Lecture Notes in Electrical Engineering 378

Yong Qin
Limin Jia
Jianghua Feng
Min An
Lijun Diao
Editors

Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation

Transportation

Springer

Lecture Notes in Electrical Engineering

Volume 378

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

Yong Qin · Limin Jia · Jianghua Feng
Min An · Lijun Diao
Editors

Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation

Transportation

Springer

Editors
Yong Qin
Beijing Jiaotong University
Beijing
China

Min An
University of Birmingham
Birmingham
UK

Limin Jia
Beijing Jiaotong University
Beijing
China

Lijun Diao
Beijing Jiaotong University
Beijing
China

Jianghua Feng
CSR Zhuzhou Institute CO., LTD.
Zhuzhou
China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-3-662-49368-7 ISBN 978-3-662-49370-0 (eBook)
DOI 10.1007/978-3-662-49370-0

Library of Congress Control Number: 2016930678

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer-Verlag GmbH Berlin Heidelberg

Contents

Contributors

Min An State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Qi An State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Xuehui An School of Civil Engineering, Tsinghua University, Beijing, China

Yazheng An Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Chang aixin Shijiazhuang Tiedao University, Shijiazhuang, China

Lifeng Bi State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Guoqiang Cai State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Guanghui Cao Henan Provincial Communications Planning, Survey and Design Institute Co., Ltd., Zhengzhou, China

Dawei Chen College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Jun Chen State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Kun Chen Tianjin Key Laboratory of Civil Aircraft Airworthiness and Maintenance, Civil Aviation University of China, TianJin, China

Ling Chen School of Physics and Telecommunications Engineering, South China Normal University, Guangzhou, China

Meihuan Chen College of Information Science and Engineering, Yanshan University, Qinhuangdao, China

Mingdian Chen State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Haidian District, Beijing, China

Peijiang Chen School of Automobile, Linyi University, Linyi, Shandong, China

Wenman Chen Management School, Wuhan University of Technology, Wuhan, People's Republic of China

Xiaojun Chen School of Economic and Management, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

Xuhong Chen CRRC Research of Electrical Technology and Material Engineering, Zhuzhou, Hunan, China

Shanmei Cheng School of Automatic, Huazhong University of Science and Technology, Wuhan, China

Xiaoqing Cheng State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Collaborative Innovation Center of Railway Traffic Safety, Beijing Jiaotong University, Beijing, China; Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Po Dai China Railway Shanhaiguan Bridge Group Co., Ltd., Qinhuangdao, China

Bo Deng Beijing Institute of System Engineering, Beijing, China

L. Deng Faculty of Information Science and Technology, Dalian Maritime University, Dalian, China

Xiaojun Deng CSR Qingdao Sifang Co., Ltd., Chengyang District, Qingdao, China; CRRC Qingdao Sifang Co., Ltd, Qingdao, China

Jianlong Ding State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Honghui Dong State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Yan Dou College of Information Science and Engineering, Yanshan University, Qinhuangdao, China

Shaohua Du CRRC Research of Electrical Technology and Material Engineering, Zhuzhou, Hunan, China

Yuejuan Du Tianjin Key Laboratory of Civil Aircraft Airworthiness and Maintenance, Civil Aviation University of China, TianJin, China

Shuai Feng State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, Haidian, China

Yun Fu China Academy of Railway Sciences, Beijing, China

Yunxiao Fu State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Collaborative Innovation Center of Railway Traffic Safety, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Jianhong Gao Mechanical Engineering School, Suzhou University of Science and Technology, Suzhou, Jiangsu, China

Xiaoxun Gao China Academy of Railway Sciences, Beijing, China

Xudong Gao School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China

Bo Gong Wuhan Institute of Marine Electric Propulsion, CSIC, Wuhan, China

Xue Han School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Dakuo He College of Information Science and Engineering, Northeastern University, Shenyang, China

Yongming He School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Bin Hong Beijing Institute of System Engineering, Beijing, China

Yiyu Hou Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Changzhen Hu Beijing Institute of Technology, Beijing, China

Yazhou Hu Beijing Institute of System Engineering, Beijing, China

Qingsong Hua Dynamic Integration and Energy Storage Systems Engineering Technology Research Center, Qingdao University, Qingdao, China

Manli Huang School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China

Wen Huang School of Automation, Nanjing University for Science and Technology, Xuanwu, Nanjing, Jiangsu, People's Republic of China

Jianhua Jia Department of Information and Communication Engineering, Tongji University, Shanghai, China

Limin Jia State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Collaborative Innovation Center of Railway Traffic Safety, Beijing Jiaotong University, Beijing, China; Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Minzheng Jia Beijing Polytechnic College, Beijing, People's Republic of China

Guotao Jiang CSR Zhuzhou Institute Co. Ltd., Zhuzhou, Hunan, China

Jie Jiang School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China

Man Jiang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Tingting Jiang School of Civil Engineering, Tsinghua University, Beijing, China

Linlin Kou State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Haidian District, Beijing, China

Da Li College of Information Science and Engineering, Northeastern University, Shenyang, China

Fachao Li School of Economics and Management, Hebei University of Science and Technology, Shijiazhuang, Hebei, China

G. Li Faculty of Information Science and Technology, Dalian Maritime University, Dalian, China

Guang Li School of Electronic and Control Engineering, Chang'an University, Xi'an, China

Haiying Li State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Hengkui Li CSR Qingdao Sifang Co., Ltd., Chengyang District, Qingdao, China; CRRC Qingdao Sifang Co., Ltd, Qingdao, China

Jian Li Beijing Traffic Control Center, Beijing, China

Man Li State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Qian Li Henan Provincial Communications Planning, Survey and Design Institute Co., Ltd., Zhengzhou, China

Wantong Li School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Xinxing Li School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China

Xuan Li School of Maritime and Transportation, Ningbo University, Ningbo, China

Yan Li State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Haidian District, Beijing, China

Yanyong Li Beijing, China

Yongjie Li School of Automobile, Linyi University, Linyi, Shandong, China

Yulong Li Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology, School of Mechanical Engineering, Guangxi University, Nanning, Guangxi, China

Zhuoyue Li State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Maoxuan Liang School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China

Shuai Lin State Key Laboratory of Rail Traffic Control and Safety, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Changyun Liu School of Civil Engineering, Tsinghua University, Beijing, China

Guangwu Liu Guangzhou Metro Corporation, Guangzhou, China

Hongzhu Liu China Railway Shanhaiguan Bridge Group Co., Ltd., Qinhuangdao, China

Keping Liu College of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, Jilin, China

N. Liu Faculty of Information Science and Technology, Dalian Maritime University, Dalian, China

Rengkui Liu School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Siyu Liu Beijing, China

Xingping Liu Beijing, China

Xinrong Liu School of Automation, Nanjing University of Science and Technology, Nanjing, China

Yanfei Liu College of Information Engineering, Capital Normal University, Beijing, China

Yuhan Liu College of Communication and Information, Changchun University of Science and Technology, Changchun, China

Xiangyang Lu CSR Zhuzhou Institute, Zhuzhou, Hunan, China

Yujing Lu State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Zhejin Lu School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China

Shen Lyu Civil Engineer College, ShenZhen University, Nan Shan District, ShenZhen, China

Jin Ma School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Yanhui Ma College of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, Jilin, China

Lijun Meng School of Electromechanical and Architectural Engineering, Jianghan University, Wuhan, Hubei, China

Lingyun Meng State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Xuelei Meng School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, Gansu, China

Yikui Mo Civil Engineer College, ShenZhen University, Nan Shan District, ShenZhen, China

Haolei Mou Tianjin Key Laboratory of Civil Aircraft Airworthiness and Maintenance, Civil Aviation University of China, TianJin, China

Lu Ni School of Electronic and Control Engineering, Chang'an University, Xi'an, China

Tao Niu China Railway Communication and Signal Survey and Design (Beijing) Co. Ltd., Beijing, China

Teng Pan School of Michigan, Shanghai Jiaotong University, Shanghai, China

Yuxiong Pan CRRC Research of Electrical Technology and Material Engineering, Zhuzhou, Hunan, China

Zhongliang Pan School of Physics and Telecommunications Engineering, South China Normal University, Guangzhou, China

Shaohuang Pang State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Huaijun Peng State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Hai Dian District, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Hai Dian District, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Hai Dian District, Beijing, China

Si Pi State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Bo Qi School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, Gansu, China

Qiongfen Qian Department 4, Air Force Early Warning Academy, Wuhan, China

Yong Qin State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Collaborative Innovation Center of Railway Traffic Safety, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Ruichang Qiu Beijing Jiaotong University (BJTU), Beijing, China

Jicheng Qu School of Economic and Management, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

Yue Qu Beijing Jiaotong University (BJTU), Beijing, China

Jingbao Ren School of Mechanical Engineering, Nangjing University of Science and Technology, Nanjing, China

Sheng Shan CSR Zhuzhou Institute, Zhuzhou, Hunan, China

Hang Shang School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China

Jingjing Shao Beijing Jiaotong University, Beijing, China

Weijuan Shao State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Hao Shi State Key Laboratory of Rail Traffic Control and Safety, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Yuantao Song School of Engineering Science, University of Chinese Academy of Sciences (UCAS), Beijing, China

Aijun Su CSR Zhuzhou Institute, Zhuzhou, Hunan, China

Fang Sun Beijing Traffic Control Center, Beijing, China

Shouxin Sun College of Information Science and Engineering, Northeastern University, Shenyang, China

Xun Sun School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

Guoping Tang CSR Zhuzhou Institute, Zhuzhou, Hunan, China

Jinjin Tang Beijing Jiaotong University, Beijing, China

Jun Tang CSR Zhuzhou Institute Co. Ltd., Zhuzhou, Hunan, China

Junqing Tang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing, China

Yi Tang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing, China

Ying Tang School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

Liang Tao School of Computer Science and Technology, Anhui University, Shushan District, Hefei City, Anhui, China

Feng Tian ShenZhen Urban Transportation Planning Center Co. Ltd, Shenzhen, China

Yudong Tian Mechanical Engineering School, Suzhou University of Science and Technology, Suzhou, Jiangsu, China

Jizhi Wan Henan Provincial Communications Planning, Survey and Design Institute Co., Ltd., Zhengzhou, China

Baohua Wang Beijing Jiaotong University (BJTU), Beijing, China

Baolong Wang School of Engineering Science, University of Chinese Academy of Sciences (UCAS), Beijing, China

Bing Wang Beijing Institute of System Engineering, Beijing, China

Dandan Wang State Key Laboratory of Rail Traffic Control and Safety, School of Traffic and Transportation, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Fuge Wang School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China

Futian Wang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Haidian District, Beijing, China

Guangxun Wang Mechanical Engineering School, Suzhou University of Science and Technology, Suzhou, Jiangsu, China

Huabin Wang School of Computer Science and Technology, Anhui University, Shushan District, Hefei, Anhui, China

Hui Wang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Kai Wang Dynamic Integration and Energy Storage Systems Engineering Technology Research Center, Qingdao University, Qingdao, China

Lu Wang School of Automation, Nanjing University of Science and Technology, Xuanwu, Nanjing, Jiangsu, China

Naiwen Wang School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China

Shujun Wang School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Xiaojun Wang Beijing Institute of Technology, Beijing, China

Xu Wang CRRC Research of Electrical Technology and Material Engineering, Zhuzhou, Hunan, China

Yafei Wang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Yanhui Wang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Zhiguo Wang School of Physics and Electrical and Electronic Engineering, Huaiyin Normal University, Huai'an, Jiangsu, China

Zixuan Wang School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China

Ziyang Wang Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Qiang Wei State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Xiukun Wei State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Shanghai Key Laboratory of Computer Software Evaluating and Testing, Shanghai Development Center of Computer Software Technology, Shanghai, China

Zheng Wen CSR Zhuzhou Institute, Zhuzhou, Hunan, China

Dingfeng Wu Beijing, China

Qihao Wu College of Information Science and Engineering, Northeastern University, Shenyang, China

Qingyun Wu Department of Information and Communication Engineering, Tongji University, Shanghai, China

Zonghua Wu Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology, School of Mechanical Engineering, Guangxi University, Nanning, Guangxi, China

Meng Xi School of Electronic and Control Engineering, Chang'an University, Xi'an, China

Jun Xia School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China

Xianpeng Xia State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing, China

Wu Xie School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China

Zongyi Xing School of Automation, Nanjing University of Science and Technology, Xuanwu, Nanjing, Jiangsu, China; School of Mechanical Engineering, Nangjing University of Science and Technology, Nanjing, China

Guangheng Xu Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology, School of Mechanical Engineering, Guangxi University, Nanning, Guangxi, China

Jie Xu State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Weixiang Xu School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Xinyue Xu State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Xiaoping Xue Department of Information and Communication Engineering, Tongji University, Shanghai, China

Dong Yan State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Shanghai Key Laboratory of Computer Software Evaluating and Testing, Shanghai Development Center of Computer Software Technology), Shanghai, China

Rui Yan Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology, School of Mechanical Engineering, Guangxi University, Nanning, Guangxi, China

Hongyu Yang Beijing Polytechnic College, Beijing, People's Republic of China

Jing Yang School of Mechanical Engineering, Nangjing University of Science and Technology, Nanjing, China

Lin Yang Beijing, China

Jinning Yang School of Economics and Management, Hebei University of Science and Technology, Shijiazhuang, Hebei, China

Weifeng Yang CSR Zhuzhou Institute, Zhuzhou, Hunan, China

Xujuan Yang Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology, School of Mechanical Engineering, Guangxi University, Nanning, Guangxi, China

Yanfang Yang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Hai Dian District, Beijing, China; School of Traffic and Transportation, Beijing Jiaotong University, Hai Dian District, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Hai Dian District, Beijing, China

Yu Yang China Academy of Railway Sciences, Beijing, China

Yuan Yang School of Automation and Information, Xi'an University of Technology, Xi'an, Shaanxi, China

Enjian Yao School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

Xiaowen Yao School of Automation, Nanjing University of Science and Technology, Nanjing, China

Houyuan Ye Management School, Wuhan University of Technology, Wuhan, People's Republic of China

Huaixian Yin School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Dynamic Integration and Energy Storage Systems Engineering Technology Research Center, Qingdao University, Qingdao, China

Kang yongzheng Shijiazhuang Tiedao University, Shijiazhuang, China

Hongfang Yu University of Electronic Science and Technology of China, Chengdu, China

Weibo Yu College of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, Jilin, China

Baojun Yuan China Railway Construction Investment Co., Ltd., Beijing, China

Lei Yuan School of Automation and Information, Xi'an University of Technology, Xi'an, Shaanxi, China

Minzheng Yuan Guangzhou Metro Corporation, Guangzhou, China

Yingying Yuan Hunan CRRC Times Electric Vehicle Co., Ltd., Zhuzhou, Hunan, China

Jun Yun Management School, Wuhan University of Technology, Wuhan, People's Republic of China

Fang Zhang Department of Information and Communication Engineering, Tongji University, Shanghai, China

Hong Zhang Beijing Institute of Technology, Beijing, China

Huimin Zhang School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China

Jinglin Zhang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China; Shanghai Key Laboratory of Computer Software Evaluating and Testing, Shanghai Development Center of Computer Software Technology, Shanghai, China

Kai Zhang College of Information Science and Engineering, Northeastern University, Shenyang, China

Ling Zhang State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Haidian District, Beijing, China

Mei Zhang Library, Linyi University, Linyi, Shandong, China

Qianmin Zhang School of Computer Science and Technology, Anhui University, Shushan District, Hefei City, Anhui, China

Qingyu Zhang School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Rui Zhang School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

Tairan Zhang CSR Zhuzhou Institute Co. Ltd., Zhuzhou, Hunan, China

Xi Zhang School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Xiaoqing Zhang School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China

Yong Zhang School of Automation, Nanjing University of Science and Technology, Xuanwu, Nanjing, Jiangsu, China

Zhen Zhang College of Information Science and Engineering, Northeastern University, Shenyang, China

Zhenyu Zhang State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Haidian District, Beijing, China

Zhilong Zhang CSR Qingdao Sifang Co., Ltd., Chengyang District, Qingdao, China; CRRC Qingdao Sifang Co., Ltd, Qingdao, China

Yong Zhao Beijing West Railway Station, Beijing, China

Zhao zhen Shijiazhuang Tiedao University, Shijiazhuang, China

Ce Zheng Beijing Jiaotong University (BJTU), Beijing, China

Hongyun Zheng School of Electronic and Information Engineering, Beijing Jiaotong University, Hai Dian, Beijing, China

Jinchu Zheng School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Liming Zheng College of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, Jilin, China

Mingming Zheng State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Yuqian Zheng College of Information Science and Engineering, Yanshan University, Qinhuangdao, China

Zhipeng Zhong School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin, Guangxi, People's Republic of China

Jian Zhou School of Computer Science and Technology, Anhui University, Shushan District, Hefei, Anhui, China

Leishan Zhou Beijing Jiaotong University, Beijing, China

Fang Zhu Department of Aircraft Maintenance and Engineering, Guangzhou Civil Aviation College, Guangzhou, China

Lei Zhu State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Haidian District, Beijing, China

Wei Zhu College of Communication and Information, Changchun University of Science and Technology, Changchun, China

Yuanzhong Zhu Beijing Polytechnic College, Beijing, People's Republic of China

Tianchun Zou Tianjin Key Laboratory of Civil Aircraft Airworthiness and Maintenance, Civil Aviation University of China, TianJin, China

Chapter 1
The Optimization Methods for Wireless Sensor Network Nodes Deployment Based on Hybrid Particle Swarm

Yan Li, Honghui Dong, Limin Jia and Junqing Tang

Abstract The energy of sensor nodes is usually limited, and it is difficult to supplement energy frequently. Multi-hop communication, which can lead to unbalanced energy consumption in the whole network and seriously affect the life cycle of wireless sensor network, is generally used for infrastructure and testing linear network. Optimization of regionalization node deployment was used to balance energy consumption among all the nodes and establish the network optimization model. Combined with engineering application, hybrid particle swarm was used as model for the linear network transmission strategy. Finally, simulation was used for modeling verification. The results of simulation showed that the energy consumption of optimizing transmission strategy is lower with balanced energy consumption among all the nodes, which aims to reduce energy consumption so that the life cycle of the whole network can be improved.

