusion strategies	(4, 1)	(4, 4)

the leading and cooperative enterprises will consider maximizing their profits; then their relationship will be as Table 3:

Under the static conditions of incomplete information, each entity wants to maximize his profits. It is assumed that leading enterprises will adopt passive diffusion strategies with an α probability, and adopt passive diffusion strategies with a $1 - \alpha$ probability. The collaborative enterprises will select passive diffusion strategies with a β probability, and select active diffusion strategies with a $1 - \beta$ probability. Then the expected profit for leading enterprises is:

$$E1(\alpha, \beta) = 2\alpha\beta + \alpha(1 - \beta) + 4(1 - \alpha)\beta + 4(1 - \alpha)(1 - \beta) = \alpha\beta - 3\alpha + 4.$$

Similarly, the expected profit for collaborative enterprises is:

$$E2(\alpha, \beta) = 2\alpha\beta + 4\alpha(1 - \beta) + 2(1 - \alpha)\beta + 4(1 - \alpha)(1 - \beta) = \alpha\beta - 3\beta + 4.$$

We can prove that $\max E1(\alpha, \beta)$, $\max E2(\alpha, \beta)$ will be $\alpha = 1/3$, $\beta = 1/3$, the probability of mixed strategies (1/3, 2/3), (1/3, 2/3) will demonstrate following probabilities:

- $Y_{\text{passive} \times \text{passive}}$ P (leading enterprises adopt passive strategies) \times P (collaborative enterprises adopt passive strategies) = $1/3 \times 1/3 = 1/9$
- $Y_{\text{passive} \times \text{active}}$ P (leading enterprises adopt passive strategies) \times P (collaborative enterprises adopt active strategies) = $1/3 \times 2/3 = 2/9$
- $Y_{\text{active} \times \text{passive}}$ P (leading enterprises adopt active strategies) \times P (collaborative enterprises adopt passive strategies) = $2/3 \times 1/3 = 2/9$
- $Y_{\text{active} \times \text{active}}$ P (leading enterprises adopt active strategies) \times P (collaborative enterprises adopt active strategies) = $2/3 \times 2/3 = 4/9$

The above analysis shows that in pursuit of maximum profit, leading and collaborative enterprises (farmers) will most likely adopt active strategies. As the above pattern shows, the probability of (active diffusion strategy \times active diffusion strategy) is 4/9, and it is far greater than that of other sets of strategies. Therefore, under the static conditions of incomplete information, the market-driven leading and collaborative enterprises in the clusters will adopt active strategies.

Based on the above analysis, we can draw the conclusions. In terms of government-led clusters, by considering the overall maximum profit, the

government will make active agricultural technology diffusion policies, and agricultural enterprises will also actively implement that policy. Even if the government adopts passive attitudes, enterprises will not wait passively, but adopt active strategies in order to maximize their profit.

In terms of market-led clusters, due to the complete effects of market mechanisms and weakened government intervention (i.e., the government is like an outsider). Then leading and collaborative enterprises are the main entities in the game and expect to gain maximum profit. Then both entities will adopt active strategies regardless of complete or incomplete information assumptions.

3 Conclusions and Suggestions

All information above shows that the successful rate of promoting Chinese agricultural research is approximately 30–40 %, while it is 70–80 % in developed countries. Therefore, we have to think about a mechanism to promote agricultural technology, and think about how to effectively promote technology diffusion. The emergence of agricultural industry clusters provides a new platform for improving the efficiency of technology diffusion. Based on the above analysis, the authors suggest that different types of industries clusters should adopt appropriate strategies in order to facilitate the improvement of agricultural technology and promote the competitiveness of Chinese agriculture in international market. The suggestions and implications of the present paper are as follows:

3.1 Active Technology Diffusion Strategies in Government-Led Industry Clusters

First, the government has to establish a regional agricultural technology diffusion system. The government has to establish and improve the research and cooperation networks involving agricultural enterprises and agricultural research institutions. The government has to provide services for industries, and avoid repeatedly research and develop outdated technology. Besides, the government has to invest in important agricultural technologies, establish technology innovation projects, develop prospective and economic projects, and reduce developing less potential technology.

Secondly, the agricultural technology promotion departments in government should manage to apply new technology to actual production. In the process of technology diffusion, an intermediary channel is like a bridge to connect supplies and demands. The government authorities at all levels should establish intermediary service agencies to ensure efficient operation of organizations and policies. On the other hand, because the development of new technology requires lots of capital and human resources, then it is indispensable for government to run risks. Only when

agricultural enterprises recognize that innovative technologies will bring greater-than-expected profits, will they adopt new technology. However, this is not conducive to the timely promotion of new technology. Therefore, the government should provide supportive policies (i.e., financial subsidies, patent protection, etc.) to motivate entities to adopt new technology so as to accelerate the promotion of technology.

3.2 Active Technology Diffusion Strategies in Market-Led Industry Clusters

Market-led clusters, especially leading enterprises, collaborative enterprises, farmers and relevant entities have to work together to participate in innovative diffusion networks.

First, leading enterprises should actively implement active technology research strategies through close cooperation with other enterprises and farmers, constantly update their knowledge, adopt latest innovative technologies, and facilitate the establishment of innovative industry clusters. After leading enterprises obtain new technology achievements, they have to manage to work with collaborative enterprises and farmers to accelerate the application of technology. At the same time, the leading enterprises can use market mechanisms to provide collaborative enterprises and farmers with business opportunities, latest technology and training. They can encourage industrial chains to work together to apply new technology just in the clusters, and make clusters have exclusive advantages in production shortly.

Secondly, collaborative enterprises and farmers have to be active in adopting new agricultural technology. Agricultural technology is highly regional. Agricultural enterprises and farmers have to have the ideas about local agriculture. On the one hand, they have to contribute their knowledge of agricultural technology; support and actively participate in research and development of new technology and adopt latest technology. On the other hand, clusters have different highly specialized divisions, including research, planting, processing, marketing, and transportation. They are a part of the whole. They should actively develop their own professional skills, and improve their technical proficiency.

Weakness of Zhang-Wang Scheme Without Using One-Way Hash Function

Zhi-Pan Wu

Abstract Zhang and Wang proposed a signature scheme without using one-way hash function and message redundancy. It based on Chang-Chang scheme and gives an improvement which overcomes the known forgery attack. In this paper, we show this scheme can not suffer forgery attack where it does not use the one-way hash function. We believe the one-way hash function still demands message recovery and redundancy.

Keywords Digital signature · One-way hash function · Message redundancy · Forgery attack

1 Introduction

With the growth of the Internet, the digital signature is becoming very important in the electronic commerce, it provides the cryptographic services on authentication and data integrity where also agree between signer and verifier, there are several digital signature scheme have been proposed [1]. In Shieh et al. [5] proposed two multi-signature schemes based on Nyberg-Rueppel scheme [4], one is parallel multi-signature scheme, and other is serial multi-signature scheme. They proposed scheme enables the specified verifier to verify and to recover the message, it applied to smaller bandwidth of data communications. Further, one-way hash function and message redundancy scheme are not used. But this scheme was insecure, it easy suffered forgery attack, hence many paper cryptanalysis and improvement [7] on Shieh et al. signature scheme. Recently, Zhang and Wang [8] presented a new digital signature scheme with message recover without using one-way hash function and message redundancy, and claimed that their scheme can resist forgery. However, in this paper, we discover their scheme still insecure, it does not resist

Z.-P. Wu (✉)

Department of Computer Science, Huizhou University, Huizhou 516007, China
e-mail: hz_wzp@163.com

forgery attack. The paper is organized as follows: Sect. 2 reviews Zhang's scheme, Sect. 3 analyzes the security of Zhang's scheme. Section 4 points out a weakness security and presents our conclusion.

2 Review of Zhang-Wang Scheme

In Zhang and Wang [8] proposed a signature scheme without using one-way hash functions. In their scheme, p and g , where p is a large prime number and g is a primitive element in $GF(p)$. The signer randomly selects his private key x , where $\gcd(x, p-1) = 1$, and computes public key $y \equiv g^x \pmod{p}$. There are two phases. The signature generation and verification phase are described below.

2.1 Signature Generation Phase

Suppose the signer wants to sign the message M , then execute the follow steps:

Step 1. The signer computes

$$s \equiv (y + M)^{M \pmod{p-1}} \pmod{p}. \quad (1)$$

Step 2. The signer chooses a random number $k \in Z_{p-1}^*$ and computes

$$r \equiv M \cdot s \cdot g^{-k} \pmod{p}. \quad (2)$$

Step 3. The signer computes t where

$$s + t \equiv x^{-1} \cdot (k - r \oplus s) \pmod{p-1}. \quad (3)$$

Step 4. The signer sends the triple parameters (s, r, t) of M to the verifier.

2.2 Verification Phase

The verifier receives the signature (s, r, t) from the signer, and can then verify the validity of signature in the as following steps:

Step 1. The verifier computes

$$M' \equiv y^{s+t} \cdot r \cdot g^{r \oplus s} \cdot s^{-1} \pmod{p}. \quad (4)$$

Step 2. The verifier checks

$$s \equiv (y + M)^{M(\bmod p-1)}(\bmod p). \quad (5)$$

If the rule holds, it shows that the signature (s, r, t) is valid.

Proof

$$\begin{aligned} M' &\equiv y^{s+t} \cdot r \cdot g^{r \oplus s} \cdot s^{-1}(\bmod p). \\ &\equiv y^{s+t} \cdot M \cdot s \cdot g^{-k} \cdot g^{r \oplus s} \cdot s^{-1}(\bmod p). \\ &\equiv g^{k-r \oplus s} \cdot M \cdot g^{-k+r \oplus s}(\bmod p). \\ &\equiv M(\bmod p). \end{aligned} \quad (6)$$

3 Misuse Order of Operation Problem

The Zhang-Wang scheme does not appear to be true, as pointed out below. According to [2] and [6], the addition has higher precedence than bitwise exclusive-or. Therefore the addition should be applied first, and then processes bitwise exclusive-or operation in a computer programming language such as C++ and so on. This rule is known as a precedence rule or order of operation. We set $(k - r = u)$ and $(-k + r = -u)$, and recompute follow equation.