Keywords Hybrid particle swarm · Network nodes · Regionalization · Transmission strategy

1.1 Introduction

Due to the dramatic increase in worldwide wireless communication and digital electronic technology, wireless sensor network, which is considered one of the most important technologies in the twenty-first century, is widely used in military applications, industrial control, environmental monitoring, and in other fields [1].

Wireless sensor node consists of perception unit, processing unit, communication unit, and energy supply unit in the wireless sensor network. Perception unit is

Y. Li · H. Dong (✉) · L. Jia · J. Tang
Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Haidian District, Beijing, China
e-mail: hhdong@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_1

responsible for information perception and conversion of data format in the region. Processing unit is used for storage, processing, and forwarding of various data. Communication unit is applied for data transceiver and wireless communication with other nodes. Energy supply unit provides the energy needed for other parts in the network to work.

Generally, wireless sensor node is provided energy by batteries to be supplied to each part to work regularly. However, due to the limited energy of a battery, how to reduce the energy consumption of wireless sensor network is considered to be of great significance. The energy consumption of wireless sensor network is used principally for perception unit, processing unit, and communication unit. As shown in Fig. 1.1, data transmission has the largest contribution to energy consumption, which is followed by data reception and free time.

As the energy consumption of the communication unit is much greater than the energy consumed by the perception unit and processing unit in wireless sensor nodes, the energy consumed for them can be neglected, while the energy consumption of communication unit is supposed to be focused on particularly.

The achievements of reducing the energy consumption of wireless sensor network normally concentrate on the acceleration of data collection and node deployment balance. Xu [2] analyzed accelerating data collection strategy and energy consumption balance among the nodes in the linear wireless sensor network. Jawhar et al. [3] studied the mobile routing algorithm and network architecture for data collection based on unmanned aerial vehicle in the linear wireless sensor network. Kundu et al. [4] researched on network energy-saving strategy based on wireless network topology protocol. Wu and Sun [5] analyzed the influence of node deployment density of linear wireless sensor network on improving the accuracy of data perception. For linear topology wireless sensor networks, Yu and Pan [6] presented a nonuniform linear topology WSN clustering routing algorithm to extend the life cycle of the network and avoid the emergence of energy voids.

In this paper, when it comes to the deployment of nodes, the life cycle of the network is to be taken into consideration to balance the energy consumption and the

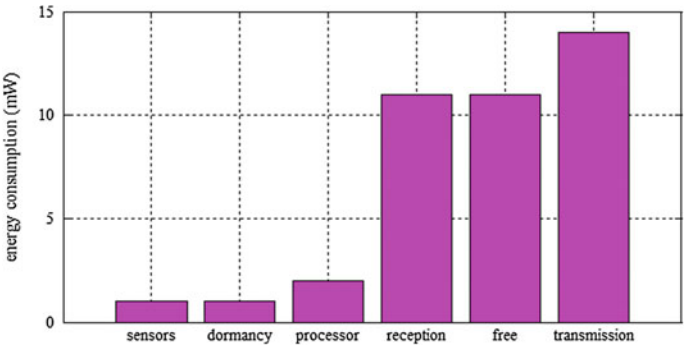


Fig. 1.1 Energy consumption distribution curve of wireless sensor node

communication load. First, nonuniform deployment scheme was used to establish the network optimization model. Then combined with engineering application, hybrid particle swarm was used to model the linear network transmission strategy, which effectively improved the life cycle of the network. Finally, simulation was used for modeling verification. The results of simulation showed that the energy consumption of optimizing transmission strategy is lower with balanced energy consumption among all the nodes, which aims to reduce energy consumption and improve the life cycle of the whole network.

1.2 Optimization of Regionalization Node Deployment

Generally, multi-hop communication is adopted for the linear detection network for infrastructure in the scheme for uniform deployment of nodes. So the nodes whose distance to base stations or convergent nodes is much closer are inevitable to forward more data and consume more energy, which will lead to "energy voids". Meanwhile, the nodes whose distance to base stations or convergent nodes is much further will remain with excess energy. Thus, the whole life of the network has a premature end. In this paper, nonuniform nodes deployment scheme was presented to balance the energy consumption among each node and improve the life cycle of the whole network.

1.2.1 Energy Consumption Model of Wireless Sensor Network

According to the relevant study of wireless communication [7], the energy consumption calculation formula of a network node transmitting bit to or receiving bit from another node has distance as follows:

$$E_{tr}(k,d) = \begin{cases} kE_{\text{elec}} + kE_{\text{amp1}}d^2 (d < d_0) \\ kE_{\text{elec}} + kE_{\text{amp2}}d^4 (d \geq d_0) \end{cases} \tag{1.1}$$

$$E_{rx}(k,d) = kE_{\text{elec}} \tag{1.2}$$

where k means the size of transmitting or receiving data, d means the distance between the two nodes, d_0 is constant whose value depends on the network environment as follows: $d_0 = \sqrt{E_{\text{apm1}}/E_{\text{amp2}}}$. E_{elec} means the energy consumption of transmitting circuit, while E_{amp} means the energy consumption of the transmitting amplifier. When the distance between transmitting node and receiving node is less than d_0, the energy consumption of transmitting data is proportional to the square of the distance, which is called free space model. On the other hand, when the distance

is more than d_0, the energy consumption of transmitting data is four square of the distance, which is called multiple attenuation model. Assume $d < d_0$. Then given the fact that direct communication is adopted, the energy consumption of transmitting k bit data is $E(k, d, 1) = 2kE_{\text{elec}} + kE_{\text{amp1}}d^2$ between the two nodes. According to the relevant study of wireless communication [7], experiments performed on randomly generated test instances are as follows:

$$d_0 = 93 \text{ m}; E_{\text{elec}} = 40 \text{ nJ/bit}; E_{\text{amp1}} = 12 \text{ pJ/bit/m}^2; E_{\text{amp2}} = 0.0014 \text{ pJ/bit/m}^4$$

If the energy consumption of wireless communication is the only one to focus on, then the relationship of energy consumption and distance between two sensor nodes transmitting a byte (8 bit) data is shown as in Fig. 1.2. From the figure, with the increase in the transmission distance, especially when the transmission distance is more than 400 m, the energy consumption of wireless communication increases quickly.

When multi-hop communication is adopted, a relaying node can be laid between two nodes. According to the inequality principle, the energy consumption is considered to be minimal when the relaying node is laid in the middle of the two nodes; and the minimum energy consumption to transmitting k bit is $E(k, d/2, 2) = 2(2kE_{\text{elec}} + kE_{\text{amp1}}(d/2)^2)$. Therefore, when $d > 2\sqrt{E_{\text{elec}}/E_{\text{amp1}}}$, multi-hop communication is more efficient to save energy without raising costs to add another node, that is, $E(k, d/2, 2) < E(k, d, 1)$. $d = 2\sqrt{E_{\text{elec}}/E_{\text{amp1}}}$ is deemed to be the minimum distance in linear network multi-hop communication.

1.2.2 Network Optimization Model

Assume the distance of linear detection area is $2L$, and the base station is in the middle of this area in the linear detection network for infrastructure. Due to the symmetry of linear network, analysis for any side of the base station is enough.

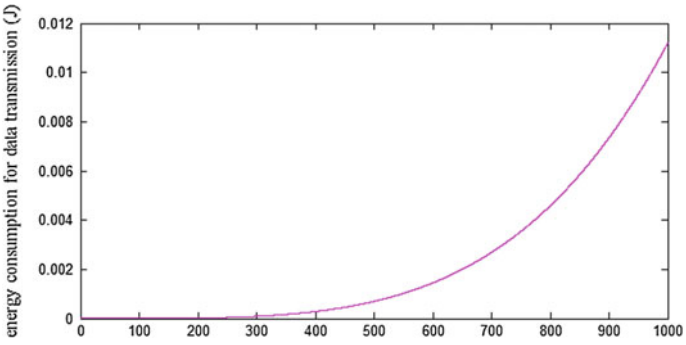


Fig. 1.2 Relationship of energy consumption and distance between two sensor nodes

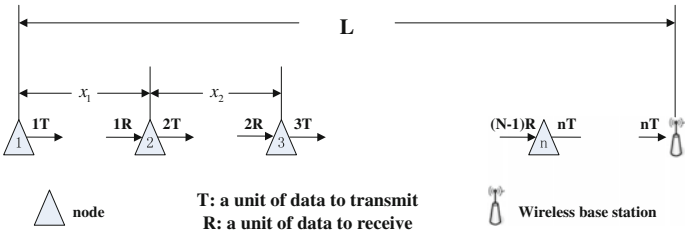


Fig. 1.3 Data node transmission flowchart of a side of base station

The topological structure of linear network is shown as in Fig. 1.3. Assume the node number is $i(1 \leq i \leq n)$, and the node number is 1 to n in turn based on the distance to the base station.

Make the following assumptions in the above network: (1) The energy of base station is enough. (2) Deployment scheme of each node is certain. Every data generation is equal and the amount of data is k. (3) Each node can only deliver data to the next node, and all the data will transmit to the base station by multi-hop communication. (4) The initial energy of each node is equal and wireless communication has the largest energy consumption. (5) The distance between arbitrary adjacent nodes is less than d_0, so the free space model is adopted.

The energy consumption of Node i to receive data from Node $i-1$ is $E_{rxi} = (i-1)kE_{\text{elec}}$, and to transmit data which include data both from Node $i-1$ and from Node i itself is $E_{txi} = ikE_{\text{elec}} + ikE_{\text{amp1}}x_i^2$. So, energy consumption of each time of Node i is $E_i = E_{rxi} + E_{txi} = (2i-1)kE_{elec} + ikE_{amp1}x_i^2$. Thus, balancing the energy consumption among all the nodes in the network is an effective way to prolong the life cycle of the network, that is, $E_1 = \cdots = E_{i-1} = E_i = \cdots = E_n$. Calculation formulas are as follows:

$$\begin{cases} \sum_{i=1}^{n} x_i = L \\ E_{\text{amp1}}[(i-1)x_{i-1}^2 - ix_i^2] = 2E_{\text{elec}}, \ 1 < i \leq n \end{cases} \tag{1.3}$$

It is obvious that formula (1.3) cannot be solved.

1.3 Optimization Methods for Deployment of Linear Network Transmission Strategy Based on Hybrid Particle Swarm

Assume there are nodes in the network, and every node can only transmit data to the next four nodes because of the limitation of distance. That is, Node 1 can only transmit data to Node 2 to Node 5, and Node 2 can only transmit data to Node 3 to Node 6, and so on.

1.3.1 Assumption of Hybrid Particle Swarm

When n is big enough to solve, the series of linear network transmission distance series is considered to be an NP problem. Then, the design process of hybrid particle swarm can be conducted as follows:

1. Encoding: particle swarm optimization (PSO) can use real number to encode, while genetic algorithm can only use the binary to encode. Assume the number of each node is 1, 2…n. According to the previous assumption, the encode range of Node 1 is [2 $\sim$ 5], the encode range of Node 2 is [3 $\sim$ 6], and the encode range of Node i is [i + 1 $\sim$ i + 4] ($1 \le i \le -4$).
2. Fitness function: the objective function is that the energy consumption of the whole network is minimum, that is, $\min \sum_{i=1}^{n} E_i$. Where $E_i = E_{tri}(s_{tri}, x_i) + E_{rxi}(s_{rxi})$. And the constraint condition is $2\sqrt{E_{\text{elec}}/E_{\text{amp1}}} < x_i < d_0$.
3. Updating: each iteration is updated with the value of the fitness function, with gradual approximation to the optimal solution.
4. Crossover: crossover and mutation were added into hybrid particle swarm compared to the traditional PSO. Select different optimal individuals for crossover to produce better solutions.
5. Mutation: random gene position will make mutation to occur with a little probability (e.g., 5 %) to generate a better solution.

1.3.2 Design of Algorithm Flow

Step 1: Population initialization. Algorithm parameter initialization includes the size of particle swarm population NIND, iteration number Maxgen, nodes set S, the set of distance between nodes X, the number of crossover cross_num, mutation probability, the set of node data quantity Q, encoding rules, the parameter of the constraint condition, and so on. Then encode the transmission distance sequence of linear network and the initial particle swarm is generated.

Step 2: Calculation for the value of fitness function. The best value and the encoding of fitness function are (tourGbest, recordGbest) = min (tourPbest, recordPbest), respectively. Then, turn to Step 7.

Step 3: Updating of particle swarm. Update the particle swarm to generate new particles. If the new particles accord with the constraint condition, then turn to the next step. Otherwise, repeat Step 3.

Step 4: Crossover with the best individual. Generate a new crossover in the range of cross_num. If the new crossover accord with the constraint condition, then turn to Step 2. Otherwise, repeat Step 4.

Step 5: Crossover with the best population. Similar to Step 4.

Step 6: Mutation of particle. Similar to Step 4.
Step 7: To determine whether to end the iteration. If the number of crossovers is reached, then end the iteration and get the result. Otherwise, turn to Step 3.

1.4 Case Study

According to the above analysis, optimization of the transmission distance sequence of the linear network is based on hybrid particle swarm. New particle can be generated by updating particle swarm, crossover with individual, crossover with population, mutation of particle, and so on. Select better particles by calculation of the value of fitness function. Search for the optimal solution by iteration repeatedly, and finally the best transmission distance sequence can be obtained.

Assume that there are 20 nodes in the linear network and uniform deployment scheme is used among all the nodes. The base station is placed on one side of the linear network. The amount of data of each node in unit time is equal. Every node can only transmit data to the next four nodes.

Since the number of unknown quantity is large in this model, enough particles are needed to ensure the diversity of population and the operation of the algorithm. Taking the solving precision and the computation time into consideration with multiple tests, the relevant parameters are set as follows: the number of particles: NIND = 20, iteration number: MAXGEN = 200, crossover probability: pc = 0.1, and mutation probability: pm = 0.05.

Comparison of random transmission distance, single-hop communication transmission, and the transmission distance sequence based on hybrid particle swarm, energy consumption of each node in the network is shown in Fig. 1.4.

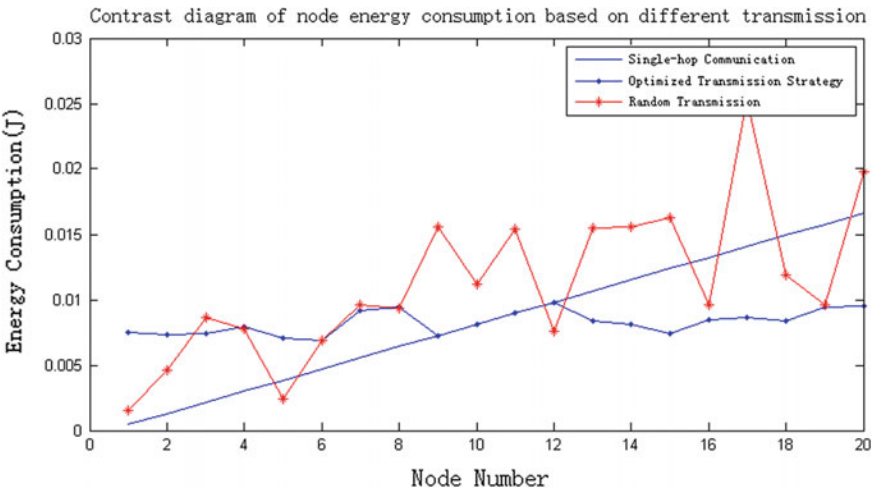


Fig. 1.4 Contrast diagram of node energy consumption based on different transmission strategies

Figure 1.4 shows that the optimized transmission strategy is much better than random transmission and single-hop communication transmission with less energy consumption. The energy consumption of each node is balanced. The goal to reduce energy consumption and improve the life cycle of the whole network was achieved. So, the result met the target expectations with better convergence and convergence results.

1.5 Conclusion

In this paper, research on transmission energy consumption of the wireless sensor network was carried out comprehensively. Optimization of regionalization node deployment was dealt with. Then combined with engineering application, hybrid particle swarm was used to the model for the linear network transmission strategy, which effectively improved the life cycle of the network. Finally, simulation was used for modeling verification. So the result is: the optimized transmission strategy is much better than random transmission and single-hop communication transmission with less energy consumption. The energy consumption of each node was balanced. The goal to reduce energy consumption and improve the life cycle of the whole network was achieved.

Acknowledgment This work is supported in part by the national high-tech development program of China (2011AA11A102) and Beijing nova program (Grant Z1211106002512027).

References

1. Xiao Y, Rayi VK, Sun B (2007) A survey of key management schemes in wireless sensor networks. Comput Commun 30(11):2314–2341
2. Xu J (2010) Research on the data collection policy for linear wireless sensor networks based on energy balance. Central South Univ. doi:10.7666/d.y1721591 (in Chinese)
3. Jawhar I, Mohamed N, Al-Jaroodi J (2014) A framework for using unmanned aerial vehicles for data collection in linear wireless sensor networks. J Intell Rob Syst 74(1–2):437–453
4. Kundu S, Roy S, Pal A (2008) A Power-aware wireless sensor network based bridge monitoring system. Networks, 2008. ICON 2008, pp 1–7
5. Wu J, Sun N (2010). Optimal sensor density in a distortion-tolerant linear wireless sensor network. Global telecommunications conference, pp 1–5
6. Yu R, Pan B (2012) Unequal clustering routing algorithm in linear WSN. J Hangzhou Dianzi Univ 32(6):97–100 (in Chinese)
7. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocol for wireless microsensor networks. Syst Sci 30(11):2314–2341
8. Estrin D (2002) Wireless sensor networks tutorial part IV: sensor network protocols. Mobicom, Westin Peachtree Plaza, Atlanta, Georgia 2002, pp 23–28
9. Chao S, Ming L, Gong Haigang (2009) ACO-based algorithm for solving energy hole problems in wireless sensor networks. J Software 20(10):2729–2743 (in Chinese)

Chapter 2
Research on Gearbox Fault Diagnosis of Urban Rail Vehicle Based on SIMPACK

Xiukun Wei, Dong Yan and Jun Chen

Abstract Gearbox, an important component of urban rail vehicles working in the most severe conditions, plays a crucial role in the urban rail vehicle fault diagnosis. It is necessary to detect the fault of gearbox at an early stage to prevent human casualties and reduce maintenance costs. Most of the data-driven fault diagnosis methods are based on an experimental platform, which has many disadvantages such as costly maintenance, limited application, and limited degree of fault simulation. In this paper, a new method based on SIMPACK is proposed and confirmed efficient by testing, which can overcome the disadvantages of the experimental platform. Meanwhile, MATLAB was applied to time–frequency domain analysis and wavelet packet analysis. These analysis results demonstrated the feasibility and reliability of this method.

Keywords Gearbox · SIMPACK · Fault diagnosis · Time–frequency domain analysis · Wavelet packet analysis

2.1 Introduction

Gearbox is a dispensable component of urban rail vehicle, and it is the place where fault occurs easily. According to statistics, 60 % of the gearbox fault occurred on gear, 19 % occurred on shaft [1], and all of the faults interact with each other.

X. Wei (✉) · D. Yan · J. Chen
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
e-mail: xkwei@bjtu.edu.cn

X. Wei · D. Yan
Shanghai Key Laboratory of Computer Software Evaluating and Testing (Shanghai Development Center of Computer Software Technology), Shanghai, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_2

The faults must be detected at an early stage to reduce maintenance costs, improve productivity, increase machine availability, and prevent human casualties [2].

With the rapid progress of modern testing and signal processing technology, researchers pay more and more attention to vibration analysis. Successful examples of fault diagnosis can be found in [3]. The literature [4] has detailed works on the implementation process of wavelet analysis and application in fault diagnosis. Time–frequency approach recognition of faults in gear tooth is presented in the literature [5]. A PCA improved algorithm of BP neural network is introduced in paper [6]. The data-driven methods that describe the above are based on an experimental platform, which has many disadvantages. In this paper, a new method based on SIMPACK is proposed for overcoming the disadvantages and achieving a fundamental change in the condition-based monitoring and fault diagnosis of gearbox.

The rest of this paper is organized as follows. In Sect. 2.2, gearbox modeling is introduced. Section 2.3 describes three kinds of gearbox faults which is analyzed by time–frequency domain. In Sect. 2.4, taking shaft-misalignment is analyzed by wavelet. The conclusions are drawn in Sect. 2.5.

2.2 Modeling of Gearbox

Parameters from ZMA080 of metro vehicle were chosen for the model. The density of gearbox case is 7.3 g/cm^2, involute helical gears is adopted by the driver gear and passive gear [7]. Simulation time is 20 s and sampling frequency is 24 kHz. Two sensors were installed at holding point of gearbox and the data in the first 5 s were removed during analysis. Parameters of gear and motor are shown in Tables 2.1 and 2.2, respectively. The basic parameters of gearbox are presented in Table 2.3. Detailed parameters setting is shown in Fig. 2.1, which is a topology graphics of gearbox. The 3-D dynamic model based on SIMPACK is shown in Fig. 2.2.

Table 2.1 Parameters of gear

Name	Normal modulus	Gear number (driver)	Gear number (passive)	Tooth profile angle	Helix angle
Value	5	19	120	20°	12.2°

Table 2.2 Basic technical parameters of the motor

Name	Rated power	Rated speed	Max speed	Starting torque	Rated torque	Max traction torque
Value	190 kW	1800 r min^{-1}	3481 r min^{-1}	1625 N m	1008 N m	1626 N m

Table 2.3 Basic parameters of gearbox

	Case body		Input shaft	Output shaft		Driver gear		Passive gear	
Size/mm	656 * 1005 * 286		L: 400	L: 400		Addendum D: 107		Addendum D: 624	
			R: 60	R: 60		Root circle D: 85		Root circle D: 601	
Mass/kg	127.6		31.5	7.3		1		1	
Rotational inertia/kg m^2	Ix	5.45	Ix	0.054	0.0029	Ix	1	Ix	1
	Iy	11.61	Iy	0.448	0.0993	Iy	1	Iy	1
	Iz	15,032	Iz	0.448	0.0993	Iz	1	Iz	1

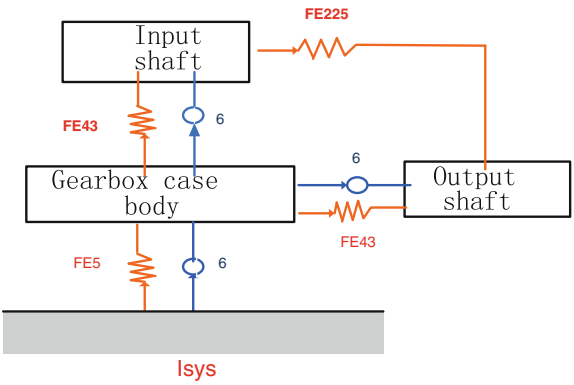


Fig. 2.1 Topology of gearbox model

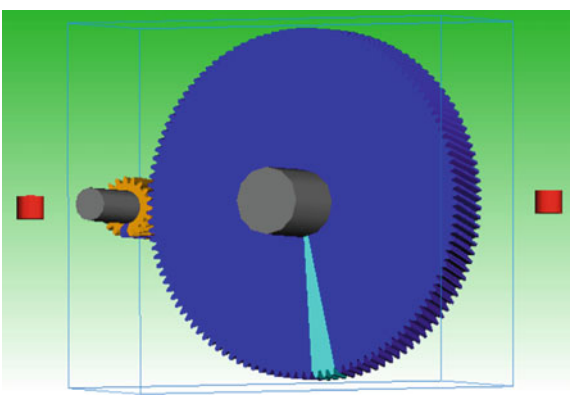


Fig. 2.2 3-D dynamic model of gearbox

2.3 Fault Diagnosis of Gearbox

2.3.1 Symmetrical Tooth Wear Analysis

The comparison of time domain waveform of acceleration signal between normal and symmetrical tooth wear can be found in Fig. 2.3, where the amplitude of acceleration is aggravated when symmetrical tooth wear occurs. The characteristic, which is presented in Fig. 2.4, shows that peaking and kurtosis change greatly. Amplitude spectrum of axial force has shown that meshing frequency and double meshing frequency increased when symmetrical tooth wear occurred in Fig. 2.5.

According to the results, the conclusions of symmetrical tooth wear are [8]:

- Energy of vibration increases when tooth wear occurs (including effective value and other indexes).
- Amplitude of meshing frequency and doubling frequency aggravates.

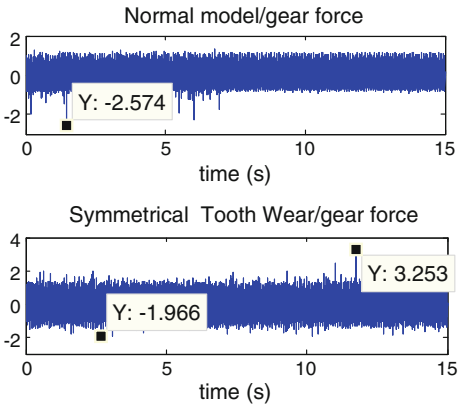


Fig. 2.3 Time domain signal of acceleration

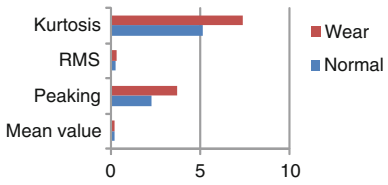


Fig. 2.4 Time domain characteristic

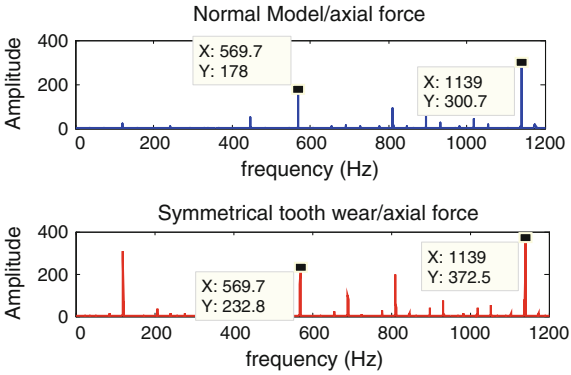


Fig. 2.5 Amplitude spectrum comparison chart of axial force

2.3.2 *Shaft Misalignment Analysis*

In Fig. 2.6, time domain waveform comparison of axial force shows that the range of axial force changes greatly and impulsive periodic signal occurs, which is the most significant difference. The gears cannot mesh with others when the shaft misalignment occurs as shown in Fig. 2.7; it also explains the phenomenon of impacted periodicity signal. The change in peak value is much higher than the other characteristics in Fig. 2.8.

Amplitude spectrum is shown in Fig. 2.9, the axial force of normal model concentrates on meshing frequency and double meshing frequency entirely, and axial force of shaft-misalignment model concentrates on rotation frequency and also its double frequency. The amplitude of failure model is higher than normal in Fig. 2.10. In addition, the carry wave frequency of the sideband, which is generated by failure model, is meshing frequency and its harmonics, and the model created fault shaft rotation frequency and double rotation frequency for interval of sideband clearly.

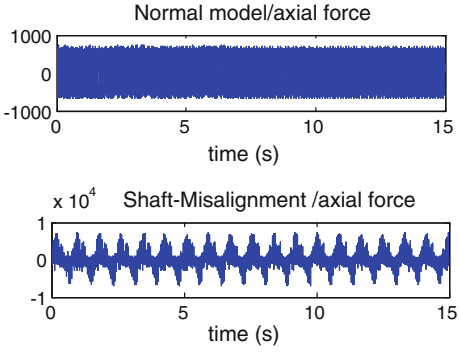


Fig. 2.6 Time domain waveform of axial force

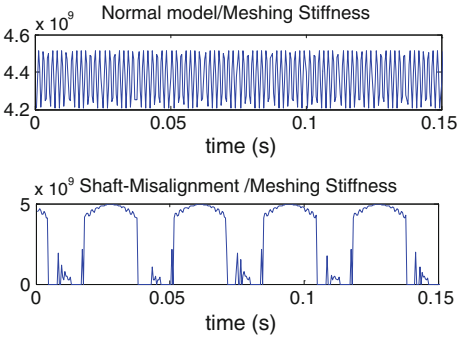


Fig. 2.7 Chart of meshing stiffness

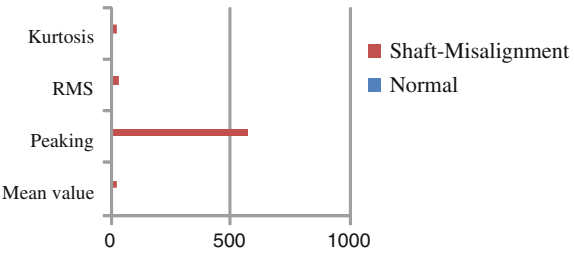


Fig. 2.8 Characteristic value of shaft misalignment

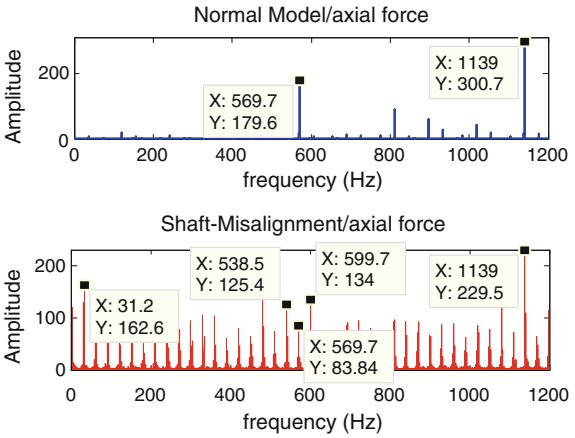


Fig. 2.9 Amplitude spectrum of axial force

According to the results, the conclusions of shaft misalignment are [8]:

- Generating sideband, which is centered by meshing frequency and double meshing frequency and its intervals are rotation frequency and double rotation frequency.
- Amplitude of frequency and double meshing frequency is higher than the normal.
- Energy of vibration increases when shaft misalignment occurs (including effective value and other characteristic values).

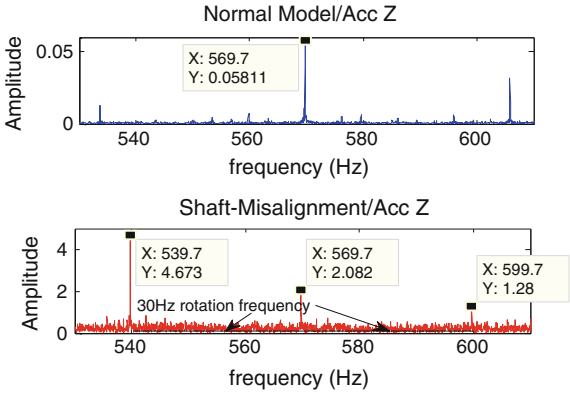


Fig. 2.10 Amplitude spectrum comparison chart of acceleration signal (z direction)

2.3.3 *Mild Bending of Shaft Analysis*

What meshing stiffness comparison Fig. 2.11 shows is a periodic change when mild bending of shaft occurs. Characteristic values are shown in Fig. 2.12, the peak value and kurtosis increase greatly in all the characteristic values.

For further analysis, increasing amplitude that corresponds to rotation frequency of fault shaft is shown in Fig. 2.13, and the carry wave frequency of the sidebands, which is generated by failure model, is meshing frequency and its harmonics. Rotation frequency of input shaft is 30 Hz, transmission ratio is 6.32, rotation frequency of output shaft is 4.7 Hz, and double frequency is 9 Hz when input shaft is failure. Amplitude spectrum comparison chart of axial force is presented in Fig. 2.14, where meshing frequency and double meshing frequency are clear.

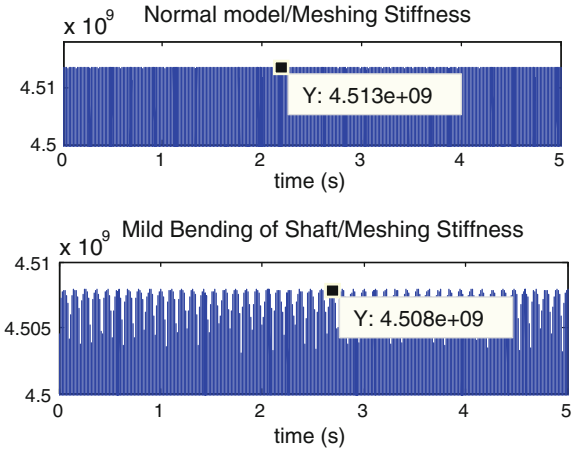


Fig. 2.11 Comparison of meshing stiffness

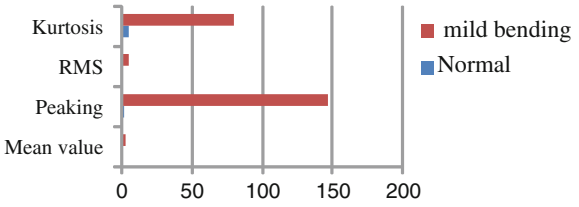


Fig. 2.12 Characteristic value of shaft misalignment

Fig. 2.13 Acceleration signal amplitude spectrum

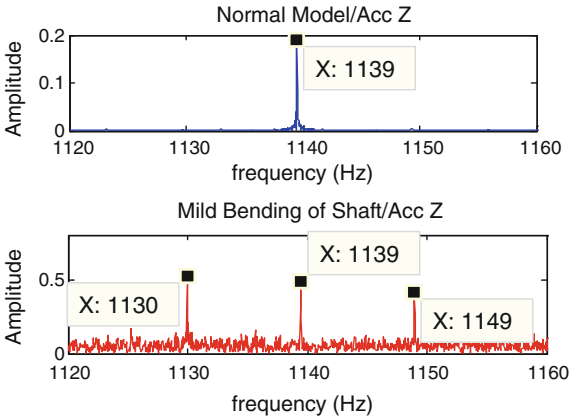
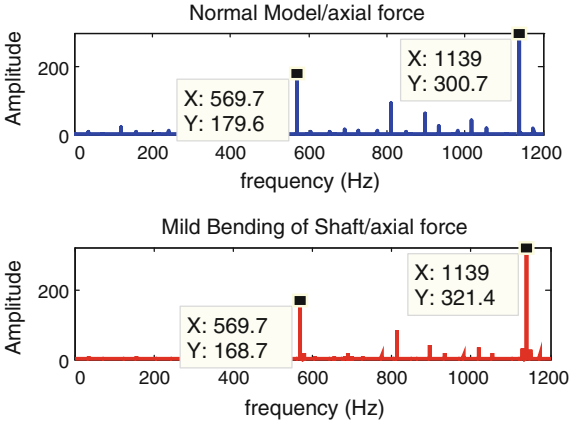


Fig. 2.14 Amplitude spectrum of axial force



According to the results, the conclusions of mild bending of shaft are [8]:

- Amplitude of meshing frequency and double frequency increases significantly.
- Gearbox system generates sidebands around the meshing frequency, and the meshing frequency and its harmonics are frequencies of carry wave when mild bending occurs.