Proof

$$\begin{aligned} M' &\stackrel{?}{\equiv} y^{s+t} \cdot r \cdot g^{r \oplus s} \cdot s^{-1}(\bmod p). \\ &\stackrel{?}{\equiv} (g^x)^{x^{-1} \cdot (k-r \oplus s)} \cdot r \cdot g^{r \oplus s} \cdot s^{-1}(\bmod p). \\ &\stackrel{?}{\equiv} g^{(k-r \oplus s)} \cdot r \cdot g^{r \oplus s} \cdot s^{-1}(\bmod p). \\ &\stackrel{?}{\equiv} g^{u \oplus s} \cdot M \cdot s \cdot g^{-k} \cdot g^{r \oplus s} \cdot s^{-1}(\bmod p). \\ &\stackrel{?}{\equiv} g^{u \oplus s} \cdot M \cdot g^{-k+r \oplus s}(\bmod p). \\ &\stackrel{?}{\equiv} g^{u \oplus s} \cdot M \cdot g^{(-u) \oplus s}(\bmod p). \\ &\neq M(\bmod p). \end{aligned} \quad (7)$$

Correction equation $x^{-1} \cdot (k - r \oplus s)(\bmod p)$ is used instead of $x^{-1} \cdot (k - (r \oplus s))(\bmod p)$ by Eq. (3). Thus, if one wants to force bitwise exclusive-or to precede addition, one writes $(k - (r \oplus s))$.

Proof

$$\begin{aligned}
M' &\equiv y^{s+t} \cdot r \cdot g^{r \oplus s} \cdot s^{-1} \pmod{p}. \\
&\equiv (g^x)^{x^{-1} \cdot (k - (r \oplus s))} \cdot M \cdot s \cdot g^{-k} \cdot s^{-1} \pmod{p}. \\
&\equiv g^{k - (r \oplus s)} \cdot M \cdot g^{-k} \cdot g^{r \oplus s} \pmod{p}. \\
&\equiv M \pmod{p}.
\end{aligned} \tag{8}$$

The other revision is $M' \equiv y^{s+t} \cdot r \cdot g^{r \oplus s} \cdot s^{-1} \pmod{p}$ is instead of $M' \equiv y^{s+t} \cdot r \cdot g^{k - (k - r \oplus s)} \cdot s^{-1} \pmod{p}$ by Eq. (4).

We also assume $(k - r = u)$ and $(-k + r = -u)$ and recompute the equation.

Proof

$$\begin{aligned}
M' &\equiv y^{s+t} \cdot r \cdot g^{k - (k - r \oplus s)} \cdot s^{-1} \pmod{p}. \\
&\equiv (g^x)^{x^{-1} \cdot (k - r \oplus s)} \cdot r \cdot g^{k - (k - r \oplus s)} \cdot s^{-1} \pmod{p}. \\
&\equiv g^{(k - r \oplus s)} \cdot M \cdot s \cdot g^{-k} \cdot g^{k - (k - r \oplus s)} \cdot s^{-1} \pmod{p}. \\
&\equiv g^{(u \oplus s)} \cdot M \cdot g^{-k} \cdot g^k \cdot g^{-(u \oplus s)} \pmod{p}. \\
&\equiv M \pmod{p}.
\end{aligned} \tag{9}$$

4 Our Attack Method

In this section, we will point out a weakness for the Zhang-Wang scheme. If attacker Eve wants to fake a message to Bob, she does not need to guess a password or challenge discrete logarithm problem. Even if the Zhang-Wang scheme used bitwise exclusive-or (XOR) operation to resist an algebra attack, a leak still exists in which the number system of two complementary numbers is the most common method of representing on the computer [3]. In the follow, we describe the practical issue that two variables do bitwise exclusive-or operation problem.

4.1 Two's Complement System

The two's complement-complement system has the advantage of not requiring that the addition and subtraction circuitry examine the signs of the operands to determine whether to add or subtract. The two's complement of zero is zero: inverting gives all ones, and adding one changes the ones back to zero. The zero's one complement is $(11111111)_2$, then add one to become $(00000000)_2$, so it is itself.

$(0)_{10} \cdot (-1)_{10} = (0)_{10}$. The $(10000000)_2$ one's complement is $(01111111)_2$, then add one to become $(10000000)_2$. $(-128)_{10} \cdot (-1)_{10} = (128)_{10}$. Also the two's complement of the most negative number representable (e.g. a one as the most-significant bit and all number bits zero) is itself. Notation:

\oplus express the bitwise exclusive-or operation.

$()_{10}$ express a decimal number system.

$()_2$ express a binary number system.

$[Pr]$ express probability.

$$(r, s) \Rightarrow \begin{cases} \text{are odd numbers, the } [Pr = \frac{1}{4}]. \\ \text{one odd and even, the } [Pr = \frac{1}{2}]. \\ \text{are even numbers, the } [Pr = \frac{1}{4}]. \end{cases}$$

For Example:

$$\begin{aligned} (193)_{10} &= (11000001)_2 \\ (249)_{10} &= (11111001)_2 \\ (193)_{10} \oplus (249)_{10} &= (00111000)_2 \\ (193)_{10} \oplus (249)_{10} &= (56)_{10} \\ (-193)_{10} &= (1111111100111111)_2 \\ (-249)_{10} &= (1111111100000111)_2 \\ (-193)_{10} \oplus (-249)_{10} &= (000000000111000)_2 \\ (-193)_{10} \oplus (-249)_{10} &= (56)_{10} \end{aligned}$$

4.2 XOR Operation

The XOR operation is a common component in design of digital logic. It is used on adder, cryptosystem or other applications. We describe the XOR boolean algebra as below. There are some definitions:

$$\bar{0} = 1 \tag{10}$$

$$\bar{1} = 0 \tag{11}$$

$$A \oplus 1 = \bar{A} \tag{12}$$

$$A \oplus 0 = A \tag{13}$$

$$A \oplus A = 0 \tag{14}$$

$$A \oplus \bar{A} = 1 \quad (15)$$

$$A \oplus B = \bar{A}B + A\bar{B} \quad (16)$$

Theorem 1 Let \oplus be an operation on the set X . It is called commutative if $A \oplus B = B \oplus A$ for all $A, B \in X$.

Proof According to Eq. (16), $A \oplus B = \bar{A}B + A\bar{B}$ (known definition) $B \oplus A = \bar{B}A + B\bar{A}$, therefore $\bar{A}B + A\bar{B} = \bar{B}A + B\bar{A}$. We obtain $A \oplus B = B \oplus A$. Thus, the XOR matches commutative law.

Theorem 2 Let \oplus be an operation in the set X . It is called associative if $(A \oplus B) \oplus C = A \oplus (B \oplus C)$ for all $A, B \in X$.

Proof

$$\begin{aligned} (A \oplus B) \oplus C &= (\bar{A}B + A\bar{B}) \oplus C. \\ &= (\bar{A}B + A\bar{B}) \oplus C. \\ &= \overline{(\bar{A}B + A\bar{B})}C + (\bar{A}B + A\bar{B})\bar{C}. \\ &= \overline{(\bar{A}B + A\bar{B})}C + (\bar{A}B + A\bar{B})\bar{C}. \\ &= \overline{(\bar{A}B)} \cdot (\overline{A\bar{B}})C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C} \\ &= (A + \bar{B})(\bar{A} + B)C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C} \\ &= A\bar{A}C + ABC + \bar{B} \cdot \bar{A}C + \bar{B}BC + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C} \\ A\bar{A} &= 0 \text{ and } B\bar{B} = 0 \\ &= ABC + \bar{B} \cdot \bar{A}C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C} \end{aligned}$$

Computing $A \oplus (B \oplus C)$

$$\begin{aligned} A \oplus (B \oplus C) &= A \oplus (\bar{B}C + B\bar{C}) \\ &= \bar{A}(\bar{B}C + B\bar{C}) + A\overline{(\bar{B}C + B\bar{C})} \\ &= \bar{A}BC + \bar{A}B\bar{C} + A\overline{(\bar{B}C)} \cdot \overline{(B\bar{C})} \\ &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + A(B + \bar{C})(\bar{B} + C) \\ &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + AB\bar{B} + A\bar{B} \cdot \bar{C} + ACB + A\bar{C}\bar{C} \\ &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + ABC + A\bar{C} \cdot \bar{B} \end{aligned}$$

$$\begin{aligned} \therefore ABC + \bar{B} \cdot \bar{A}C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C} &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + ABC + A\bar{C} \cdot \bar{B} \\ \therefore (A \oplus B) \oplus C &= A \oplus (B \oplus C) \end{aligned}$$

Here, the XOR matches associative law.

Theorem 3 Let $A = B$, $A \oplus B = \overbrace{0000\dots0000}^{bits}$.

Proof According to (14), $A \oplus A = 0$, therefore $A \oplus B = \overbrace{0000\dots0000}^{bits}$.

Theorem 4 If A, B are odd numbers, $(A) \oplus (-A) = \overbrace{1111\dots1110}^{bits}$, $(B) \oplus (-B) = \overbrace{1111\dots1110}^{bits}$. $(A \oplus B) = (-A \oplus -B)$.

Proof According to Theorem 3, if $A = B$, then

$(A \oplus B) \oplus (-A \oplus -B) = \overbrace{0000\dots0000}^{bits}$. From Theorem 1 commutative law and Theorem 2 associative law, we rewrite this equation $(A \oplus B) \oplus (-A \oplus -B) = (A \oplus -A) \oplus (B \oplus -B)$. According to Theorem 4,

$A \oplus -A = B \oplus -B$. From Theorem 3, $(A \oplus B) \oplus (-A \oplus -B) = \overbrace{0000\dots0000}^{bits}$. We get $(A \oplus B) = (-A \oplus -B)$.