- Energy of vibration increases when mild bending of shaft occurs (including effective value and other characteristic values).

2.4 Wavelet Analysis

A partially magnified point S30 is shown in Fig. 2.15 (1), and the wavelet packet decomposition of vibrating signal is shown in Fig. 2.16 (2). Vibration intensity of each fault model layer is higher than normal layer, especially point S30. The phenomenon is caused by low frequency of 30 Hz of fault shaft rotation frequency.

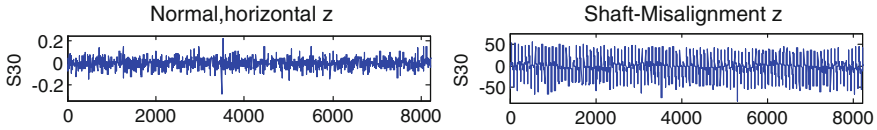


Fig. 2.15 Wavelet packet decomposition of vibrating sign (1)

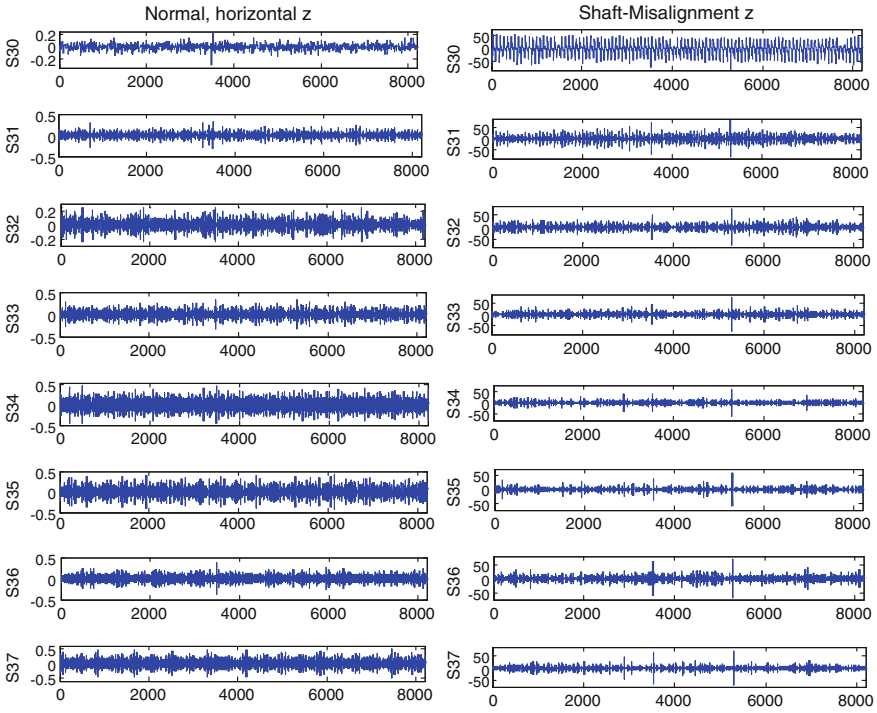


Fig. 2.16 Wavelet packet decomposition of vibrating sign (2)

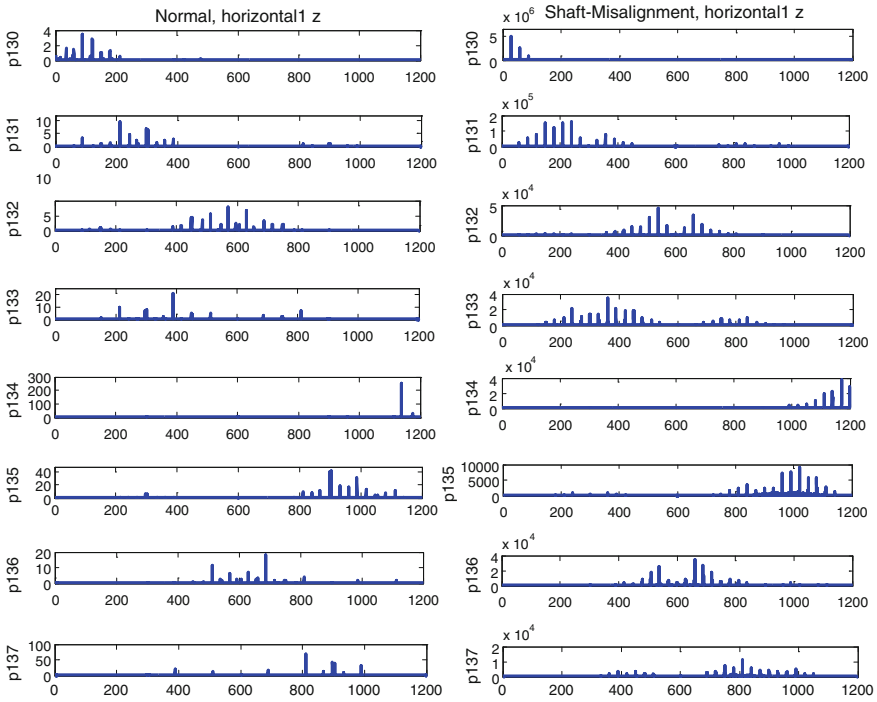


Fig. 2.17 Power spectrum of each layer signal of vibrating signal

From power calculation and analysis of signal of each layer, it is shown that the power of shaft-misalignment signal is far higher than the power of normal signal. The phenomenon is caused by increased vibration of the gear system when failure happens, and the results are shown in Fig. 2.17.

The energy distribution histograms of signal of each layer are shown in Figs. 2.18 and 2.19. It can be seen that the energy of normal model mainly concentrates on meshing frequency and double meshing frequency; in other words, higher frequency energy accounts for a larger proportion in the total energy. Energy of rotation frequency of shaft misalignment increases greatly, its rotation frequency is 30 Hz, which belongs to low frequency band 0–300 Hz, and the low frequency band energy accounts for nearly 60 % of the total energy.

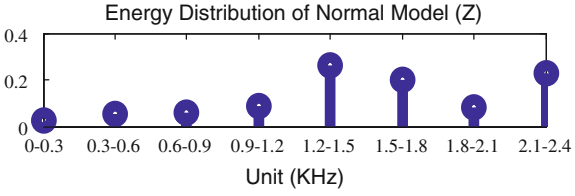


Fig. 2.18 Energy distribution of normal

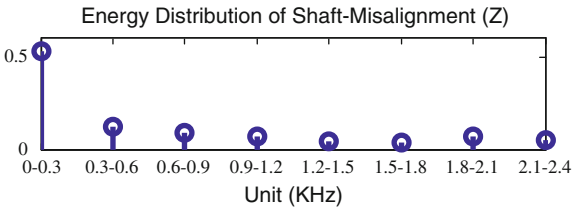


Fig. 2.19 Energy distribution of axle misalignment

2.5 Conclusions

In this paper, simulation based on SIMPACK is introduced, and the applicability of the method is verified by calculations and analyses. The conclusions in this paper are similar to gearbox analysis book [8]. Further extension of this work is under way to concentrate on more kinds of gearbox faults, such as pitting, broken, and invalidation. Additional suggested lines of work can be done on gearbox fault feature extraction techniques and methods. In this way, the method of fault diagnosis provided by SIMPACK will be more accurate and flexible.

Acknowledgments This work is partly supported by Shanghai Key Laboratory of Computer Software Evaluating and Testing (Grant number: SSTL2015_01), Chinese National Key Technologies R&D program (Contract No. 2015BAG13B01) and State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2014ZZ005).

References

1. Lan F (2006) Study on diagnosis of typical gear faults. Sichuan Metall 28:47–48 (in Chinese)
2. Li B, Zhang P, Tian H, Mi S, Liu D, Ren G (2011) A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox, Expert Syst App 38:10000–10009
3. Dempsey PJ (2008) Integrating oil debris and vibration gear damage detection technologies using fuzzy logic. In: NASA, US
4. Inalpolat M, Kahraman A (2009) A theoretical and experimental investigation of modulation sidebands of planetary gear sets. J Sound Vibr 323:677–696
5. Meltzera G, Dienb NP (2004) Fault diagnosis in gears operating under non-stationary rotational speed using polar wavelet amplitude maps. Mech Syst Signal Process 18:985–992
6. Kong Q, Huang F (2011) Research of gearbox fault diagnosis based on the PCA—improved algorithm of BP neural network. Instrum Technol 9:16–19 (in Chinese)
7. Pu Q, Zhou J (2011) Bogie gear box development of metro vehicle ZMA080 of Type A. Electr Locomotives Mass Transit Veh 4(34) (in Chinese)
8. Ding K, Li W, Zhu X (2006) Practical technology of fault diagnosis of gear and gear box. China Machine Press, Beijing (in Chinese)

Chapter 3
A Novel NP Balance Strategy for NPC Inverter Based on SPWM

Bo Gong and Shanmei Cheng

Abstract The neutral point (NP) voltage fluctuation is an innate problem for NPC three-level (3L) inverter. The NP voltage variation is analyzed, the relationship between NP voltage variation and NP current is studied, and a new control method is proposed in this paper. A closed-loop system is built by superimposing the voltage offset on the regulation phase in the proposed method. It is easy to balance the NP voltage by regulating the value of the voltage offset and the adjusting angle; the control structure is simple. Simulation results verify that the proposed control method has excellent ability to maintain the NP voltage balance.

Keywords NP voltage variation · NPC · Voltage offset · SPWM

3.1 Introduction

NPC topology for 3L inverter is most commonly used in medium-voltage applications [1–4]. However, the neutral point voltage fluctuation is an innate problem of NPC 3L inverter. The different capacitors parameters, unbalance loads, and different modulation methods will cause the NP voltage to change, make the NP voltage shift which may destroy the capacitors and power the IGBT, and cause deterioration of the operational condition in the system. Therefore, the control method of NP voltage fluctuation is widely studied and many control strategies have been proposed to solve the problem [5–8].

This paper presents detailed analysis of the relationship between NP voltage variation and NP current, a new control strategy which is used is proposed to solve

B. Gong (✉)
Wuhan Institute of Marine Electric Propulsion, CSIC, Wuhan, China
e-mail: 151701728@163.com

S. Cheng
School of Automatic, Huazhong University of Science and Technology, Wuhan, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_3

the NP voltage unbalance. A voltage offset, in the proposed method, is superimposed on the modulation waves of SPWM. The relationship between the shift of NP voltage and the voltage offset with different conditions is studied; the NP voltage regulation speed is determined by the amplitude of the offset, and the regulation direction is determined by the sign of the offset. It is easy to make the NP voltage balance by regulating the sign and the amplitude of the voltage offset. The simulation results based on MATLAB indicate that the proposed method has excellent ability to maintain the NP voltage balance.

3.2 NP Voltage Analysis

In a PWM cycle, the NP current i_o is

$$i_o = i_a d_{ao} + i_b d_{bo} + i_c d_{co} \tag{3.1}$$

where d_{ao}, d_{bo}, d_{co} are the "O" state time of each phases in a PWM cycle. Set $j = a$, b, c, and the "O" state time can be described by the following formula:

$$d_{jo} = \begin{cases} 1 - u_j & (u_j \geq 0) \\ 1 + u_j & (u_j < 0) \end{cases} \tag{3.2}$$

Therefore, the shift in the NP voltage is

$$\Delta U_c = U_{c1} - U_{c2} = \frac{i_o}{C} = \frac{i_a d_{ao} + i_b d_{bo} + i_c d_{co}}{C} \tag{3.3}$$

By the above expression, it is easy to know that the shift in the NP voltage is caused by the NP current.

3.3 NP Voltage Control Method

One phase among the three phases is chosen to be the regulation phase u_m and the regulation phase has the maximum absolute value of all the phases. u_m is determined by the following expression:

$$u_m = \begin{cases} \min\{u_j\} & \max\{u_j\} < abs(\min\{u_j\}) \\ \max\{u_j\} & \max\{u_j\} > abs(\min\{u_j\}) \end{cases} \quad j = a, b, c \tag{3.4}$$

The voltage offset Δu is superimposed on u_m to make the NP voltage balance

$$u'_m = u_m + \Delta u \tag{3.5}$$

Set $\Delta\theta$ as the adjusting angle, and the changed modulation waveform is shown as follows:

$$u'_a = \begin{cases} u_a + \Delta u & (3\times\pi/2 - \Delta\theta) < \omega t < (3\times\pi/2 + \Delta\theta) \quad \text{or} \quad (\pi/2 - \Delta\theta) < \omega t < (\pi/2 + \Delta\theta) \\ u_a & (\pi/2 + \Delta\theta) \le \omega t \le (3\times\pi/2 - \Delta\theta) \quad \text{or} \quad (-\pi/2 + \Delta\theta) \le \omega t \le (\pi/2 - \Delta\theta) \end{cases} \tag{3.6}$$

$$0 \le \Delta\theta \le \pi/2 \tag{3.7}$$

ΔU_{ca} is the shift in the NP voltage in phase A, ΔU_{ca} is expressed as follows:

$$\begin{aligned} \Delta U_{ca} &= \int_0^{2\pi} \frac{i_a d'_{ao}}{C} \mathrm{d}(\omega t) \\ &= \int_0^{\pi} \frac{i_a \times (1 - u_j)}{C} \mathrm{d}(\omega t) - \int_{\frac{\pi}{2}-\Delta\theta}^{\frac{\pi}{2}+\Delta\theta} \frac{i_a \times \Delta u}{C} \mathrm{d}(\omega t) \\ &\quad + \int_{\pi}^{2\pi} \frac{i_a \times (1 + u_j)}{C} \mathrm{d}(\omega t) + \int_{\frac{3\pi}{2}-\Delta\theta}^{\frac{3\pi}{2}+\Delta\theta} \frac{i_a \times \Delta u}{C} \mathrm{d}(\omega t) \\ &= -\frac{4 \times I_m \times \Delta u \times \sin\Delta\theta \times \cos\varphi}{C} \end{aligned} \tag{3.8}$$

The shift in the NP voltage ΔU_c is

$$\begin{aligned} \Delta U_c &= \Delta U_{ca} + \Delta U_{cb} + \Delta U_{cc} = \int_0^{2\pi} \frac{i_a d'_{ao}}{C} \mathrm{d}(\omega t) + \int_0^{2\pi} \frac{i_b d'_{bo}}{C} \mathrm{d}(\omega t) + \int_0^{2\pi} \frac{i_c d'_{co}}{C} \mathrm{d}(\omega t) \\ &= -\frac{12 \times I_m \times \Delta u \times \sin\Delta\theta \times \cos\varphi}{C} \end{aligned} \tag{3.9}$$

As shown in Figs. 3.1 and 3.2, ΔU_c is directly proportional to $|\Delta u|$, when cosφ and $\Delta\theta$ are constant. And ΔU_c is proportional to $\Delta\theta$, when the cosφ and Δu are constant. Therefore, the NP voltage can be controlled by Δu and $\Delta\theta$.

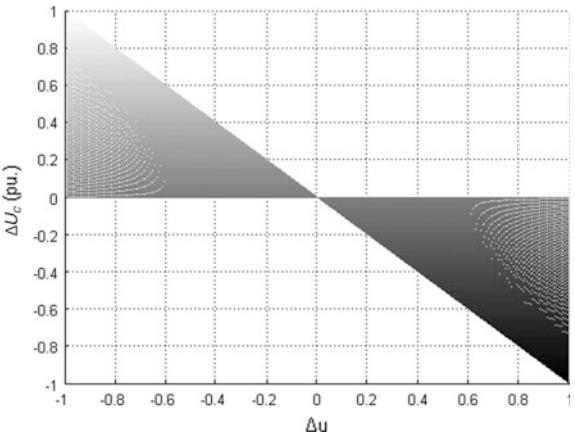


Fig. 3.1 Relationship between ΔU_c and Δu

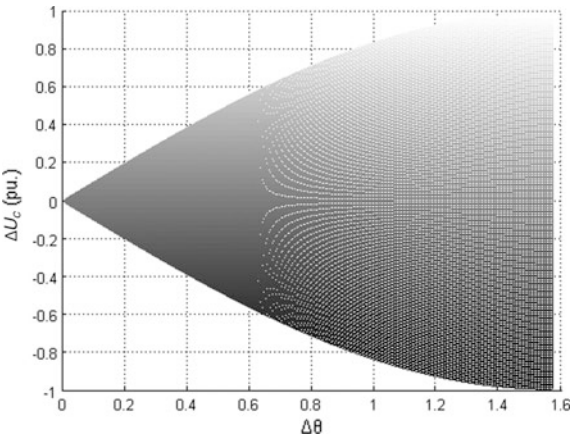


Fig. 3.2 Relationship between ΔU_c and $\Delta\theta$

3.4 Simulation Results

To demonstrate the validity of the new control method, a mode of NPC inverter is set up in this paper. In the MATLAB-Simulink mode the parameters are as follows: the amplitude of the voltage Udc and switch frequency fs of IGBT are 540 V and 1 kHz, respectively. At first, the amplitude of the capacitors' voltage is half of Udc. For the load, the resistance is 60 Ω and the inductance is 33 mH. The NP voltage is unbalanced by a parallel 2000 Ω resistor with the lower capacitor. The voltage control runs when the time in the simulation is 0.1 s.

The control results are shown in Fig. 3.3. When the NP voltage control method works, the regulation in the shift of NP voltages is quick and the NP voltage is controlled in the expected value.

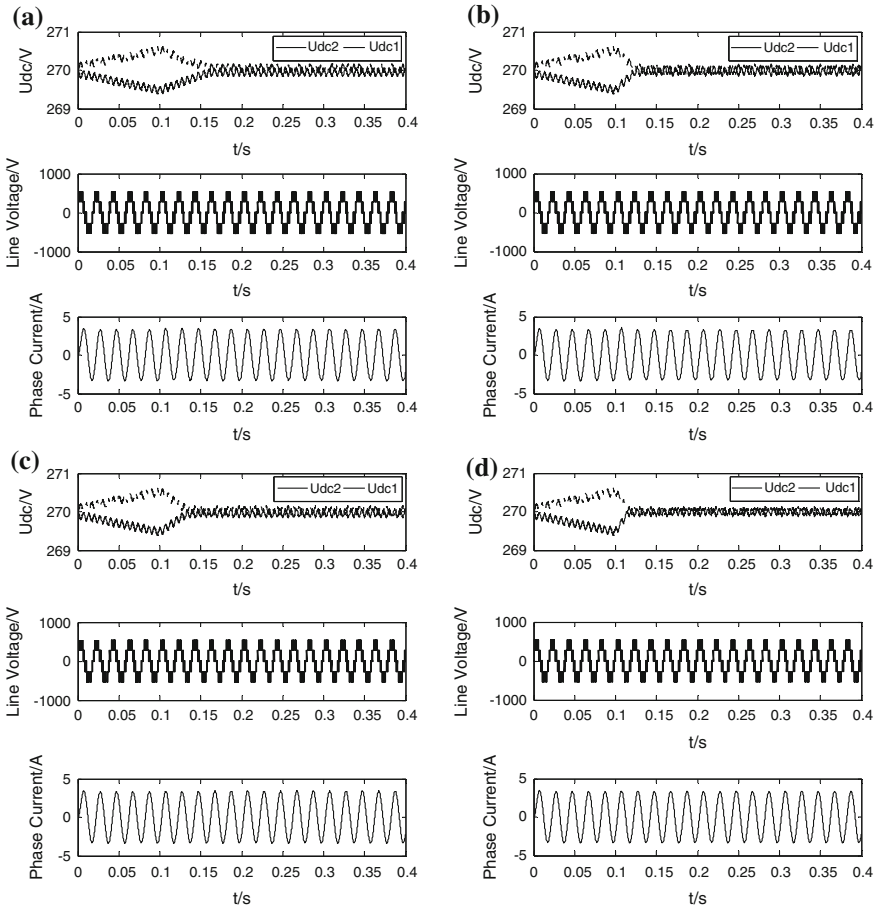


Fig. 3.3 Control results using neutral point control strategy. **a** $m = 0.9$, $\Delta u = 0.05$, $\Delta\theta = \pi/3$; **b** $m = 0.9$, $\Delta u = 0.1$, $\Delta\theta = \pi/3$; **c** $m = 0.9$, $\Delta u = 0.05$, $\Delta\theta = \pi/2$; **d** $m = 0.9$, $\Delta u = 0.1$, $\Delta\theta = \pi/2$

In Fig. 3.3a, b, the NP voltage control results by adding the voltage offset are shown when the modulation indices, $\Delta\theta$ and cos φ, are kept constant. It is obvious that the larger the voltage offset is, the stronger the NP voltage adjusting is. In Fig. 3.3a, c, the voltage control results are shown by changing the adjusting angle when the modulation indices, voltage offset, and cosφ are the same. It can be observed that the effect of the NP voltage regulation is proportional to $\Delta\theta$, which is consistent with the analysis.

3.5 Conclusions

This paper is concerned with NP voltage control method for PWM NPC inverters. The shift in NP voltage is studied. A closed-loop system is built by superimposing the voltage offset on the regulation phase in the proposed method, and the simulation results illustrate that the control method has good voltage regulation ability.

References

1. Celanovic N, Boroyevich D (2000) A comprehensive study of neutral-point voltage balancing problem in three-level neutral-point-clamped voltage source PWM inverters. IEEE Trans Power Electron 15(2):242–249
2. McGrath BP, Holmes DG (2002) Multicarrier PWM strategies for multilevel inverters. IEEE Trans Industr Electron 49(4):858–867
3. Kouro S, Lezana P, Angulo M, Rodriguez J (2008) Multicarrier PWM with DC-Link ripple feedforward compensation for multilevel inverters [J]. IEEE Trans Power Electron 23(1):52–59
4. Abu-Rub H, Holtz J, Rodriguez J, Baoming G (2010) Medium-voltage multilevel converters-state of the art, challenges, and requirements in industrial applications. IEEE Trans Industr Electron 57(8):2581–2595
5. Busquets-Monge S, Alepuz S, Rocabert J, Bordonau J (2009) Pulse Width modulations for the comprehensive capacitor voltage balance of N-level three-leg diode-clamped converters. IEEE Trans Power Electron 24(5):1364–1375
6. Yongdong L, Chenchen W (2010) Analysis and calculation of zero-sequence voltage considering neutral-point potential balancing in three-level NPC converters. IEEE Trans Industr Electron 57(7):2262–2271
7. Cobreces S, Bordonau J, Salaet J, Bueno EJ, Rodriguez FJ (2009) Exact linearization nonlinear neutral-point voltage control for single-phase three-level npc converters. IEEE Trans Power Electron 24(10):2357–2362
8. Mekhilef S, Kadir MN (2010) Voltage control of three-stage hybrid multilevel inverter using vector transformation. IEEE Trans Power Electron 25(10):2599–2606

Chapter 4
Research of Improved Adaptive Median Filter Algorithm

Weibo Yu, Yanhui Ma, Liming Zheng and Keping Liu

Abstract This paper proposes an improved adaptive median filter algorithm for highly corrupted images with salt-and-pepper noise. First of all, the improved adaptive median filter makes full use of the relationship between the gray value of the current pixel and the gray values of the neighborhood pixels. It calculates the corresponding threshold by adaptive processing. Second, all the pixels are classified into signal and noise pixels using the relationship of the difference between maximum and minimum values of the pixels and threshold. Finally, it deals with the noise and signal point using different methods. Experimental results show that the proposed algorithm has better performance for capabilities of noise removal, and preserves the image edge detail information, especially effective for the cases when the images are highly corrupted.

Keywords Threshold · Image denoising · Median filter · PSNR · Adaptive median filter

4.1 Introduction

Digital image plays an important role in many fields of science and technology, such as bioengineering, medicine, defect detection, face recognition, industrial size detection, etc. Size detection is an effective noncontact detection technology. It has been widely applied in on line detection of various kinds of artifacts. Impulsive noises can be commonly found in the sensor or transmission channel during acquisition and transfer procedures for digital signal images. Salt-and-pepper noise [1, 2] is one typical kind of impulsive noise. Factors of salt-and-pepper noise mainly include sensor and digital image equipment, electrical system factors, environmental factors in the process of image generation, etc. So it must process the

W. Yu (✉) · Y. Ma · L. Zheng · K. Liu
College of Electrical and Electronic Engineering, Changchun University of Technology, No. 2055 Yanan Street, Changchun, Jilin, China
e-mail: yu_weibo@126.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_4

noise of image; its objective is to improve the image quality and signal-to-noise ratio, and reduce the influence of image noise on the final measurement results. In order to reduce noise and not to make it fuzzy, the image must be filtered. Nonlinear filtering algorithm represented by median filter uses statistical properties of the sort of image and it is better to keep the image details. However, the median filter basically does not work when density is greater than 0.5. In order to solve this problem, some improved median filter algorithms have been proposed such as weighted median filter [3], adaptive median filter, etc. However, these algorithms treat all pixels equally, as a result these noise-free pixels could be altered sometimes. In this paper, an improved adaptive filter algorithm is proposed based on threshold. The improved adaptive filter algorithm can not only get rid of the noise with high density, but also keeps the details of the image.

4.2 The Basic Principle of Median Filter Algorithm

Median filter is nonlinear filtering technique based on the theory of order statistics [4, 5]. It was established in 1971 by Turky. Median filter, the most prominently used impulse noise removing filter, provides a better way of removal of impulse noise from corrupted images by replacing every pixel of the image by the median value of the pixels from a neighborhood chosen around the corresponding pixel. Median filtering can not only effectively remove the noise of the isolated points, but can also overcome the image detail fuzzy problem [6].

Median filter is defined as follows:

Let $x_1, x_2, \ldots, x_n$ be an array. The array is arranged in ascending order.

$$y = \text{med}(x_1, x_2, \ldots, x_n) = \begin{cases} x_{i(n+1/2)} & n \text{ is odd} \\ \frac{1}{2}\left[x_{i(n+1/2)} + x_{i(n/2+1)}\right] & \text{even } n \end{cases} \tag{4.1}$$

where y is a median of array.

In the one-dimensional case, the median filter is a sliding window with odd pixels. The pixel value of the window center is replaced by the median of the window pixels. In the two-dimensional case, it chooses some form of two-dimensional sliding window, and then the pixel gray values are arranged in ascending order. The output of filter is

$$y_{ij} = \text{med}(x_{ij}) = \text{med}\left\{x_{(i+r)(j+s)}\right\} \tag{4.2}$$

Here r, s belongs to the two-dimensional filtering window and med is a median in an array.

Two-dimensional median filters have many different window shapes and different shapes have different filtering effects. The filtering effect is not ideal if the

template is too small. It may result in the loss of boundary information if the template is too large and the image processing speed is slow because of increasing computational complexity. According to the contents and requirements of the image, it should select the appropriate window shape. In addition, the sparse distribution template saves time.

As a nonlinear filtering technique, median filter replaces the value of pixel by the median of the gray levels in the neighborhood of that pixel. The median filter not only removes noise, but also well protects the edge of the image and sharpens the detail. But the median filter changes the gray value of the image to a certain extent. Sometimes, it loses details and small target areas of the image and requires a lot of sorting operation.

4.3 The Adaptive Median Filter Algorithm Based on Threshold

4.3.1 Adaptive Median Filter Algorithm

The center pixel value is replaced by the median of the window gray levels. It is easy to filter some detailed information. In addition, it is invalid when the noise number is greater than the half the window size. To remedy these shortcomings mentioned above, it must increase the filtering window so that the image details are preserved. In adaptive median filtering algorithm [7, 8], first, all the pixels are separated into signal pixels and noise pixels. Then it processes noise and signal points by different processing methods. It removes the noise by changing the size of the filter window. Adaptive median filter algorithm gets a single value by changing the size of the window in denoising process. The single value is used to replace the center pixel value of the window.

For a size $M \times N$ of the image, let S_{xy} be the filter window of a true image at pixel location (x, y), $f(x, y)$ be the gray level at pixel location (x, y), $S_{\max}$ be the largest filter window, $f_{\min}$ be the minimum gray value, $f_{\max}$ be the maximum gray value, f_{med} be the median of the filter window. The steps of the adaptive median filter algorithm are elucidated as follows:

Step 1: Set initialize filter window size $w = 3$.

Step 2: Compute the minimum value $f_{\min}$, the maximum value $f_{\max}$, and the median value f_{med} of the current window pixel.

Step 3: If $f_{\min} < f_{\text{med}} < f_{\max}$, then go to Step 5. Otherwise increase the window size by 2.

Step 4: If $w \leq S_{\max}$, then go to Step 2. Otherwise it is considered as corrupted pixel and replace by f_{med}.

Step 5: If $f_{\min} < f(x, y) < f_{\max}$, then it is considered as uncorrupted pixel and is retained. Otherwise it is considered as corrupted pixel and replace by f_{med}.

4.3.2 Improved Adaptive Median Filter Algorithm

The adaptive median filter is effective to smooth the image of the salt-and-pepper and maintains excellent image edge and detail information. But with increase in noise density, adaptive median filter is not satisfactory. From the adaptive median filtering steps, it judges noise and signal point, which are based on the minimum and maximum values. It is easy to cause an edge or the high frequency part of the miscarriage of justice. Thus it is not good to keep image edge details.

In order to keep the image edge and texture details, it must distinguish the noise points and signal points before the filtering process. In this paper an adaptive median filter algorithm based on threshold [9] is proposed. In the polluted area, the difference in the gray value between noise pixels and other pixels in the window is larger. In the unpolluted area, however, the difference in the gray value between noise pixels and other pixels in the window is smaller. So the variable threshold with the change of the template window is adopted. The threshold could be obtained adaptively in each window.

The concrete implementation process of the dynamic threshold is elucidated as follows:

1. Compute the mean and standard except the center pixel in the template window.

$$f_{\text{ave}} = \sum_{i=1}^{n} f(x,y)/n \tag{4.3}$$

$$f_{\text{var}} = \sqrt{\sum_{i=1}^{n} \left(f(x,y) - f_{\text{ave}}\right)^2 / n} \tag{4.4}$$

2. The center pixel is subtracted from the other pixels in the template window and takes the absolute value $|\Delta f(i)|, i = 1, 2, \ldots n$.
3. Compute F_1, F_2.

$$F_1 = \frac{1}{n} \sum_{i=1}^{n} |\Delta f(i)| \tag{4.5}$$

4. Compute threshold T.

$$T = (F_1 + F_2) \times \frac{f_{\text{ave}}}{f_{\text{ave}} + f_{\text{var}}} \tag{4.6}$$

This kind of dynamic threshold method and adaptive median filtering method combine to form an improved adaptive median filter algorithm. The steps of the

adaptive median filter algorithm based on dynamic threshold are elucidated as follows:

Step 1: Set initial filter window size $w = 3$.

Step 2: Compute the minimum value $f_{\min}$ and the maximum value $f_{\max}$ of the current window pixels. If $f_{\max} - f_{\min} > T$ (T is threshold), then go to step 3. Otherwise move to the next pixel and go to Step 1.

Step 3: Find out all pixel values between $f_{\max}$ and $f_{\min}$ of the window. Let n be the number of finding all pixels values. If $n \geq w$, compute the median of these pixels. If $n < w$, increase the window size by 2 and go to Step 1.

Step 4: If the current pixel $f(x, y) = f_{\max} || f(x, y) = f_{\min}$, then it is considered as corrupted pixel and replace by f_{med}. Otherwise it is retained, and go to Step 1.