Theorem 5 If A, B are even numbers, $4|A, B$ and $8 \nmid A, B$. $(A \oplus B) = (-A \oplus -B)$.

Proof We assume A and B are n bits even numbers, when $4|A$ and $8 \nmid A$,

$A = \overbrace{****\dots100}^{bits}$. When $4|B$ and $8 \nmid B$, $B = \overbrace{****\dots100}^{bits}$. $(A \oplus B) = \overbrace{****\dots000}^{bits}$. Suppose $-A = \overbrace{####\dots100}^{bits}$, and $-B = \overbrace{####\dots100}^{bits}$. $(-A \oplus -B) = \overbrace{♣♣♣♣\dots000}^{bits}$. $(A \oplus B) \oplus (-A \oplus -B) = \overbrace{****\dots000}^{bits} \oplus \overbrace{♣♣♣♣\dots000}^{bits} = 0$. Therefore $(A \oplus B) = (-A \oplus -B)$.

4.3 Our Attack

From above statement, the attacker Eve can easily fake the valid signature (r', s', t) in the following steps.

- Step 1. Eve sets $r' = -r$.
- Step 2. Eve sets $s' = -s$.
- Step 3. Eve sends (r', s', t) signatures to Bob, and successful executes the forgery attack.

Proof

$$\begin{aligned}
 M'' &\stackrel{?}{\equiv} y^{s'+t} \cdot (r') \cdot g^{(r' \oplus s')} \cdot (s)^{-1} \pmod{p}. \\
 &\equiv g^{s'+t} \cdot M \cdot (s') \cdot g^{-k} \cdot g^{(r' \oplus s')} \cdot (s')^{-1} \pmod{p}. \\
 &\equiv g^{k-(r' \oplus s')} \cdot M \cdot g^{-k+(r' \oplus s')} \pmod{p}. \\
 &\equiv M' \pmod{p}.
 \end{aligned} \tag{17}$$

5 Conclusion

Several programming languages use precedence levels that conform to the order of operation used in mathematical precedence. In general, the arithmetic is always higher than bitwise logical operation on precedence. If a designer/developer misused or misunderstood this situation, it may cause a dangerous problem. We clearly described some examples of this case in the paper. It is a good way to prevent multiplicative property of algebra attack using the XOR operation. However, according to our analysis, the Zhang-Wang scheme is still insecure.

Acknowledgements The authors would like to thank the reviewers for their comments that help improve the manuscript. The author also thanks Chenglian Liu for his useful suggestion.

References

1. Chang CC, Chang YF (2004) Signing a digital signature without using one-way hash functions and message redundancy schemes. *IEEE Commun Lett* 8(8):485–487
2. Kruse RL, Ryba AJ (2000) *Data structures and program design in C++*. Prentice Hall, NJ
3. Liu C, Chen S, Sun S (2012) Security of analysis mutual authentication and key exchange for low power wireless communications. *Energy Procedia* 17(Part A):644–649
4. Nyberg K, Rueppel RA (1994) Message recovery for signature schemes based on the discrete logarithm problem. In: *Advances in Cryptology—EUROCRYPT'94*
5. Shieh SP, Lin CT, Yang WB, Sun HM (2000, July) Digital multisignature schemes for authenticating delegates in mobile code systems. *IEEE Trans Veh Technol*
6. Wikipedia: Order of operations. Website (2009). http://en.wikipedia.org/wiki/Order_of_operations
7. Zhang F et al (2005) Cryptanalysis of Chang et al. signature scheme with message recovery. *IEEE Commun Lett* 9(4):358–359
8. Zhang J, Wang Y (2005) An improved signature scheme without using one-way hash functions. *Appl Math Comput* 170(2):905–908

Weakness of an ElGamal-Like Cryptosystem for Enciphering Large Messages

Jie Fang, Chenglian Liu and Jieling Wu

Abstract In 2002, Hwang et al. proposed an ElGamal-like cryptosystem for enciphering large message where it modified from ElGamal cryptosystem. They believe their scheme is based on the difficulty of finding the composite exclusive-or operation. Although, they used bitwise exclusive-or to against multiplicative attack. For this scheme, it is still insecure. In this paper, we give a proof to certain that we claimed.

Keywords Discrete logarithms · Elgamal cryptosystem · Bitwise Exclusive-OR

1 Introduction

A well-known public key cryptosystem ElGamal algorithm was proposed in 1985 which it based on discrete logarithms problem. The assumption is variety with RSA assume on factoring large integer numbers. In 2002, Hwang et al. [2] presented an ElGamal-like cryptosystem which it improved the original method to encrypt a large plaintext. Their method are both the Diffie–Hellman distribution and the ElGamal scheme. Lyuu et al. [5] firstly gave an attack that they assume if the number of plaintexts n exceeds the order of 2 modulus q and the prime p is chosen to be of the form $2^e q$, where e is a positive integer and q is prime number. After, Wang et al. also demonstrated a vulnerable in 2006. They mentioned to select the

J. Fang (✉)

School of Electronics and Information Engineering, Fuqing Branch
of Fujian Normal University, Fuqing 350300, China
e-mail: fangjie_1@foxmail.com

C. Liu

Department of Computer Science, Huizhou University,
Huizhou 516007, China
e-mail: chenglian.liu@gmail.com

J. Wu

Department of Economics and Management, Huizhou University,
Huizhou 516007, China
e-mail: jieling.wu@hotmail.com

prime p such that the smallest positive integer T for $2^{T+1} \equiv 2 \pmod{p-1}$ is as large as possible upon on Carmichael number assumption. In this paper, we simply showed a variety method which the exclusive-or be used in some situation case of cryptosystem is dangerous. Section 2 briefly review the ElGamal-like scheme, and Sect. 3 is our comment. The conclusion draws in final section.

2 Review of Hwang–Chang–Hwang’s Scheme

In 2002, Hwang et al. [2] proposed an elgamal-like cryptosystem for enciphering large messages scheme, Lyuu et al. [5] and Wang et al. [6] pointed out some attacks, and then propose a practical anonymous user authentication scheme with security. The detailed as below:

2.1 The ElGamal Cryptosystem

The ElGamal [1] cryptosystem was proposed in 1985, it based on discrete logarithms. Let p is a large prime number, and g is primitive root where $g \in \mathbb{Z}_p$, and compute the public key $y_i \equiv g^x \pmod{p}$. The x denotes secret key. Here p , g and y are public information, the x_i and r are private information. If user u_i want to deliver the message m ($0 \leq m \leq p-1$) to u_j , u_i randomly selects an integer r and then encrypts m as below:

$$b \equiv g^r \pmod{p}. \quad (1)$$

$$c \equiv m \cdot y_i^r \pmod{p}. \quad (2)$$

u_i sends (b, c) to u_j . When u_j receives (b, c) , u_j decrypts c as follows:

$$m \equiv c \cdot (b^{x_j})^{-1} \pmod{p}. \quad (3)$$

The cipher c depends on both plaintext m and the random integer r , a different random number r will obtain a different ciphertext c from same plaintext m . There are two restriction in the ElGamal cryptosystem; one is random number r can not be repeated and others is message m must be less than $p-1$.

2.2 Hwang et al.’s Scheme

Assume p is a large prime number such as 513 bits and g is a primitive element of $GF(p)$. Each user u_i randomly choose his private key $x_i \in \mathbb{Z}_p$ and computes the public key $y_i \equiv g^{x_i} \pmod{p}$. p , g and y_i are published information. Any user wants to deliver the message m_i to u_i by following steps:

Step 1. Break plaintext m_i into t pieces m_1, m_2, \dots, m_t , each piece of length being 512 bits.

Step 2. Generate two random numbers r_1 and r_2 , where $1 < r_1, r_2 \leq p - 1$, and compute b_1 and b_2 as follows:

$$b_1 \equiv g^{r_1} \pmod{p}. \quad (4)$$

$$b_2 \equiv g^{r_2} \pmod{p}. \quad (5)$$

Step 3. Compute $C_j, j = 1, 2, \dots, t$ as follows:

$$C_j \equiv m_j \cdot (y_i^{r_1} \oplus (y_i^{r_2})^{2^j}) \pmod{p}. \quad (6)$$

Step 4. Send $\{b_1, b_2, C_j, j = 1, 2, \dots, t\}$ to the receiver through a public channel.

After receiving $\{b_1, b_2, C_j, j = 1, 2, \dots, t\}$ from the sender, the receiver recovers the plaintext m_i from following:

$$m_j \equiv C_j \cdot (b_1^{x_i} \oplus (b_2^{x_i})^{2^j})^{-1} \pmod{p}. \quad (7)$$

3 Our Comment

In this section, we will introduce two point views; one is logical bitwise exclusive-or operation, the other one is order of operations in computer system. We introduce the precedence properties in some programming languages.

3.1 The Exclusive-OR Operation Issue

The XOR operation is a common component in design of digital logical. It is used on adder, cryptosystem or other application. We described the XOR boolean algebra as below.

Notation:

\oplus : express bitwise exclusive-or operation.

$()_{10}$: express decimal number system.

$()_2$: express binary number system.

$[Pr]$: express probability.

There are some axioms known of definition:

$$\bar{0} = 1 \quad (8)$$

$$\bar{1} = 0 \quad (9)$$

$$A \oplus 1 = \bar{A} \quad (10)$$

$$A \oplus 0 = A \quad (11)$$

$$A \oplus A = 0 \quad (12)$$

$$A \oplus \bar{A} = 1 \quad (13)$$

$$A \oplus B = \bar{A}B + A\bar{B} \quad (14)$$

Theorem 1 Let \oplus be an operation on the set X . It is called commutative if $A \oplus B = B \oplus A$ for all $A, B \in X$.

Proof According to Eq. (14), $A \oplus B = \bar{A}B + A\bar{B}$ (known definition), $B \oplus A = \bar{B}A + B\bar{A}$, therefore $\bar{A}B + A\bar{B} = \bar{B}A + B\bar{A}$. We obtain $A \oplus B = B \oplus A$.

Thus, the XOR matches commutative law.

Theorem 2 Let \oplus be an operation in the set X . It is called associative if $(A \oplus B) \oplus C = A \oplus (B \oplus C)$ for all $A, B \in X$.

Proof

$$\begin{aligned} (A \oplus B) \oplus C &= (\bar{A}B + A\bar{B}) \oplus C. \\ &= (\bar{A}B + A\bar{B}) \oplus C. \\ &= \overline{(\bar{A}B + A\bar{B})}C + (\bar{A}B + A\bar{B})\bar{C}. \\ &= (\bar{A}\bar{B} + \bar{A}B)C + (\bar{A}B + A\bar{B})\bar{C}. \\ &= \overline{(\bar{A}B)} \cdot (\bar{A}\bar{B})C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C}. \\ &= (A + \bar{B})(\bar{A} + B)C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C}. \\ &= A\bar{A}C + ABC + \bar{B} \cdot \bar{A}C + \bar{B}BC + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C}. \\ A\bar{A} &= 0 \text{ and } B\bar{B} = 0. \\ &= ABC + \bar{B} \cdot \bar{A}C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C}. \end{aligned}$$

Computing $A \oplus (B \oplus C)$.

$$\begin{aligned} &= A \oplus (B \oplus C) = A \oplus (\bar{B}C + B\bar{C}). \\ &= \bar{A}(\bar{B}C + B\bar{C}) + A(\overline{\bar{B}C + B\bar{C}}). \\ &= \bar{A}BC + \bar{A}BC + A(\overline{\bar{B}C}) \cdot (\overline{B\bar{C}}). \\ &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + A(B + \bar{C})(\bar{B} + C). \\ &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + A\bar{B}\bar{B} + A\bar{B} \cdot \bar{C} + ACB + A\bar{C}C. \\ &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + ABC + A\bar{C} \cdot \bar{B}. \\ \therefore ABC + \bar{B} \cdot \bar{A}C + \bar{A}B\bar{C} + A\bar{B} \cdot \bar{C} &= \bar{A} \cdot \bar{B}C + \bar{A}B\bar{C} + ABC + A\bar{C} \cdot \bar{B}. \\ \therefore (A \oplus B) \oplus C &= A \oplus (B \oplus C). \end{aligned}$$

Here, the XOR matches associative law.

Theorem 3 Let $A = B$, $A \oplus B = \overbrace{0000 \dots 0000}^{bits}$.

Proof According to Eq. (12), $A \oplus A = 0$, therefore $A \oplus B = \overbrace{0000 \dots 0000}^{bits}$.

Theorem 4 If A, B are both odd numbers, $(A) \oplus (-A) = \overbrace{1111 \dots 1110}^{bits}$, $(B) \oplus (-B) = \overbrace{1111 \dots 1110}^{bits}$. $(A \oplus B) = (-A \oplus -B)$.

Proof According to Theorem 3, if $A = B$, then $(A \oplus B) \oplus (-A \oplus -B) = \overbrace{0000 \dots 0000}^{bits}$. From Theorem 1 commutative law and Theorem 2 associative law, we rewrite this equation $(A \oplus B) \oplus (-A \oplus -B) = (A \oplus -A) \oplus (B \oplus -B)$.

According to Theorem 4, $A \oplus -A = B \oplus -B$.

From Theorem 3, $(A \oplus B) \oplus (-A \oplus -B) = \overbrace{0000 \dots 0000}^{bits}$.

$$\therefore (A \oplus B) = (-A \oplus -B).$$

Theorem 5 If A, B are even numbers, $4|A, B$ and $8|A, B$. $(A \oplus B) = (-A \oplus -B)$.

Proof We assume A and B are n bits numbers, when $4|A$ and $8|A$,

$$A = \overbrace{***** \dots 100}^{bits}. \quad \text{When } 4|B \text{ and } 8|B, B = \overbrace{***** \dots 100}^{bits}.$$

$$(A \oplus B) = \overbrace{***** \dots 000}^{bits}.$$

$$\text{Assume } -A = \overbrace{***** \dots 100}^{bits}, \quad \text{and} \quad -B = \overbrace{***** \dots 100}^{bits}.$$

$$(-A \oplus -B) = \overbrace{***** \dots 000}^{bits}.$$

$$(A \oplus B) \oplus (-A \oplus -B) = \overbrace{***** \dots 000}^{bits} \oplus \overbrace{***** \dots 000}^{bits} = 0.$$

$$\therefore (A \oplus B) = (-A \oplus -B).$$

3.2 Security Analysis

To Chien’s scheme, it exists a vulnerable which it combines XOR operation with two’s complement number system in computer, the related article be found in [3, 4, 7]. In the follows, we describe the practical issue that two variables do bitwise exclusive-or operation problem.

$$(y_i^{r_1}, (y_i^{r_2})^{2^j}) \Rightarrow \begin{cases} \text{both odd numbers, the } [pr = \frac{1}{4}]. \\ \text{one odd and even numbers, } [pr = \frac{1}{2}]. \\ \text{both even numbers, the } [pr = \frac{1}{4}]. \end{cases}$$

For Example (Tables 1 and 2).

The attacker Eve can easy to fake the valid parameters $(y_i^{r_1}, (y_i^{r_2})^{2^j})$ where $y_i^{r_1} \oplus (y_i^{r_2})^{2^j} \stackrel{?}{\equiv} (-y_i^{r_1} \oplus -(y_i^{r_2})^{2^j})$. She does follow steps:

Step 1. Eve sets $-r = -y_i^{r_1}$.

Step 2. Eve sets $-s = -(y_i^{r_2})^{2^j}$.

$$\begin{aligned} C_j &\equiv m_j \cdot (r \oplus s) \pmod{p} \\ &\stackrel{?}{\equiv} m_j \cdot (-r \oplus -s) \pmod{p}. \end{aligned} \tag{15}$$

From Theorems 1 to 5, we obtain $C_j \equiv m_j \cdot (r \oplus s) \pmod{p} \stackrel{?}{\equiv} m_j \cdot (-r \oplus -s) \pmod{p}$ are both odd numbers or even numbers where they matches specified rules. Now, it clearly describes from Eq. (15). Eve may forge successful for her attack.

Table 1 Example of both odd numbers

$(187)_{10} = (10111011)_2$
$(241)_{10} = (11110001)_2$
$(187)_{10} \oplus (241)_{10} = (01001010)_2$
$(187)_{10} \oplus (241)_{10} = (74)_{10}$
$(-87)_{10} = (111111101000101)_2$
$(-241)_{10} = (111111100001111)_2$
$(-187)_{10} \oplus (-241)_{10} = (000000001001010)_2$
$(-187)_{10} \oplus (-241)_{10} = (74)_{10}$

Table 2 Example of both even numbers

$(108)_{10} = (01101100)_2$
$(116)_{10} = (01110100)_2$
$(108)_{10} \oplus (116)_{10} = (00011000)_2$
$(108)_{10} \oplus (116)_{10} = (24)_{10}$
$(-108)_{10} = (111111110010100)_2$
$(-116)_{10} = (111111110001100)_2$
$(-108)_{10} \oplus (-116)_{10} = (00011000)_2$
$(-108)_{10} \oplus (-116)_{10} = (24)_{10}$

4 Conclusion

We clearly described an example of this case in the paper. If designer used bitwise exclusive-or operation in two's complement number system, it may cause others dangerous problem. Thus, the Hwang et al.'s scheme is insecure. From previous section, the author give this result.

Acknowledgements This work is partially supported by the project from Department of Education of Fujian Province under the number JA12351 with JK2013062 and also partially supported from Guandong province special funds to foster the students' science and technology innovation under the number 139785 and 139802 (Climbing program funds).

References

1. ElGamal T (1985) A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans Inf Theory* IT-31(4):469–472
2. Hwang MS, Chang CC, Hwang KF (2002) An ElGamal-like cryptosystem for enciphering large messages. *IEEE Trans Knowl Data Eng* 14:445
3. Liu C, Chen S, Sun S (2012) Security of analysis mutual authentication and key exchange for low power wireless communications. *Energy Proc* 17, Part A(0):644–649
4. Liu C, Zhang J (2009) Security analysis of zhang-wang digital signature scheme. *Commun CCISA* 15(4):24–29
5. Lyuu YD, Wu ML (2004) Cryptanalysis of an Elgamal-like cryptosystem for enciphering large messages. *WSEAS Trans Inf Sci Appl* 1(4):1079–1081
6. Wang MN, Yen SM, Wu CD, Lin CT (2006) Cryptanalysis on an Elgamal-like cryptosystem for encrypting large messages. In: *Proceedings of the 6th WSEAS international conference on applied informatics and communications*, pp 418–422
7. Xu L, Liu C, Wang N (2010) Comment on an improved signature without using one-way hash functions. In: *2010 International conference on cyber-enabled distributed computing and knowledge discovery (CyberC)*, pp 441–443

Study of Kindergartner Work Pressure Based on Fuzzy Inference System

Jie Fang

Abstract Since the change of economical level and population structure recently in Taiwan, a phenomenon of non-marriage and fertility declination is getting obvious. Therefore, parents emphasize the high quality of childcare education. This study mainly explores the relationship between childcare staffs working pressure as well as their efficiency in kindergartens and the management of children interpersonal conflict. With the results of a questionnaire and a rule base by specialists, the teacher working efficiency and pressure index can be obtained through the fuzzy inference system. From this study, we can conclude the childcare staff working pressure is not identical because of their various individual backgrounds and the management of children interpersonal conflict is totally different in each kindergarten. This study result can be submitted to administrators for reference.