4.4 Experimental Results and Analysis

To validate the efficiency of the proposed method, we adopt the 512 × 512 standard Lena gray image as the test image based on MATLAB R2008a. Let 7 × 7 is the largest template window. Images are corrupted by salt-and-pepper noise at various noise densities. The results are shown in Figs. 4.1. and 4.2. Figure 4.1 shows the

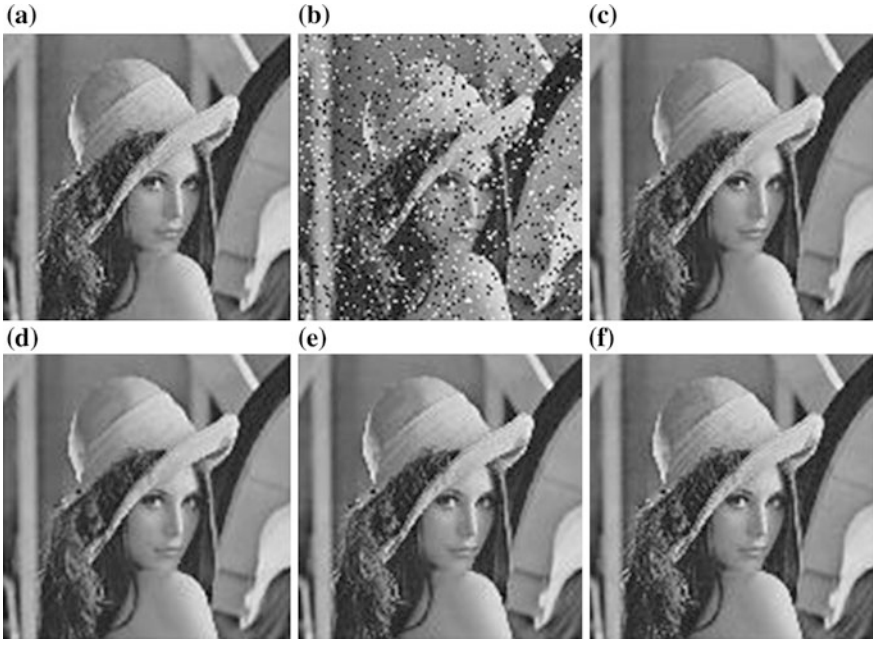


Fig. 4.1 Restoration results using various filters for image Lena corrupted by 10 % random-value noise. **a** Original image. **b** Corrupted image. **c** 3 * 3 Median filter. **d** 5 * 5 Median filter. **e** Adaptive median filter. **f** Improved adaptive median filter

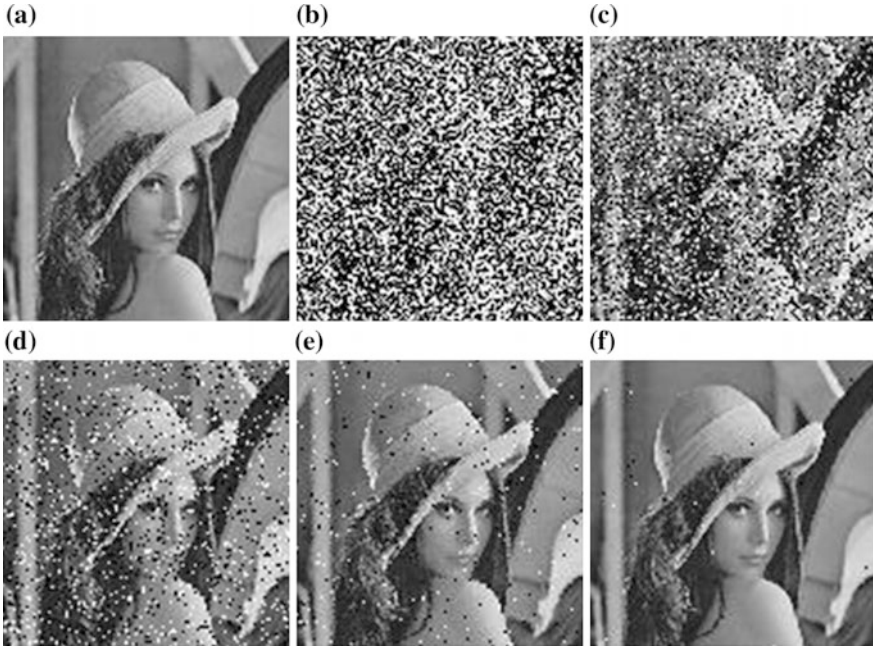


Fig. 4.2 Restoration results using various filters for image Lena corrupted by 10 % random-value noise. **a** Original image. **b** Corrupted image. **c** 3 * 3 Median filter. **d** 5 * 5 Median filter. **e** Adaptive median filter. **f** Improved adaptive median filter

restoration results using various filters for image Lena corrupted by 10 % random-value noise. Figure 4.2 shows the restoration results using various filters for image Lena corrupted by 70 % random-value noise.

From Fig. 4.1c–f, both the adaptive median filter and improved adaptive median filter not only remove the noise, but also can preserve the edge and detail information about the image when the noise density is small. The 3×3 median filter is good to maintain the image details, but it is obvious that Fig. 4.1c still has a very small amount of noise. The 5×5 median filter blurs a few details. From Fig. 4.2c–e, the median filtering algorithm is basically a failure and removing noise effect is not ideal by adaptive median filter. From Fig. 4.2f, the Lena image filtered by improved adaptive median filter algorithm proposed in this paper has better image quality than those by other methods. The image edge and detailed information is preserved well and the noise is basically removed. Finally, we can draw the conclusion that the proposed method can give better visual effect of noise restoration than adaptive median filter when the noise density is larger.

From the image point of view, with the increase in noise density, image clarity is reduced after adaptive median filter. However, the image is still relatively clear by improved adaptive median filter.

In order to evaluate the effect of filter algorithm, the peak signal noise ratio (PSNR) [10, 11] is used as the objective evaluation parameters. While its value is

Table 4.1 Comparison of PSNR values of different algorithm for image at different noise densities

Noise density (%)	Image denoising algorithm			
	3 * 3 MF	5 * 5 MF	AMF	IAMF
10	76.2126	69.8295	78.4287	90.3677
30	54.2034	62.7939	69.2851	82.6631
50	35.0041	52.2871	61.4155	75.5726
70	22.9548	32.2012	45.0059	61.5960

greater, the effect of denoising is better. It is often used in image processing, especially in image denoising. The obtained PSNR is shown in Table 4.1.

From Table 4.1, we can see that the PSNR of the improved adaptive median filter algorithm is bigger than the adaptive median filter and median filter when the noise density is the same. The PSNR moderating trend of the median filter is larger with the increase in noise density. The PSNR moderating trend of adaptive median filter algorithm and improved adaptive median filter algorithm is resembled. But the PSNR value of the improved adaptive median filter is higher than adaptive median filter in the same high density. From Table 4.1 it is seen that the improved adaptive median filter algorithm is effective for high density noise removal.

4.5 Conclusion

In this paper an improved adaptive median filter for edge and detail information preservation is proposed. The median filter and adaptive median filter fail to preserve the necessary details when noise level is bigger. With the increase in noise density, improved adaptive median filter algorithm gives better performance in comparison with all the discussed algorithms in this paper. The improved adaptive median filter not only removes the high density noise of the image, but also protects more image details and information. Experimental results show that the improved adaptive median filter improves the accuracy of noise and the fidelity of image filtering, and has better performance on the high density noise.

Acknowledgements This work is supported by Jilin Provincial Science and Technology Development Project (20140204018GX) and Jilin Provincial Youth Leading and Team Project of Science and Technology Innovation (20150519009JH). Weibo Yu is the corresponding author.

References

1. Chen C, Ding Y, Liu D et al (2010) Adaptive switch seighted mean filtering for salt and pepper noise removal. Comput Eng 36(4):210–212 (in Chinese)
2. Chen T (2014) Digital Image processing, 2nd edn. Tsinghua University Press, Beijing (in Chinese)

3. Deng X, Xiong Y, Peng H (2009) Effective adaptive weighted median filter algorithm. Comput Eng Appl 45(35):185–187 (in Chinese)
4. Ruan Q (2013) Digital Image Processing. Electronic industry press, Beijing (in Chinese)
5. Tukey J W (1974) Nonlinear (Nonsuperposable) methods for smoothing data. In: Proceedings of the IEEE electronics and aerospace systems conference, New York, IEEE, pp 673–681
6. Zhang X, Ge L (2007) A well adaptive median filtering algorithm and its application. Microcomput Inf 23(6):217–218, 237 (in Chinese)
7. Ni C, Ye M, Chen X (2006) An improved adaptive median filter algorithm. J Image Graphics 11(5):672–678 (in Chinese)
8. Wang X, Li F (2010) Improved adaptive median filtering Comput Eng Appl 46(3):175–176, 218 (in Chinese)
9. Kong F, Li Z (2009) Improvement on switching median filter. Taiyuan Sci-tech 186(7):80–81, 84 (in Chinese)
10. Huang B, Lu Z, Ma C et al (2011) Improved adaptive median filtering algorithm. J Comput Appl 31(7):1835–1837 (in Chinese)
11. Guo H, Xie G (2007) An improved method of adaptive median filter. J Image Graphics 12 (7):1185–1188 (in Chinese)

Chapter 5
Time-Dependent Reliability Sensitivity Analysis for Performance Degradation Mechanism

Shaohua Du, Yingying Yuan, Yuxiong Pan, Xu Wang and Xuhong Chen

Abstract Reliability sensitivity analysis (SA) is a method to identify the relationship between the change in reliability and the change in the characteristics of uncertain variables. Very few methods can be applied for the reliability SA for long-term degeneracy mechanism, especially when the time-dependent limit state function regarding the interested performance is implicit. This paper proposes a new method to compute the reliability sensitivity measures. First, a surrogate model called time-dependent polynomial chaos expansion (PCE) is employed to approximate the uncertain output of the model and the extensive probabilistic collocation method (PCM) and moving least squares (MLS) method-based algorithm is proposed to compute the coefficients of the time-dependent PCE. Then, the explicit time-dependent limit state function can be obtained and the reliability sensitivity measures can be calculated straightforwardly. Finally, this approach is applied to an engineering case and the time-dependent reliability sensitivity measures are obtained for the long-term degeneracy mechanism model.

Keywords Reliability sensitivity · PCE · Time-dependent · Degradation

5.1 Introduction

Reliability of the mechanism is defined as the probability that a kinematic mechanism will perform its intended movement function during a specified period of time under stated conditions. From a design engineering viewpoint, in order to maximize the mechanism performance under constrains of target mechanism reliability, the reliability-based design optimization (RBDO) should be employed [1].

S. Du (✉) · Y. Pan · X. Wang · X. Chen
CRRC Research of Electrical Technology and Material Engineering, Zhuzhou 412001, Hunan, China
e-mail: dush@teg.cn

Y. Yuan
Hunan CRRC Times Electric Vehicle Co., Ltd., Zhuzhou 412007, Hunan, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_5

However, the computation time for RBDO greatly depends on the number of random design variables [2]. Reliability sensitivity analysis (SA) aims at quantifying the respective effects of uncertain parameters onto the reliability of a model, and helps the designer to answer the following question: (1) Which uncertain parameters have the highest contribution to reliability, (2) which uncertain parameters are not so important to reliability that can be ignored when selecting the design variables, and (3) how will reliability be affected if the mean or the standard deviation of a particular variable is changed [3]? Thus, the RBDO for kinematic mechanism can benefit from the answers of the above questions.

In recent decades, many studies have been devoted to time-dependent RBDO [1, 4, 5]. The inputs (such as design parameters, loads) of mechanism vary with time, so the interested output and reliability is time-dependent consequently. In this case, calculating the time-dependent mechanism reliability sensitivity measures is much more time-consuming than static mechanism reliability SA.

However, the time-dependent reliability SA for performance degradation mechanism is rarely investigated so far, some literatures propose [6, 7] the dynamic reliability SA concentrating on the structural reliability of mechanical components. This is based on the assumption that the distribution of the load or the maximum load is known, thus the time-dependent limit state function is explicit. But it is hardly applicable for performance degradation mechanism, because the distribution of the performance is unknown since it is affected by a lot of uncertain factors.

In order to meet the needs of time-dependent RBDO for performance degradation mechanism, this paper proposes a polynomial chaos expansion (PCE)-based method to compute the time-dependent mechanism reliability sensitivity measures. First, the complicated and computational model can be replaced by the surrogate model of time-dependent PCE, the performance of the mechanism can be computed analytically. Then, the limit state function can be constructed. After that, time-dependent mechanism reliability sensitivity measures can be obtained straightforwardly.

5.2 Time-Dependent Reliability Sensitivity Analysis

5.2.1 Performance Degradation Model for the Mechanism

For the performance degradation model, the interested output of mechanism G may be influenced and changed with time. This can be defined in Eq. (5.1.1), where $\mathbf{G}$ is a vector of performance variables, $\mathbf{X} = [X_1, X_2, \ldots, X_n]$ is a vector of random variables, $\mathbf{Y}(t) = [Y_1(t), Y_2(t), \ldots, Y_m(t)]$ is a vector of stochastic process of the degradation parameters, and $\mathbf{t} = (1, 2, \ldots, n)$ is the vector of the number of tasks.

$$\mathbf{G} = g(\mathbf{X}, \mathbf{Y}(t), \mathbf{t}) \tag{5.1.1}$$

5.2.2 Time-Dependent Reliability

As for performance degradation models, the reliability of the model is time dependent. In Eq. (5.1.1), since the $\mathbf{G}$ is a random variable, if we define the threshold of a failure is z, then the time-dependent limit state function can be defined as:

$$\mathbf{F} = \mathbf{G} - z = g(\mathbf{X}, \mathbf{Y}(t), \mathbf{t}) - z \tag{5.1.2}$$

The failure state and safe state can also be defined as follows:

$$\begin{cases} F > 0 & \text{failure state} \\ F \leq 0 & \text{safe state} \end{cases} \tag{5.1.3}$$

The time-dependent probability of failure over time interval $[t_0, t_s]$ can be written as

$$P_f(t_0, t_s) = P_r\{F = g(\mathbf{X}, \mathbf{Y}(t), \mathbf{t}) - z > 0, \quad \exists t \in [t_0, t_s]\} \tag{5.1.4}$$

Thus, the time-dependent reliability can defined as follows correspondingly:

$$R(t_0, t_s) = \Pr\{F = g(\mathbf{X}, \mathbf{Y}(t), \mathbf{t}) - z \leq 0, \quad \forall t \in [t_0, t_s]\} \tag{5.1.5}$$

5.2.3 Time-Dependent Reliability Sensitivity Measures

The reliability sensitivity measures are defined as the derivative of the failure probability with respect to distribution parameters. When the time factor t is involved, the time-dependent reliability sensitivity measures can be defined as

$$S\mu_{x_i}(t) = \frac{\partial P_f(t_0, t_s)}{\partial \mu_{x_i}}, \quad t \in [t_0, t_s] \tag{5.1.6}$$

$$S\sigma_{x_i}(t) = \frac{\partial P_f(t_0, t_s)}{\partial \sigma_{x_i}}, \quad t \in [t_0, t_s] \tag{5.1.7}$$

where $S\mu_{x_i}(t)/S\sigma_{x_i}(t)$ means the reliability sensitivity of the mean/the standard deviation of variable x_i at the time instant t. μ_{x_i} and σ_{x_i} represent the mean and the standard deviation of the variable, respectively.

There are a number of ways to compute the reliability sensitivity measures, e.g., the first order and second moment method (FOSM), Monte Carlo simulation (MCS) based on linear regression method, MC integration method, and finite difference method. Among these methods, FOSM is a widely used method. At each

time instant, Eqs. (5.1.6) and (5.1.7) can be seen as the point reliability sensitivity. If the random variables are mutually independent, it is easily shown that [8]:

$$\frac{\partial p_f}{\partial \mu_{x_i}} = -\frac{\left(\frac{\partial F}{\partial x_i}\right)_{\boldsymbol{\mu}_\mathbf{x}}}{\sqrt{2\pi}\sigma_F} \exp\left[-\frac{1}{2}\left(\frac{\mu_F}{\sigma_F}\right)^2\right] \tag{5.1.8}$$

$$\frac{\partial p_f}{\partial \sigma_{x_i}} = \frac{\left(\frac{\partial F}{\partial x_i}\right)^2_{\boldsymbol{\mu}_\mathbf{x}} \sigma_{x_i}\mu_F}{\sqrt{2\pi}\sigma_F^3} \exp\left[-\frac{1}{2}\left(\frac{\mu_F}{\sigma_F}\right)^2\right] \tag{5.1.9}$$

where $\boldsymbol{\mu}_\mathbf{x} = (\mu_{x_1}, \mu_{x_2}, \ldots, \mu_{x_n})$ stands for the mean of the variables, μ_F/σ_F means the mean/the standard deviation of the limit state function.

$$\mu_F = F(\mu_{x_1}, \mu_{x_2}, \ldots, \mu_{x_n}) \tag{5.1.10}$$

$$\sigma_F = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial F}{\partial x_i}\right)^2_{\boldsymbol{\mu}_\mathbf{x}} \sigma_{x_i}^2} \tag{5.1.11}$$

And the reliability indicator can be defined as

$$\beta = \frac{\mu_F}{\sigma_F} \tag{5.1.12}$$

Thus, the time-dependent reliability sensitivity can be calculated as follows:

$$S\mu_{x_i}(t) = -\frac{\left(\frac{\partial \mathbf{F}}{\partial x_i}\right)_{\boldsymbol{\mu}_\mathbf{x}}}{\sqrt{2\pi}\sigma_F(t)} \exp\left[-\frac{1}{2}\left(\frac{\mu_F(t)}{\sigma_F(t)}\right)^2\right] \tag{5.1.13}$$

$$S\sigma_{x_i}(t) = \frac{\left(\frac{\partial \mathbf{F}}{\partial x_i}\right)^2_{\boldsymbol{\mu}_\mathbf{x}} \sigma_{x_i}\mu_F(t)}{\sqrt{2\pi}\sigma_F^3(t)} \exp\left[-\frac{1}{2}\left(\frac{\mu_F(t)}{\sigma_F(t)}\right)^2\right] \tag{5.1.14}$$

5.3 Time-Dependent PCE for the Model Output

In Eqs. (5.1.13) and (5.1.14), the performance function **G** must be evaluated to obtain the time-dependent reliability sensitivity measures. As a performance degradation mechanism, the performance function is implicit or is too complex for explicit evaluation. Thus, a surrogate model called PCE is employed here.