Keywords Digital signature · One-Way hash function · Message redundancy · Forgery attack

1 Introduction

Education is a long-term program for a nation. The childcare education is even a foundation for the various grades of educations. The childcare staff attitude and behavior is the model for children. If the childcare staffs unfairly manage the interpersonal conflict between children because of their working pressure, the children in a kindergarten may have unhealthy personal relationship. Furthermore, if the failure case dealing with a child personal conflict results in factors for working pressure such as harmful interaction between childcare staffs and parents,

J. Fang (✉)

School of Electronics and Information Engineering, Fuqing Branch of Fujian Normal University, Fuqing 350300, China
e-mail: fangjie_1@foxmail.com

unrespectable environment in a kindergarten, and mutual rejection among colleagues, the teaching quality will be influenced over a long period of time. Therefore, we choose the childcare staffs servicing in the Taipei Area for this survey to explore the relationship between the working pressure and managing the children personal conflict [1]. After realizing the potential crisis in the childcare field, we can ponder a resolution for administrative decision of childcare education, cultivation of childcare teachers, and teaching environment as well as a reference of future research to enhance the quality of childcare education.

2 Review of Environment in Taiwan

Phylis Lan Lin shows in *Pressure and Adaptation*: the definition of pressure is that an individual will feel some grade or category of threat under some contexts. When an individual perceives pressure, he/she will feel uncomfortable. The uncomfortable degree will be different among humans because of an individual's body health, culture, social value, and regulation. Therefore, an individual must pay additional vigor to keep the balance of body and mind. Selye makes researches about pressure and believes that the reaction from the stimulation of various pressure has a common nature that an individual has a request to recover the normal status; an individual must exhaust extra physical or mental energy to respond when this pressure source menaces an individual [2]. By this hypothesis, the feeling of pressure is not identical for each individual but the additional load from pressure is applicable to everybody. When we inspect the related documents about teachers working pressure and solution in domestic or overseas researches, the objects in studies are teachers in primary schools, junior high schools, or colleges but there are few researches about childcare staffs in kindergartens. Meanwhile, these contents of researches emphasize the situations in working pressure but there is no exploration in the childcare staffs working pressure and managing children's interpersonal conflict.

2.1 Nouns Definition

For distinctly define the changeable meaning in this study, the conceptual definition is explained below:

Childcare staff The childcare staffs are personnel servicing in the public or private kindergartens around Taipei Area (Taipei City, Taipei County, and Keelung City). They might be administrative personnel for childcare, teachers, teachers with administration tasks, and assistant childcare staffs.

Working Pressure and Status Beehr and Newman believe that working pressure is a context that a worker is forced to deviate the normal situation in body and mind because the interaction between related working factors and a worker alters a worker's mental and physical status. Caplan believes that working pressure is a stressful phenomenon on an individual because of some characteristics in the working environment. Yung-Chen Huang expresses that working pressure is an unbalanced status because pressure from working factors prompts the quest of external circumstance and internal ability of an individual to produce difference in mutual action. By the above mentioned descriptions, working pressure is a status that the mutual influence between "related working factors" and "an individual" stimulates an individual to produce adaptive reaction. If an individual cannot control stress from contextual factors or the feeling of difference of unbalance, working pressure will occur. Childcare staffs pressure is that they feel the negative feeling such as nervousness, threat, upset, anxiety, and unhappiness, when they work in teaching, however, face the menace in workplace and fail to satisfy themselves expectation. The reacting course of action, feeling, spirit, or physiology that influences the working performance will be different because of an individual's previous experience, personal characteristics, or specific recognition.

Interpersonal conflict between children The simple definition for the interpersonal conflict between children is "when a child does or says something to stimulate another child's objection, the conflict occurs". A more sophisticated explanation is the occurrence of a conflict needs three dimensions:

1. Previous event simulating the conflict.
2. Opposite opinion on both sides or unilateral side.
3. Object to the opposite opinion.

A strategy for conflict management A strategy for conflict management is that a teacher has to choose a method to resolve the conflict or complaint between children because of disagreement of opinion. By the score distribution in the questionnaire of "A strategy that a teacher deals with the interpersonal conflict between children" edited by Liang-Yin Lin, the strategy for conflict management is divided into four categories, "problem resolution", "mediation", "indirect intervention", and "authoritative arbitration". Their definitions are:

1. A strategy of problem resolution: Conflicting parties face their problems, frankly communicate, and integrate opinions from both sides to get a satisfactory formula for solving a problem or obtaining a win-win effect on both sides.
2. A strategy of mediation: The preschool staffs, as a third party, mediate the conflict between children. The staffs should: find the essential of problems, listen but do not criticize, suggest a method to mediate the different opinions, and retain an objective stance. The methods are compromise, negotiation, satisfaction, reciprocation, concession, reminder, and suggestion.

3. A strategy of indirect intervention: The preschool staffs do not directly deal with the conflict between children but retain the right of resolution to children themselves after they make a complaint. (For example, ask children to represent, think, and transmit teachersreminder.) The methods are: assign children themselves to resolve or ask the third party (children with good social skills, parents, or colleagues) to handle.
4. strategy of authoritative arbitration: Ask children to respect and comply with the authority from teachers. The preschool staffs should use some skillful strategies to deal with childrens conflict for repressing it but they cannot eliminate the origin resulting in this conflict. The methods are remove or replacement of conflict, isolation, and authoritative order.

2.2 A Table for Major Regions of Pressure Source

It is not surprised that the pressure on a busy administrator could be multiplied in virtue of intensive competition. Kafry and Pines show that the most positive way to face the pressure for an individual is located on zone 1 in the matrix below among many methods to avoid or solve pressure (Table 1).

By the aforementioned pressure from each dimension and the trend of intensive competition, the childcare staffs have to adequately manage self-pressure to prevent mental collapse. This study will explore the major reasons of pressure formation.

2.3 Study Limited and Rang

Range of the study: Taipei City, Taipei County, and Keelung City. Defined objects of the study: childcare staff of the public or private kindergartens in Taipei Area.

Table 1 Region of pressure source

	Action	Avoidance
Direct	Zone 1: change the source of pressure, face pressure, and overcome it with a positive attitude	Zone 2: ignore the source of pressure, avoid pressure, and flee
Indirect	Zone 3: discuss the source of pressure, change yourself, and focus on other things to transfer the attention	Zone 4: addict alcohol and drug, have a sickness, and be listless

3 Questionnaire Design

3.1 Questionnaire Structure

Design the structure of a questionnaire by the background of an individual, the status of job, and the conflict between children. The scheme is shown in Fig. 1.

3.2 Study of Sampling

The target population of the sampling in this study is the childcare staff in the public and private kindergartens of Taipei Area (Taipei City, Taipei County, and Keelung City). The number of issued questionnaire is 300 and feedback of them is 277. The effective questionnaire is 265.

3.2.1 Basic Information of an Individual Teacher

- Educational background: senior high school, senior vocational school, college, university, and graduate school.

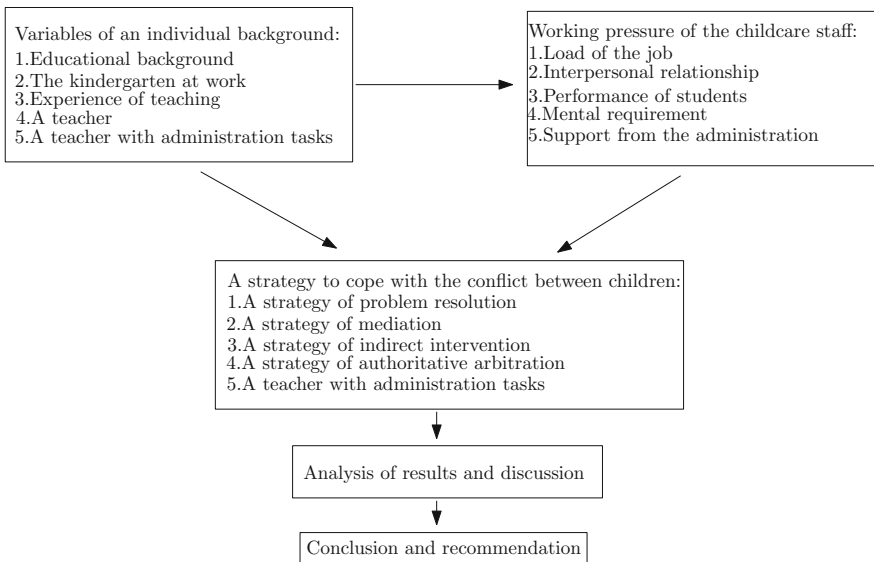


Fig. 1 The structure of a questionnaire

- Nature of a kindergarten: the public kindergarten (the primary school with a kindergarten) and the private kindergarten.
- Scale of a kindergarten: below 5 classes, 6–10 classes, 11–15 classes, 16–20 classes, and 21 classes or more.
- Working experience: 1–5 years, 6–10 years, 11–15 years, and 16 years or more.
- Present duty: a teacher, an administrator, or a teacher with administrative tasks.

A teacher's working pressure and status table The main structure is based on the kindergarten teachers' teaching and their working pressure table in the questionnaire for adaptation of a teacher's body and mind. The researcher divides a teacher's working status into five categories: load of the job, interpersonal relationship, performance of students, support from the administration, and mental requirement. The categories are explained below [3]:

- Load of the job: the feeling a teacher has for load of the job.
- Interpersonal relationship: the interaction between the teacher and the parents, colleagues, or children.
- Performance of students: each student's behavior or performance in the schoolwork.
- Support from the administration: support from the principal and the administration.
- Mental requirement: a request that a teacher believes in her profession.

A strategy that a teacher cope with the conflict between children There are total 30 topics in the questionnaire. The researcher divides the strategy into communication, mediation, nonintervention, and authoritative arbitration. The question number is listed below [4]:

- Communication: number 4, 10, 15, 16, 22.
- Mediation: number 3, 5, 6, 9, 11, 21, 23, 27, 29.
- Nonintervention: number 7, 12, 14, 18, 19, 24, 30.
- Authoritative arbitration: number 2, 8, 14, 16, 20, 26, 28.
- The finding is inducted by integrating the results:
 - (1) There is remarkable relationship between the individual backgrounds of childcare staffs in the kindergartens and their working pressure.
 - (2) There is difference in working pressure among childcare staffs with different educational backgrounds.
 - (3) There is difference in working pressure among childcare staffs working in kindergartens with different nature.
 - (4) There is difference in working pressure among childcare staffs working in kindergartens with different scale.
 - (5) There is difference in working pressure among childcare staffs with different working experience.
 - (6) There is difference in working pressure among childcare staffs with different duties.
 - (7) There is relationship between working pressure and handling the interpersonal conflict between children.

4 An Assessment of Working Efficiency Using Fuzzy Logic Inference

4.1 The Concept of the Fuzzy Logic and Inference

The development of various researches applying the fuzzy logic theory has been very rapid in recent tens of years since Zadeh [5], the founder of the fuzzy logic, proposed the thesis, Fuzzy Sets, in 1965. The issues about Uncertainty and Imprecision can be solved by the fuzzy logic. It can simulate the human language and thinking to closely and efficiently deal with a question. The logic of a human thinking is very swift and flexible. For simulating a humans thinking logic in a policy decision or assessment, it is necessary to use a mathematic model with a humans thinking logic. Therefore, if we can use some rules to describe the procedure of decision making for some specific cases, the fuzzy logic will become an intact solution by efficiently applying the knowledge. In a word, the fuzzy logic is an applied science of transferring the description of the natural human language in thinking of policy decision to a mathematic model. This model includes three major procedures: fuzzification, inference, and defuzzification. For example, an intact structure of the fuzzy logic shows in Fig. 2. When all input variables transfers to oral languages, a set of rules describing by the type of “If-Then” will infer the final assessment result in the procedure of the “fuzzy inference”. And the results will be other language values of variables of output language. An assessment system about working pressure can be constructed by a suitable rule that specialists build for the questionnaire and the fuzzy logic inference so an administrator can fully realize his/her teachers pressure condition at work.

4.2 Establishment of a Specific Function and a Rule Base

We use Mat-Lab 7.0 as a tool to sequentially build several membership functions by an individual background, working status, and conflict management. The results are shown in Fig. 3.

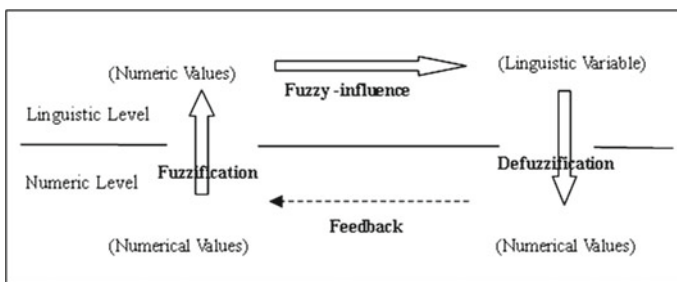


Fig. 2 Flow char of fuzzy inference model

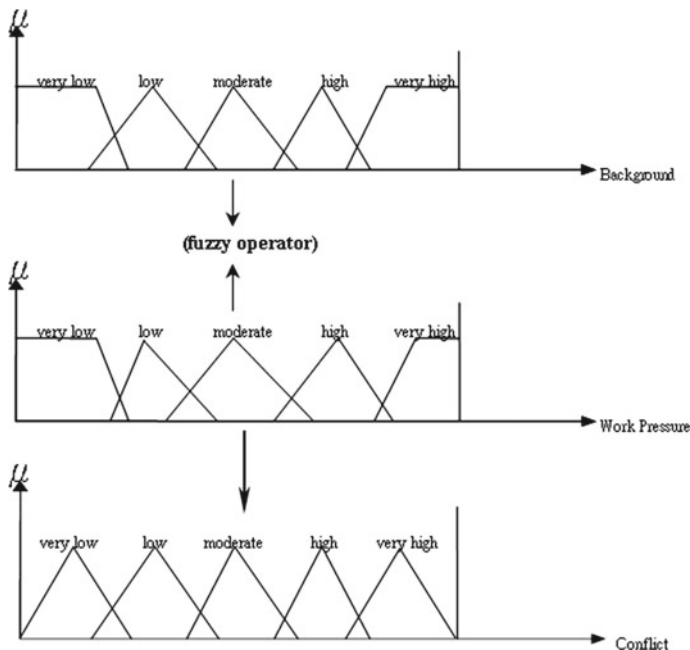


Fig. 3 Membership function

An individual background represents his/her education as well as working experience and they are divided into three grades. Working status is an employees performance in workplace and they are divided into three grades. The interpersonal conflict includes the interaction with students parents, children, and colleagues and they are also divided into three grades. There are five grades for the working efficiency [6]. Therefore, there are 27 rules (3 cubed) mapping to 5 grades. By experience and suggestion of specialists, we build a set of rule base by the factors which influence working efficiency. The intact base is listed in Table 2.

If (Personal Background = Good) and (Active Status = Poor) and (Personnel Conflict = Moderate) then (Working Press = Poor) For example: if a teacher with a perfect individual background has a poor working status and her interpersonal conflict management is moderate, her working efficiency may be not ideal. The educational background of an individual is not a critical factor. For example, if a teacher with a moderate educational background has a good working status and a moderate interpersonal conflict management; her working efficiency may be good. The rule number 20 in the rule base clearly indicates it.

Table 2 A rule base for working efficiency

Linguistic variables								
Rule No	Personal background		Active status		Personal management		Working efficiency	
Linguistic values								
1	Good	1	Good	1	Good	1	Very high	2
2	Good	1	Good	1	Moderate	0	High	1
3	Good	1	Good	1	Poor	-1	Low	-1
4	Good	1	Poor	-1	Good	1	Moderate	0
5	Good	1	Poor	-1	Moderate	0	Low	-1
6	Good	1	Poor	-1	Poor	-1	Very low	-2
7	Good	1	Moderate	0	Good	1	Low	-1
8	Good	1	Moderate	0	Moderate	0	Moderate	0
9	Good	1	Moderate	0	Poor	-1	Low	-1
10	Poor	-1	Good	1	Good	1	Very high	2
11	Poor	-1	Good	1	Moderate	0	High	1
12	Poor	-1	Good	1	Poor	-1	Moderate	0
13	Poor	-1	Poor	-1	Good	1	High	1
14	Poor	-1	Poor	-1	Moderate	0	Low	-1
15	Poor	-1	Poor	-1	Poor	-1	Very low	-2
16	Poor	-1	Moderate	0	Good	1	High	1
17	Poor	-1	Moderate	0	Moderate	0	Moderate	0
18	Poor	-1	Moderate	0	Poor	-1	Low	-1
19	Moderate	0	Good	1	Good	1	Very high	2
20	Moderate	0	Good	1	Moderate	0	High	1
21	Moderate	0	Good	1	Poor	-1	Low	-1
22	Moderate	0	Poor	-1	Good	1	Moderate	0
23	Moderate	0	Poor	-1	Moderate	0	Low	-1
24	Moderate	0	Poor	-1	Poor	-1	Very Low	-2
25	Moderate	0	Moderate	0	Good	1	Low	-1
26	Moderate	0	Moderate	0	Moderate	0	Moderate	0
27	Moderate	0	Moderate	0	Poor	-1	Low	-1

5 Conclusion

Since the reduction of fertility rate is obvious recently in Taiwan, most parents spoil their children excessively. Besides, for reinforcing the internal management and enhancing teachers efficiency, the kindergartens assign many tasks to their teachers. The pressure which teachers must face includes not only from their administrators and parents but also their professional skills, working experience, and conflict management between children. This study for the questionnaires written by teachers indicates teachers working pressure and efficiency through a rule base established

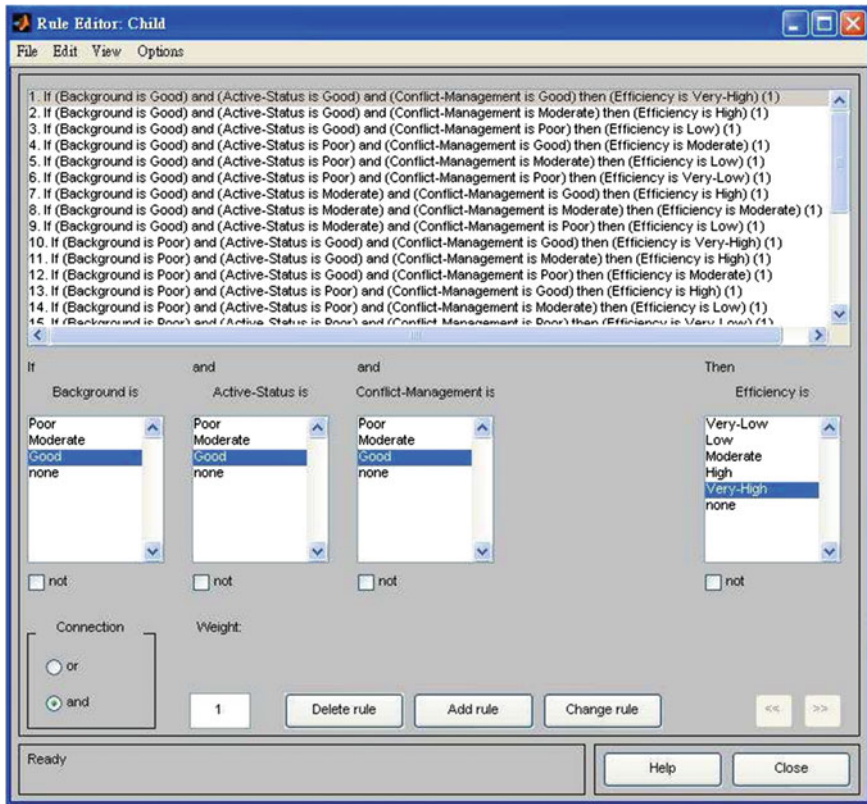


Fig. 4 Rule editor

by specialists and the fuzzy inference system. However, we still not analyze the relationship among teachers working environments, working pressure, and the interpersonal conflict of children. We hope that this exploration will be included in the future research to make up the deficiency in this study (see Fig. 4).