5.3.1 Time-Dependent PCE

PCE is a promising surrogate model that uses a set of orthogonal polynomial basis to approximate the random space of the system response. It is more accurate and effective to deal with the uncertainty compared to the conventional surrogate models, e.g., response surface method (RSM), Kriging method, and artificial neural networks method (ANN) [9]. The polynomial chaos of a performance degradation model response can be described as follows:

$$G = g(t, \boldsymbol{\xi}) = a_0(t) + \sum_{i_1=1}^{\infty} a_{i_1}(t)\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}(t)\Gamma_2(\xi_{i_1}, \xi_{i_2}) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}(t)\Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) \tag{5.1.15}$$

where $G = g(t, \boldsymbol{\xi})$ is the random system response, $a_i(t)$ is the coefficient of PCE, $\Gamma_p(\xi_{i_1}, \ldots, \xi_{i_p})$ is the polynomial of the selected basis, and p is the polynomial degree. In practical engineering, Eq. (5.1.15) can be simplified as follows:

$$g(t, \boldsymbol{\xi}) = \sum_{j=0}^{N_c-1} a_j(t)\psi_j(\boldsymbol{\xi}), \tag{5.1.16}$$

where $\psi_j(\boldsymbol{\xi}) = \prod_{i=1}^{p} \psi_{m_i}^{j}(\xi_i) = \Gamma_p(\xi_{i_1}, \ldots, \xi_{i_p})$ and N_c is the total number of PCE coefficients, which can be calculated as:

$$N_c = 1 + \frac{n!}{(n-1)!} + \frac{(n+1)!}{(n-1)!2!} + \cdots + \frac{(n-1+p)!}{(n-1)!p!} = \frac{(n+p)!}{n!p!}, \tag{5.1.17}$$

where n is the number of random variables in the model.

5.3.2 Computation of the Time-Dependent PCE Coefficients

The probabilistic collocation method (PCM) is a widely used method to compute the PCE coefficient, however, collocation methods are inherently unstable, especially with PCE of high order [10]. This paper proposes an extensive PCM based on regression, which needs to select a number of points equaling *twice* the number of coefficients, and it is more stable than the conventional method.

Suppose the mechanism will perform *Nc* tasks. The proposed method requires choosing m time points among the *Nc* tasks, i.e., $T = \{T = t_1, t_2, \ldots, t_m\}$, and constructing PCEs at each point. At $t\,(t \in T)$ time point, we can use the extensive PCM to compute the coefficients of PCE and the matrix form is Eq. (5.1.18), where

$\boldsymbol{\xi}_j$ represents the jth set of collocation points. Equation (5.1.18) can be written in a compact form of Eq. (5.1.19). The vector of coefficients can be derived as Eq. (5.1.20). At each time point, the coefficients of PCE can be calculated according to Eq. (5.1.20) (see Eq. 5.1.21), where $\mathrm{r}^{t_i}(\boldsymbol{\Xi})$ is the vector of responses at t_i, and $\mathbf{a}^{t_i}$ is the vector of the coefficients of PCE at t_i. Equation (5.1.21) can be simplified as Eq (5.1.22).

$$\begin{bmatrix} g(t,\xi_1) \\ g(t,\xi_2) \\ \vdots \\ g(t,\xi_{2N_c}) \end{bmatrix} = \begin{bmatrix} \psi_0(\xi_1) & \psi_1(\xi_1) & \cdots & \psi_{N_c-1}(\xi_1) \\ \psi_0(\xi_2) & \psi_1(\xi_2) & \vdots & \psi_{N_c-1}(\xi_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_0(\xi_{2N_c}) & \psi_1(\xi_{2N_c}) & \cdots & \psi_{N_c-1}(\xi_{2N_c}) \end{bmatrix} \times \begin{bmatrix} a_0(t) \\ a_1(t) \\ \vdots \\ a_{N_c-1}(t) \end{bmatrix} \tag{5.1.18}$$

$$\mathrm{r}(\mathrm{t},\boldsymbol{\Xi}) = \boldsymbol{\psi}(\boldsymbol{\Xi})\mathrm{a}(\mathrm{t}) \tag{5.1.19}$$

$$\mathrm{a}(\mathrm{t}) = \boldsymbol{\psi}^{-1}(\boldsymbol{\Xi}) = \mathrm{D}(\boldsymbol{\Xi})\,\mathrm{r}(\mathrm{t},\boldsymbol{\Xi}) \tag{5.1.20}$$

$$\begin{aligned} \mathbf{a}^{t_1} &= \mathrm{D}(\boldsymbol{\Xi})\,\mathrm{r}^{t_1}(\boldsymbol{\Xi}), \\ \mathbf{a}^{t_2} &= \mathrm{D}(\boldsymbol{\Xi})\,\mathrm{r}^{t_2}(\boldsymbol{\Xi}), \\ &\vdots \\ \mathbf{a}^{t_m} &= \mathrm{D}(\boldsymbol{\Xi})\,\mathrm{r}^{t_m}(\boldsymbol{\Xi}), \end{aligned} \tag{5.1.21}$$

$$\Lambda = \mathrm{D}(\boldsymbol{\Xi})\mathrm{R}(\boldsymbol{\Xi}), \tag{5.1.22}$$

After the calculation of Λ, a new matrix $\mathbf{H}$ of PCEs coefficients is created. Each row represents coefficients of the PCE, and each column represents a specific coefficient varying with steps. $\mathbf{H}$ can be then expressed in a column form

$$\mathbf{H} = [\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_{N_c-1}], \tag{5.1.23}$$

where $\mathbf{a}_0$ is the vector of the first coefficients of PCEs along with the steps. According to [11], the approximation functions of coefficients are derived as

$$\hat{\mathbf{f}}(\mathrm{t}) = \boldsymbol{\Phi}\,(\mathrm{t})\mathrm{H} \tag{5.1.24}$$

where $\hat{\mathbf{f}}(\mathbf{t}) = [\hat{f}_0(\mathbf{t}), \hat{f}_1(\mathbf{t}), \ldots, \hat{f}_{N_c-1}(\mathbf{t})]$, $\hat{f}_i(\mathbf{t})$ is defined as the approximation function of the ith coefficient. Then the conventional PCE can be translated into time-dependent PCE, which is written as follows:

$$G = g(\mathbf{t},\boldsymbol{\xi}) = \sum_{j=0}^{N_c-1} \widehat{f}_j(\mathbf{t})\boldsymbol{\psi}_j(\boldsymbol{\xi}). \tag{5.1.25}$$

5.4 Case Study

The engineering case is an airborne retractable mechanism (see Fig. 5.1). The process where the mechanism moves from I position to II position and moves back is defined as a task. At the initial design stage, the positions of and orientation of the front arm, rear arm, the link in pin-and-lug hinge and hydraulic actuator at IIposition are determined by requirements from higher level system. Therefore, the link length of pin-and-lug hinge depends on the X_{A}, the length of the hydraulic actuator depends on the X_{B}, the length of rear arm and front arm depends on the Y_{C} and Y_{D}, respectively. Thus, $X_{\mathrm{A}}, X_{\mathrm{B}}, Y_{\mathrm{C}}/Y_{\mathrm{D}}$ are all significant design variables. When the system is performing, it is under the weight of the functional device. The top hinge (pin-and-lug hinge) is working in a nonlubricated environment. As a result, the radius of the lug R_{lug} is bigger for the wear of the hinge.

The maximum driving force (MDF) is generated by the hydraulic actuator, while the output resistance force is resulted from loads and imposed on the hydraulic actuator. During the whole running process, the resistance force (RF) varies with the change of input displacement, i.e., the elongation of hydraulic actuator. If the maximum RF (MRF) is larger than the MDF, the seizure failure will happen. Thus, the time-dependent limit state function concerning seizure failure can be given as

$$\mathbf{F} = \gamma - Q(\mathbf{X}, \mathbf{Y}(t), \mathbf{t}) \tag{5.1.26}$$

where $Q(\mathbf{X}, \mathbf{Y}(t), \mathbf{t})$ is the stochastic process of the MRF, $\gamma \sim N(78{,}500, 10)$ is the distribution of MDF.

In order to compute the wear depth Δh of the top hinge, the multidiscipline simulation model including kinematics model built by ADAMS, FE model built by ANSYS, and Archard wear model is constructed (see Fig. 5.2).

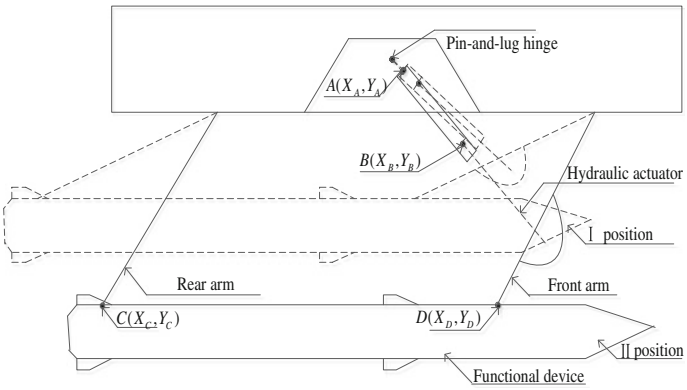


Fig. 5.1 Schematic of the airborne retractable

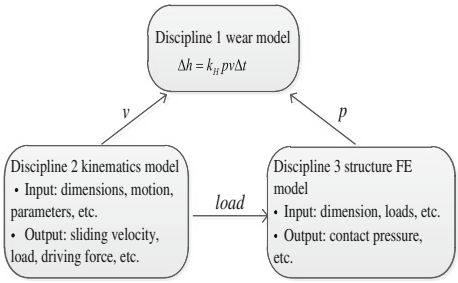


Fig. 5.2 Schematic of the computation of mechanical system

Table 5.1 The distribution of the coordinate

Symbol	Description	Unit	Distribution
X_A	Coordinate X of point A	mm	$N(-524.2641, 3^2)$
X_B	Coordinate X of point B	mm	$N(49.4975, 0.5^2)$
Y_C	Coordinate Y of point C	mm	$N(148.4924, 1^2)$
R_{lug}	Outer radius of the lug	mm	$U(15, 15.3)$

In this case, the interested output of the mechanism, MRF, is approximated by the PCE regardingthe uncertain variables $X_A, X_B, Y_C/Y_D$ and R_{lug} (listed in Table 5.1), and the reliability sensitivity of the four parameters is studied.

5.4.1 Time-Dependent PCE for the MRF

It can be seen that the degree of the PCE is $p = 3$ which is sufficient from Tables 5.2 and 5.3. The responses at every 100 steps are selected to build the PCEs, which means that the PCE are constructed at Step 1, 100, 200, and so on till 1000. Figure 5.3 is the first coefficients of time-dependent PCE. The blue points are computed by the extensive PCM and the red curve is approximated by the moving least squares (MLS). The other 34 coefficients of time-dependent PCE are also obtained by the same method.

Then the time-dependent PCEs are constructed completely. The comparative results of MRF are obtained at selected time steps which are 1, 250, 450, 650, and

Table 5.2 Comparative results of mean of the MRF along the time axis

Method	Samples	The mean of MRF ($\times 10^4$)				
		1 step	250 step	450 step	600 step	850 step
MCS	10,000	3.8263	5.1363	5.4478	7.2875	7.8872
2-order PCE	30	3.8569	5.1542	5.4601	7.2043	7.8635
Relative error	–	0.0080	0.0035	0.0023	0.0114	0.0030
3-order PCE	70	3.8344	5.1433	5.4296	7.2786	7.8711
Relative error	–	0.0021	0.0014	0.0033	0.0012	0.0020

Table 5.3 Comparative results of the variance of the MRF along the time axis

Method	Samples	The variance of MRF ($\times 10^5$)				
		1step	250 step	450 step	600 step	850 step
MCS	10,000	7.8452	8.1325	6.9467	7.6652	9.1752
2-order PCE	30	7.1056	7.9413	7.0489	7.3354	9.0854
Relative error	–	0.0943	0.0235	0.0147	0.0430	0.0098
3-order PCE	70	7.6955	8.1654	7.0159	7.8857	9.4387
Relative error	–	0.0191	0.0040	0.0010	0.0288	0.0287

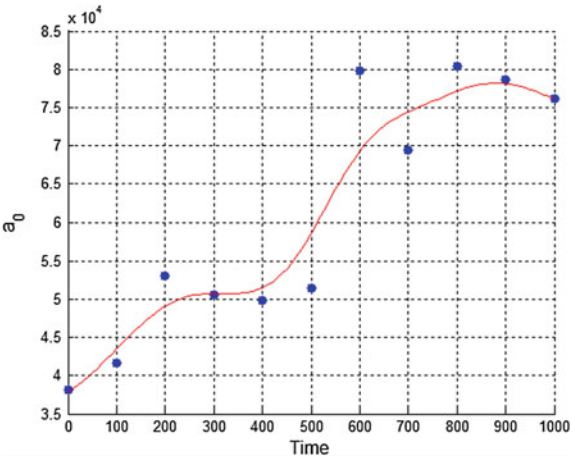


Fig. 5.3 MLS approximation curve for a_0 of the MRF

850 steps. The MCS with 10,000 samples is employed as a benchmark for the comparison. The statistics including the mean and the variance of the MRF, are shown in Tables 5.2 and 5.3, respectively. It can be seen that the results approximated by the MLS is very close to the results from MCS, which indicates the proposed method is feasible and efficient.

5.4.2 Time-Dependent Reliability Sensitivity Measures

After the time-dependent PCE for MRF is available, time-dependent limit state function can be obtained and the time-dependent reliability sensitivity measures of the four uncertain parameters' mean values can be calculated according to Eqs. (5.1.13) and (5.1.14). Figures 5.5 and 5.6 list the four uncertain parameters' time-dependent sensitivity values, respectively.

From Fig. 5.4, we can see that the reliability indicator is decreasing, which means the reliability of the mechanism is becoming worse along the time axis. Figures 5.5 and 5.6 indicate the reliability sensitivity of the four parameters is also

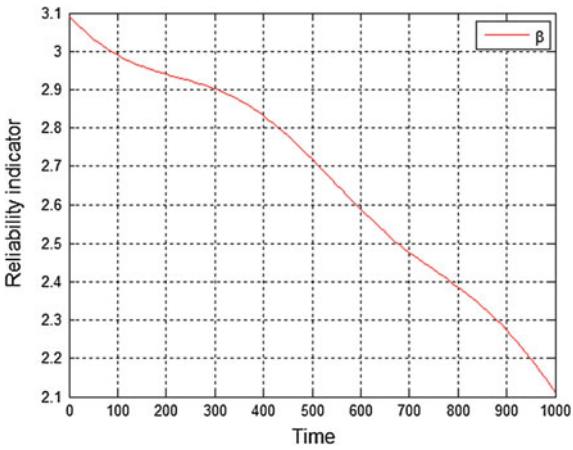


Fig. 5.4 The curve of reliability indicator β

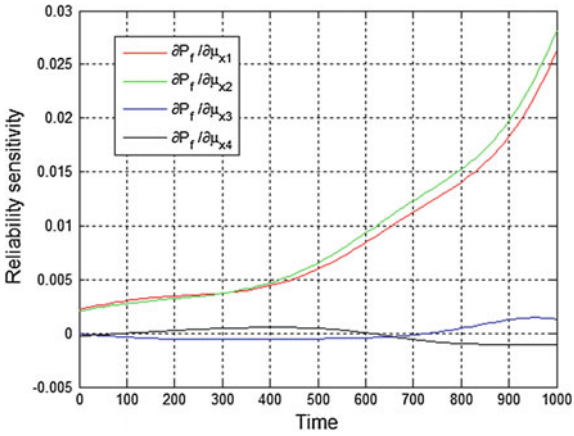


Fig. 5.5 The curve of the mean of the reliability sensitivity

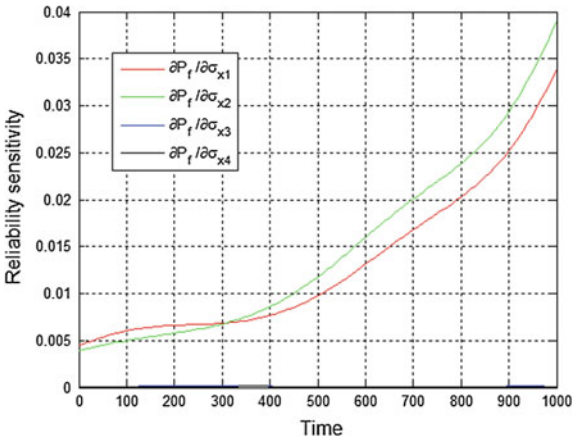


Fig. 5.6 The curve of the standard deviation of the reliability sensitivity

time-dependent. As for the mean reliability sensitivity (see Fig. 5.6), Y_c and R_{lug} have little effect on the reliability, which means we can ignore these two parameters in the work of time-dependent RBDO and reduce the dimensionality of the problem. Besides, the mean value of Y_c have a positive effect on the reliability during 1–600 tasks, and have a negative effect on the reliability during 600–1000 tasks and the designers should pay attention to this phenomenon.