Acknowledgements The authors would like to thank the reviewers for their comments that help improve the manuscript. The author also thanks Chenglian Liu for his useful suggestion.

References

1. Akhter F, Hobbs D, Maamar Z (2003) How users perceive trust in virtual environment. In: International conference on information and knowledge engineering. Las Vegas, Nevada, pp 23–26
2. Black M (1937) Vagueness: an exercise in logical analysis. *Philos Sci* 4:427–455
3. Cox E (1994) *The fuzzy systems handbook a practitioner's guide to building, using, and maintaining fuzzy systems*. Academic Press, Cambridge

4. David JL (2000) You've got surveys. *Am Demographics* 42–44
5. Zadeh L (1965) Fuzzy sets. *Inf Control* 8:338–353
6. Lukasiewicz J (1970) Philosophical remarks on many-valued systems of propositional logic
reprinted in selected works. North-Holland, Amsterdam

Comment on ‘The Hermite-Hadamard Inequality for R -Convex Functions’

Zhi-Pan Wu

Abstract We found an error in the theorem 2.3 from original paper [Gholamreza Zabandan, Abasalt Bodaghi and Adem Kilicman. *Journal of Inequalities and Applications*, DOI: [10.1186/1029-242X-2012-215](https://doi.org/10.1186/1029-242X-2012-215)]. Hence, the correct statement of the equation is provided in this paper.

Keywords Hermite-Hadamard inequality · Convex functions

1 Introduction

Zabandan et al. [1] presented their contribution the Hermite-Hadamard inequality for r -convex functions. To Theorem 2.3, there is an error. The author therefore corrects the equation as follows.

Theorem 2.3 Let $f : [a, b] \rightarrow (0, \infty)$ be r -convex and $r \geq 0$. Then the following inequalities hold:
$$\frac{1}{b-a} \int_a^b f(x) dx \leq \begin{cases} \frac{r}{r+1} \left(\frac{f^{r+1}(b) - f^{r+1}(a)}{f^r(b) - f^r(a)} \right), & r \neq 0, \\ [f(b) - f(a)] \ln \frac{f(b)}{f(a)}, & r = 0. \end{cases}$$

Proof First, let $r > 0$. Since f is r -convex, for all $t \in [0, 1]$, we have $f(ta + (1-t)b) \leq [tf^r(a) + (1-t)f^r(b)]^{\frac{1}{r}}$.

It is easy to observe that

Z.-P. Wu (✉)

Department of Computer Science, Huizhou University, Huizhou 516007, China
e-mail: hz_wzp@163.com

$$\begin{aligned} \frac{1}{b-a} \int_a^b f(x)dx &= \int_0^1 f(ta + (1-t)b)dt \\ &\leq \int_0^1 [tf^r(a) + (1-t)f^r(b)]^{\frac{1}{r}} dt \\ &= \int_0^1 [t(f^r(a) - f^r(b)) + f^r(b)]^{\frac{1}{r}} dt \end{aligned}$$

By substitution $t(f^r(a) - f^r(b)) + f^r(b) = z$, we obtain

$$\begin{aligned} \frac{1}{b-a} \int_a^b f(x)dx &\leq \frac{1}{f^r(b) - f^r(a)} \int_{f^r(b)}^{f^r(a)} z^{\frac{1}{r}} dz \\ &= \frac{1}{f^r(b) - f^r(a)} \cdot \frac{1}{1 + \frac{1}{r}} \left[z^{1 + \frac{1}{r}} \right]_{f^r(a)}^{f^r(b)} \\ &= \frac{r}{r+1} \left(\frac{f^{r+1}(b) - f^{r+1}(a)}{f^r(b) - f^r(a)} \right). \end{aligned}$$

For $r = 0$, we have $f(tx + (1-t)y) \leq [f(x)]^t [f(y)]^{1-t}$.
 Since,

$$\begin{aligned} \frac{1}{b-a} \int_a^b f(x)dx &= \int_0^1 f(ta + (1-t)b)dt \\ &\leq \int_0^1 [f(a)]^t [f(b)]^{1-t} dt \\ &= f(b) \int_0^1 \left[\frac{f(a)}{f(b)} \right]^t dt \\ &= f(b) \left[\frac{f(a)}{f(b)} \right]^t \ln \frac{f(a)}{f(b)} \Big|_0^1 \\ &= [f(b) - f(a)] \ln \frac{f(b)}{f(a)}. \end{aligned}$$

Correction equation $f(b) \left[\frac{f(a)}{f(b)} \right]^t \ln \frac{f(a)}{f(b)} \Big|_0^1$ is used instead of $f(b) \left[\frac{f(a)}{f(b)} \right]^t / \ln \frac{f(a)}{f(b)} \Big|_0^1$.

The detailed be description following.

Let $x = ta + (1-t)b$, since $x = a, t = 1$; while $x = b, t = 0$.

$$\begin{aligned}
 \frac{1}{b-a} \int_a^b f(x)dx &= \frac{1}{b-a} \int_0^1 f(ta + (1-t)b)d(ta + (1-t)b) \\
 &= \frac{1}{b-a} \int_0^1 f(ta + (1-t)b) \cdot (a-b)dt \\
 &= \int_0^1 f(ta + (1-t)b)dt \\
 &\leq \int_0^1 [f(a)]^t [f(b)]^{1-t} dt \\
 &= f(b) \int_0^1 \left[\frac{f(a)}{f(b)} \right]^t dt \\
 &= f(b) \left[\frac{f(a)}{f(b)} \right]^t / \ln \frac{f(a)}{f(b)} \Big|_0^1 \\
 &= \frac{f(b) - f(a)}{\ln \frac{f(b)}{f(a)}}.
 \end{aligned}$$

We can rewrite another proof under formula.

$$\begin{aligned}
 \text{Let } \frac{f(a)}{f(b)} &= k. \\
 \frac{1}{b-a} \int_a^b f(x)dx &= \int_0^1 f(ta + (1-t)b)dt \\
 &= f(b) \int_0^1 (k)^t dt \\
 &= f(b) \int_0^1 (e^{\ln k})^t dt \\
 &= f(b) \int_0^1 e^{t \ln k} \\
 &= f(b) \frac{1}{k} \cdot e^{t \ln k} \Big|_0^1 \\
 &= f(b) \frac{1}{k} k^t \Big|_0^1.
 \end{aligned}$$

2 Conclusion

As the original erratum in theorem 2.3, Zabandan et al.’s results derived an error in their computation. The wrong result, may gets an incorrect assumption. The author gives this correction for his presentation.

Acknowledgements The authors would like to thank the reviewers for their comments that help improve the manuscript. The author also thanks Chenglian Liu for his useful suggestion.

Reference

1. Zabandan G, Bodaghi A, Kilicman A (2012) The hermite-hadamard inequality for r-convex functions. *J Inequalities Appl* 2012:1–8. doi:[10.1186/1029-242X-2012-215](https://doi.org/10.1186/1029-242X-2012-215)

Author Index

A

Alismaili, Salim, 597
An, Zexin, 451

C

Cai, Zhaoquan, 373
Cao, Hongbing, 1131
Chang, Che-Chang, 1185
Chang, Cheng-Ming, 1091
Chang, Jieh-Ren, 1065
Chang, Kuan-Wu, 899
Chang, Lu, 865
Changn, Jieh-Ren, 1073
Chang, Ya-Chen, 681
Chan, Ho-Hsiang, 937
Chan, Ta-Chien, 899
Chan, Wei-Chen, 617
Chan, Yu-Wei, 617, 743
Chen, Chih-Chen, 661, 671
Chen, Ching-Chung, 783
Chen, Ching-Ken, 713
Chen, Der-Chin, 721, 731, 773, 783
Chen, Dongmin, 15
Chen, Fang-Tzu, 1185
Cheng, Chung-Yu, 993
Chen, Guey-Shya, 801
Chen, Haiyan, 7, 15
Chen, Huan-Chung, 1065
Chen, Hui-Qian, 801
Chen, Jianqi, 395
Chen, Jing-Min, 705, 713
Chen, Jin-Jia, 743
Chen, Jiun-Ting, 681
Chen, Jun-Yan, 1
Chen, Lu, 459
Chen, Mu-Song, 651, 697
Chen, Ping-I, 917
Chen, Pin-Liang, 927
Chen, Wei, 791, 865

Chen, Wei-Hsin, 773
Chen, Wenbo, 441
Chen, Wuhui, 583
Chen, Yong, 97, 1167
Chen, You-Shyang, 1023, 1033, 1041, 1051, 1059
Chen, Yu-Luen, 661, 671
Chen, Yung-Hui, 1, 117, 541
Chen, Yu Wen, 503
Chin, Kai-Yi, 363
Chiu, Chiung-Lin, 1051, 1059
Cho, Ta-Hsiung, 761
Chou, Chih-Hong, 661
Chou, Jerome Chih-Lung, 985, 1017
Chou, Shih-Wei, 671
Chou, Wen-Kuang, 551
Chuang, Chu-Chun, 1
Chuang, Huan-Ming, 1023, 1033, 1041
Chu, Chi Nung, 959, 1007

D

Deng, Lawrence Y., 117
Ding, Min-Hui, 801
Ding, Ning, 325
Ding, XianShu, 43
Dong, MingChui, 379, 491
Duan, Yanlin, 841, 855
Du, Huan, 1159, 1173

F

Fang, Jie, 1225, 1233
Feng, Guiliang, 161, 185, 389
Feng, Shuo, 185
Fu, BinBin, 379

G

Gao, Rui, 459
Gao, Yingning, 215
Gou, Jianping, 43, 103

Guo, Jian, 225, 231
 Guo, Jianhua, 173
 Guo, Ran, 491
 Guo, Xifeng, 515

H

Han, Deqiang, 417
 Hao, Li, 405, 451, 515
 Ho, Chih-Ching, 975
 Ho, Ching-Wei, 77
 Hong, Guan-Syuan, 753
 Hong, Zi-Min, 705
 Ho, Tze-Yee, 651
 Hsiao, Chin Tsai, 503
 Hsiao, Heng-Chih, 661
 Hsieh, Hsiang-Chin, 363
 Hsieh, Meng-Yen, 551
 Hsieh, Ya-Hui, 721, 731
 Hsu, Sung-Pin, 671
 Huang, Chien-Sheng, 753
 Huang, Ching-Lien, 1, 117, 541
 Huang, Chun-Hong, 881
 Huang, Hailang, 1123
 Huang, He-Liang, 1209
 Huang, Jun, 491
 Huang, Kuang-Lung, 743
 Huang, Li, 1201
 Huang, Rongsheng, 389
 Huang, Shengjian, 817
 Huang, Sun-Jen, 1097
 Huang, Yuanyuan, 205, 471
 Huang, Yung-Fa, 689
 Huang, Yu Ting, 959
 Hu, HaiFeng, 65
 Hui, Lin, 85
 Hung, Chan-Hsiang, 651
 Hung, Che-Lun, 53, 239
 Hung, Chia-Liang, 985
 Hung, Hsiu-Yen, 1085
 Hung, Jason C., 523, 801, 881
 Huo, Li-fang, 353
 Hu, XiangYang, 491
 Hwang, Chi-Pan, 651, 697

J

Jiang, Hong, 1123
 Jiang, Jiulei, 261
 Jiang, Yan-Ru, 1
 Jie, Yang, 155
 Juang, Hung-chi, 1073

K

Kao, Jui-Hung, 899
 Kao, Wen-Hsing, 279, 289, 801

Kumara, Banage T.G.S., 583
 Ku, Tsun, 927

L

Lai, Feipei, 899
 Lai, Ruey-Gang, 949
 Lan, Anyi, 389
 Lee, Ko-Fong, 363
 Lee, Mike Y.J., 985
 Lee, Min-Feng, 801
 Lee, Shih-Chieh, 721, 731, 773, 783
 Lee, Wei-Chun, 117
 Lei, Zhou, 173
 Liang, Chao, 173
 Liang, Junhua, 141, 311, 531
 Liang, Song, 865
 Liao, Ming-Hong, 551
 Liaw, Horng-Twu, 887, 975
 Li, Baoliang, 481
 Li, Bin, 1123, 1167
 Li, Bo, 389
 Li, Gui-Lin, 551
 Li, Jing, 337
 Li, Kuan-Ching, 551
 Li, Kuo-Pin, 279, 289
 Li, Liangyi, 1167
 Li, Mengxiang, 597
 LI, Min, 301
 Lin, Bo-Cheng, 899
 Lin, Chien-Ku, 1023, 1033, 1041
 Lin, Chin-Hung, 617
 Lin, Ching-Huang, 697, 753, 761
 Lin, Chun-Yuan, 53
 Lin, Chyuan-Yuh, 1041
 Lin, Edgar Chia-Han, 575
 Lin, Hong-Wun, 1065
 Lin, Hsien-Chang, 761
 Lin, Kuan-Cheng, 523
 Lin, Ping, 429
 Lin, Renyi, 865
 Lin, Ruei-Yang, 289
 Liou, Bo-Shen, 289
 Liu, Chenglian, 1225
 Liu, Fangfang, 7, 15
 Liu, Feijing, 561
 Liu, Haijian, 865
 Liu, Minghe, 405
 Liu, Naidi, 531
 Liu, Te-Shu, 743
 Liu, Tongjuan, 841, 855
 Liu, Wei, 325, 561
 Liu, Wenjuan, 1115
 Liu, Wenwen, 261
 Liu, Yang, 141, 311, 337, 1147

- Liu, Yaxu, 141, 337
 Liu, Yaya, 261
 Liu, Yingqi, 841, 855
 Liu, Zongtian, 325
 Li, Wei-Chiang, 525
 Li, Weimin, 195, 215
 Li, Xunfeng, 195
 Li, Yen-Chen, 671
 Li, Zhenni, 585
 Li, Zhenyu, 1147, 1153
 Lo, Chih-Min, 1085
 Lo, Chin-Wen, 279
 Luo, Weizhong, 373
 Luo, Xiangfeng, 1179
 Luo, Xin, 471
 Lu, Yiping, 161, 185
- M**
 Ma, Hongxing, 103
 Ma, Jinfeng, 1107
 Ma, Ying, 43
 Ma, Zhaohui, 53
 Mei, Lin, 271, 1147, 1153, 1159
- N**
 Ngorsed, Manoon, 627
 Nian, Yanyun, 459
- P**
 Paik, Incheon, 583
 Pan, Jin-Gu, 917
- Q**
 Qian, Quan, 239
 Qi, Liang, 817, 827
 Qin, Chenglei, 1131
 Qin, Jing, 161, 185
 Qiu, Dong, 1193
- R**
 Rao, YunBo, 43
 Renjie, Wu, 133
- S**
 Sekar, Booma Devi, 491
 Shang, Zhihao, 441
 Shen, Jun, 597
 Shih, Ying-Ying, 671
 Song, Junjie, 827
 Song, Zhenwen, 865
 Suesaowaluk, Poonphon, 627
 Su, Gon-Jong, 993
 Sun, Peng, 531
 Sun, Wei-Zen, 9899
- Sun, Xinghua, 451, 515, 531
 Sun, Yuan, 225, 231
 Su, Pin-Yu, 1
 Su, Zhen, 811
 Syu, Han-Ci, 1
- T**
 Tanaka, Takazumi, 583
 Tang, Hengliang, 225, 231
 Tang, Qianjin, 271, 1107
 Tang, Yang, 791
 Tang, Zhiwei, 97, 1123, 1167
 Tan, Yue, 325, 561
 Tao, Ye, 251
 Tian, Xining, 441
 Tian, Yuan, 429
 Tsai, Fang-Chi, 1
 Tsai, Jui Pin, 503
 Tsai, Tzu-Chieh, 937, 993
 Tsan, Yu-Tse, 617
 Tseng, Cheng-Chieh, 761
 Tseng, Chun-Hsiung, 1, 117, 541
- V**
 Vai, Mang I., 379
- W**
 Wang, Changsheng, 173
 Wang, Cong, 301
 Wang, Hsiao-Hsi, 53
 Wang, Hsuan-Fu, 651, 661, 671, 689, 697,
 753, 761
 Wang, Jiawei, 141
 Wang, Kuei Min, 85
 Wang, Lung-Cheng, 541
 Wang, Shih-Chuan, 661
 Wang, Xiangrong, 481
 Wang, Xiaosheng, 1139
 Wang, Xiaoyu, 451
 Wang, Xili, 103
 Wang, Xu, 561
 Wang, Xuan, 311
 Wang, Yu-Bing, 77
 Wang, Zongxia, 417
 Wan, Jiang-Feng, 429
 Wan, Tian-Long John, 541
 Wei, Xiao, 15, 1131
 Weng, Jing-De, 881
 Weng, Martin M., 881
 Weng, Shaolin, 639
 Wen, Zhiqiang, 97
 Wu, Bo-Han, 1097
 Wu, Hongxi, 405
 Wu, Jieling, 1225

Wu, Qiu-Ming, 1193
 Wu, Wei-Chen, 887
 Wu, Yixuan, 271, 1107
 Wu, Zhi-Pan, 1217, 1245
 Wu, Zhizong, 271, 1107
 Wu, Zongheng, 639

X

Xiaoling, Guo, 133, 155
 Xiao, Zhang, 133, 155
 Xiong, Taisong, 205, 471
 Xue, Li, 1131
 Xu, Yi, 215
 Xu, Zheng, 7, 15, 271, 1115, 1139, 1147,
 1153, 1159, 1173

Y

Yang, Chang-Lin, 541
 Yang, Cheng-Chih, 689
 Yang, Guanghui, 429
 Yang, Hongbin, 173
 Yang, Hsien-Wei, 279
 Yang, Jeng-Chi, 289
 Yang, Lei, 395, 429
 Yang, Liangbin, 33
 Yang, LiangBin, 65
 Yang, Mei-Fang, 1051
 Yang, Ping-Che, 927
 Yang, Po-Hui, 705, 713
 Yang, Yan, 301
 Yan, Jeng-Chi, 279
 Yan, Xuesong, 395
 Yan, Zhiguo, 1153, 1159, 1173
 Yao, HaoDong, 379
 Yao, Shixuan, 481

Yeh, Feng-Ming, 721, 731, 773, 783
 Ye, Yongfei, 337, 405, 451, 531
 Yingxu, Lai, 251
 Yin, Xiaobo, 1179
 Yiwen, Zhao, 251
 Yuan, Wei, 811
 Yu, Cheng-Hsien, 949, 965
 Yu, Feng, 827
 Yu, Zhi-ting, 239

Z

Zeng, Qingjun, 865
 Zeng, Zhi, 373
 Zhang, Chunlei, 515
 Zhang, Li-ming, 353
 Zhang, Qiang, 417
 Zhang, Rui, 239
 Zhang, Sheng, 429
 Zhang, Shunxiang, 1115, 1139
 Zhang, Xiao, 141, 161, 311, 389, 405
 Zhang, Xunchao, 395
 Zhang, Yaling, 395
 Zhang, Yujia, 325, 561
 Zhao, Mingru, 225, 231
 Zhao, Xian, 827
 Zhao, Xi-qing, 353
 Zhao, Zhangzhi, 337
 Zhao, Zhisheng, 141, 311, 405
 Zheng, Chun-Hua, 1209
 Zheng, Huiru, 639
 Zheng, Lingxiang, 639
 Zhou, Qingguo, 395, 429, 459
 Zhou, Qingyuan, 7
 Zhou, Wencheng, 639
 Zhou, Xinli, 23, 65