5.5 Conclusion

This paper conducted the time-dependent reliability analysis for long-term degradation mechanism. The research on time-dependent RBDO and time-dependent reliability analysis can benefit from this work. The time-dependent PCE is employed when computing the reliability sensitivity measures. The extension PCM and MLS based time-dependent PCE coefficients computation method is proposed. The greatest advantage of this method is that only little simulation work is needed to construct the time-dependent PCE and we can obtain better approximation of the output of the long-term degradation mechanism. The work of computation of the reliability measures can also benefit from the proposed algorithm. The future work of this paper is to analyze more complex mechanism models to test the effectiveness of the proposed method.

References

1. Hu Z, Li H, Du X et al (2013) Simulation-based time-dependent reliability analysis for composite hydrokinetic turbine blades. Struct Multi Optim 47(5):765–781
2. Xiongming L, Zhenghui W (2011) Research on reliability sensitivity for ranking the importance of random influential factors. In: 2011 IEEE 2nd international conference on software engineering and service science (ICSESS), IEEE, pp 235–238
3. Sues RH, Cesare MA (2005) System reliability and sensitivity factors via the MPPSS method. Probab Eng Mech 20(2):148–157
4. Hu Z, Du X (2014) Lifetime cost optimization with time-dependent reliability. Eng Optim 46(10):1389–1410
5. Wang Y, Zeng S, Guo J (2013) Time-dependent reliability-based design optimization utilizing nonintrusive polynomial chaos. J Appl Math 2013(4):561–575
6. Wang XG, Zhang YM, Yan YF et al (2008) Dynamic reliability sensitivity analysis of torsion bar. Adv Mater Res 44:275–282
7. Wang XG, Wang BY, Zhu LS et al (2011) Dynamic reliability sensitivity design of mechanical components with arbitrary distribution parameters. Adv Mater Res 199:487–494
8. Melchers RE, Ahammed M (2004) A fast approximate method for parameter sensitivity estimation in Monte Carlo structural reliability. Comput Struct 82(1):55–61
9. Wang Y, Zeng S, Guo J (2013) Time-dependent reliability-based design optimization utilizing nonintrusive polynomial chaos. J Appl Math 2013(4):561–575
10. Isukapalli SS (1999) Uncertainty analysis of transport-transformation models. The State University of New Jersey, Rutgers
11. Guo J, Du S, Wang Y, et al (2014) Time-dependent global sensitivity analysis for long-term degeneracy model using polynomial chaos. Adv Mech Eng 2014(10):1–16

Chapter 6
Research on Fault Detection and Isolation of Railway Vehicle Suspension System by Impulse Track Irregularity

Xiukun Wei, Jinglin Zhang and Qiang Wei

Abstract The suspension system is one of the key components of the city rail train, whose status is directly related to the comfort, reliability, and security of the train. Urban rail train suspension system generally uses passive detection; the detection is achieved mainly by artificial periodic inspection, which wastes manpower, material, and financial resources. This paper uses the PLS (partial least squares) algorithm to perform the fault detection experiment of suspension system of urban rail vehicle under track irregularity. Under the effect of track irregularity, the experiment of suspension system of urban rail vehicle is carried out using the D–S (Dempster–Shafer) evidence theory. The research results show that the PLS algorithm can be used for fault detection in transit and fault warning, to ensure driving safety. D–S evidence theory can completely separate the fault component type.

Keywords Track irregularity · Fault detection · Fault isolation · PLS · D–S

6.1 Introduction

With the rapid development of society, the safety and efficiency of transportation are paid more attention than before. In recent years, the number of urban subways in some large cities continues to increase, especially, in Tokyo, New York, Paris, and other large cities around the world. At the same time, the safety of city rail transit has gained more and more attention. The repair of the train is to eliminate all possible security risks and to prolong the service life of the train. The purpose is to

X. Wei (✉) · J. Zhang · Q. Wei
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University,
Beijing, China
e-mail: xkwei@bjtu.edu.cn

X. Wei · J. Zhang
Shanghai Key Laboratory of Computer Software Evaluating and Testing, Shanghai Development Center of Computer Software Technology, Shanghai, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_6

ensure the safe operation of city rail trains, so as to reduce the operating costs and improve the economic benefits of city rail train [1].

The train suspension system is a key part of city rail train, and their health is directly related to the comfort, reliability, and security of the train. In the past 20–30 years, theoretical study of the dynamics of rail train, from the mechanical engineering to the comprehensive discipline including sensor, electronic, and computer technology has been made. The literature [2] has surveyed electronic, computer technology in the application of suspension and operation of the device, but is less applied in monitoring. The literature [3] has reviewed the possible prospects for the development of the inspection system, and helped developers of the state detection identify the scope of application of the technology. Through the simulation, the literature [4] has proposed the model based on the state monitoring system of the rail train. At present, the fault diagnosis method has made great breakthrough and development. The literature [5] has used the method of model driven and data driven to study the problem of fault detection and analyses at the same time. In addition, the literature [6] has offered a new method for fault detection and condition monitoring of train suspension system. The technology uses the symmetry of the train (suspension) configuration, discusses the dynamic interaction between different bogies due to the failure of suspension components, and also studies the linear and bilinear damper.

We can learn from the above literature, the fault detection methods for suspension system of urban rail vehicle continue to improve. The goal of this paper is to improve safe operation of the city rail transit by the train suspension system, discover the hidden faults in city rail train, and reduce the losses caused by accidents. First, we introduce the train vehicle suspension system. Then, using the PLS method, we research the rail train's two-line suspension system fault detection under the track irregularity excitation. The mechanism of track irregularity excitation is introduced. The principle of PLS method is explained, and the algorithm programming is realized by MATLAB software. Finally, we use the D–S evidence theory, and research on the rail train's two-line suspension system fault isolation under the track irregularity excitation. The principle of D–S evidence theory is introduced. At the same time, we conducted city rail train suspension system fault isolation based on the experimental platform for the actual data of track irregularity excitation, and obtained the fault isolation results according D–S evidence theory.

6.2 The Vehicle Suspension System

To brace the car-body and bogie, rail vehicles usually have a primary suspension (between the wheel sets and the bogie frame) and a secondary suspension (between the bogie frame and the car-body), to control the attitude of the car-body with respect to the track surface and to isolate the forces generated by the track unevenness at the wheels for ensuring safety, stability, and comfort. The vertical suspension system is described in Fig. 6.1, including four secondary dampers, four

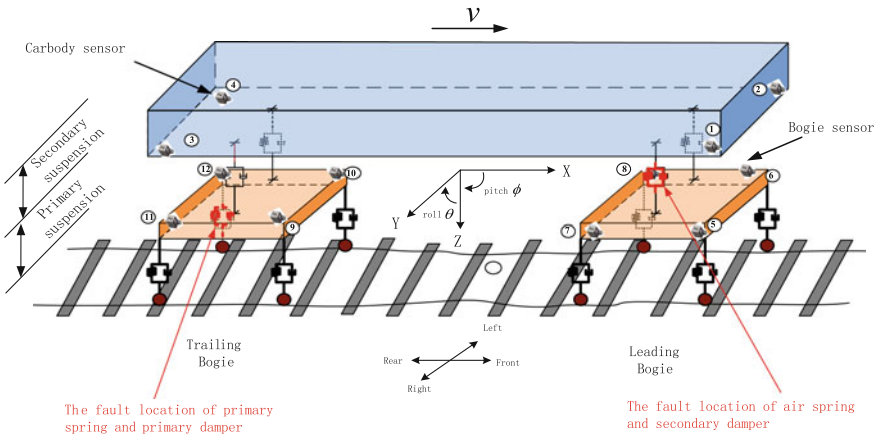


Fig. 6.1 Vertical suspension system and the sensor configuration

air springs, eight primary dampers, and eight primary springs. The red components in Fig. 6.1 represent the fault components during the simulation study. The accelerometers are numbered and take eight acceleration signals as diagnostic input.

In this paper, the D–S evidence theory and the PLS are used for fault diagnosis. The basic idea of PLS is that the system (or process) is driven by a small number of latent variables. These latent variables cannot be measured directly, but can be estimated from the data. They are linear combinations of the process observed variables and the quality variables [7]. As shown in types (6.1) and (6.2).

$$X = \hat{X} + E = \sum_{a=1}^{A} t_a p_a^T + E = TP^T + E \tag{6.1}$$

$$Y = \hat{Y} + F = \sum_{a=1}^{A} t_a q_a^T + F = TQ^T + F \tag{6.2}$$

Dempster–Shafer (D–S) evidence theory transfers the relationship between proposition and aggregate into one-to-one relationship and transfers the uncertainty of proposition into the uncertainty of aggregate. The detailed principle of the D–S evidence theory can be found in [8–11].

6.3 Fault Detection of Suspension System

Track irregularity is the deviation of position and shape caused by its structure and external factors. Rail vehicle wheels will be influenced by unsmooth track and other excitations, and generate continuous vibration. Due to the universal track

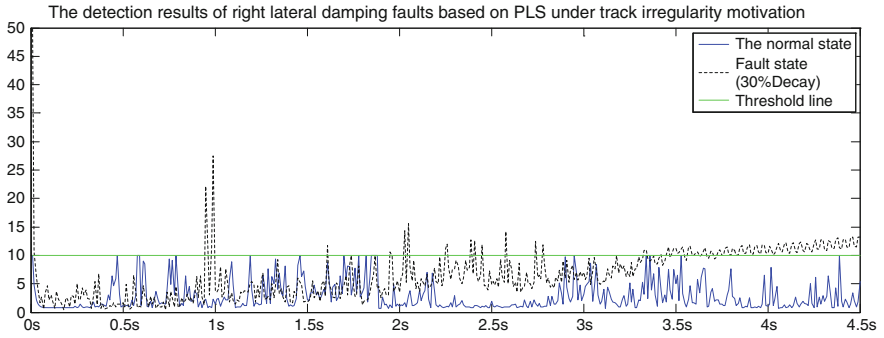


Fig. 6.2 The detection results of *right lateral* damping faults based on PLS under track irregularity motivation

irregularity, the fault diagnosis of suspension system based on track irregularity can realize real-time detection during operation.

Fault detection procedures of city rail train suspension system based on PLS algorithm is developed in MATLAB, fault detection is done on the processed data, and statistical index for every two seconds is recorded. If there are more than 20 points beyond the threshold line, it is thought that there is a fault which is detected and makes a fault alarm.

When the vertical and lateral damper of secondary suspension system under the influence of track irregularity fails, the fault detection results based on PLS are shown in Fig. 6.2. The green thick solid line represents the alarm threshold, the blue thin line indicates the detection index of suspension system under normal condition, different colors of the dotted lines represent the test results of component performance in corresponding suspension systems under minor fault, middle fault, and serious fault.

For air spring failure, the fault detection results based on PLS are shown in Fig. 6.3. The green thick solid line represents the alarm threshold, the blue thin line indicates the detection index of suspension system air spring under normal condition, different colors of the dotted lines represent the test results of component performance in corresponding suspension systems under minor fault, middle fault, and serious fault.

From the fault detection results in Fig. 6.2, we learn that under the excitation of track irregularity, when there is a minor fault on front lateral damper, the fault can be detected at the fifth second, the warning delay is long. The alarm distance is long when the fault occurs, explains the general detection effect. The occurrence of moderate and severe fault, the fault can be detected within the first second, which is a good detection effect.

From the fault detection results in Fig. 6.3 we can learn that under the excitation of track irregularity, when there is fault on the left rear air spring, the fault can be detected at the fourth second, indicating that the warning time is slight delayed and

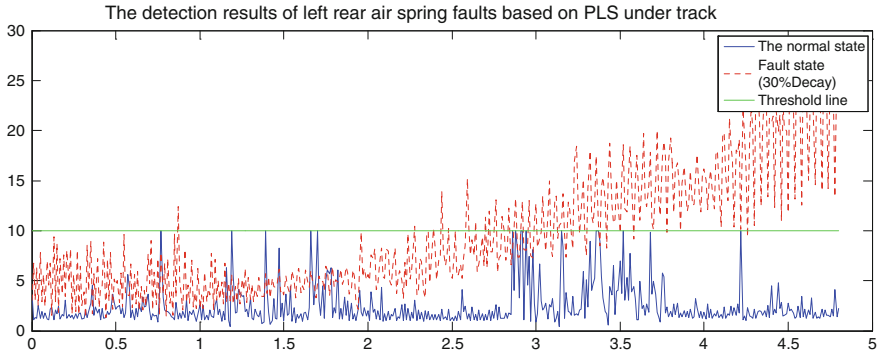


Fig. 6.3 The detection results of *left* rear air spring faults based on PLS under track irregularity motivation

the detection effect is good, when serious fault occurs, the fault can be detected within the first second, which is a good detection effect.

Through the 20 groups of experiments, the time from failure to alarm and the success rate of PLS in fault detection of city rail train suspension system with different work conditions is calculated. The success rate under the excitation of track irregularity based on PLS method is $P_{I,PLS} = n/20$, n is the number of successful fault detections. The alarm time is $T_{I,PLS} = \sum t_i/20$, t_i is the detection time from the occurence of fault to the occurence of alarm.

6.4 Fault Isolation of Suspension System

Fault diagnosis includes fault detection and fault isolation. The previous chapter uses fault detection method, which accomplishes the fault detection of rail train suspension system, and its effect is obvious. This chapter completes fault isolation of suspension system that bases on the D–S evidence theory, which means to determine the location and level of the fault, and finally to complete the fault diagnosis of the suspension system which is motivated by track irregularity.

When a fault occurs on the train suspension system, some changes will happen in time-domain sequences, which are called the fault features. These changes will become more significant in the frequency domain. The residual sequences in the time domain are generated, and the relative features through fast Fourier transform (FFT) are obtained, which will be matched with the similarity as signature sequences. This paper considers two different typical faults which includes damper and air spring of secondary suspension system in the right-front side.

The four degrees amplitude-frequency of faults in damper and air spring of secondary suspension system in the right-front side are sketched in Figs. 6.4 and 6.5. It is clear that there are some obvious differences between the amplitude

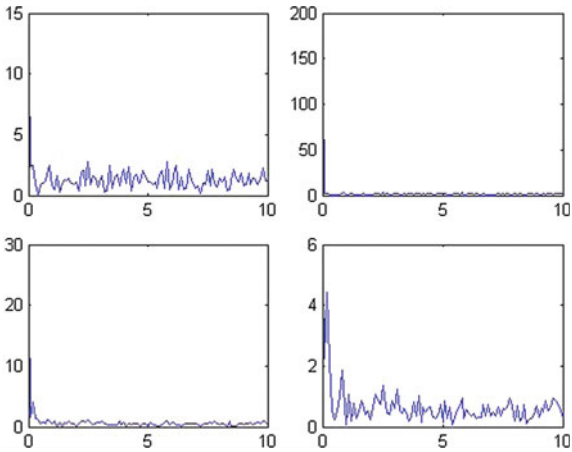


Fig. 6.4 Amplitude-frequency characters of the faults in air spring of secondary suspension system in the *right-front side*

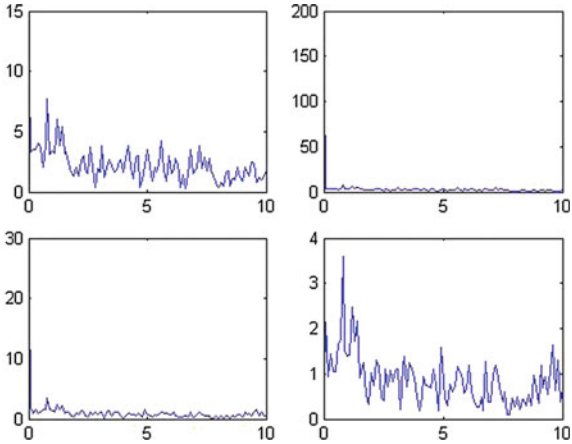


Fig. 6.5 Amplitude-frequency characters of the faults in tamper of secondary suspension system in the *right-front side*

frequency sequences of these two faults in energy distributions of different frequency bands, which is the precondition of fault isolation based on the D–S evidence theory.

The acceleration data of train body from the excitation of track irregularity is collected. According to the type and level of the faults, establishing standard feature library of suspension system with the excitation of track irregularityincludes two different fault features.

Data fusion is accomplished by the D–S evidence theory in MATLAB. The standard feature library which bases on excitation of track is loaded irregularity. Then, merge the data of different fault conditions which comes from simulation experiment platform of suspension system with standard feature library that is stimulated by track step. And finally the fault isolation results are obtained.

Group the faults that need to be isolated and reconcile them with standard features. For example, when a fault has occurred on the damper of secondary suspension system in the right-front side, obtain the data T1 (t_c_fr_z_20) corresponding to F1, and merge it with standard feature library by the D–S evidence theory. The fault isolation results are obtained.

Base on the D–S evidence theory to extract features of the fault T1 from suspension systems of urban rail train, which is stimulated by excitation of track irregularity. The results are consistent with the fault. Therefore, the fault on the damper of secondary suspension system in the right-front side is sketched in Fig. 6.6.

The accuracy of fault isolation is calculated that bases on the results discussed above. According to the scope of fault isolation, this paper has done statistics about accuracy of different kinds of fault isolations. Based on the results of fault isolation, this paper can determine which component or which part of the train has a fault. Besides, the location and level of a fault can also be determined. The success rate refers to the ratio between the number of successful fault isolation and the total number of test, which is sketched in Table 6.1

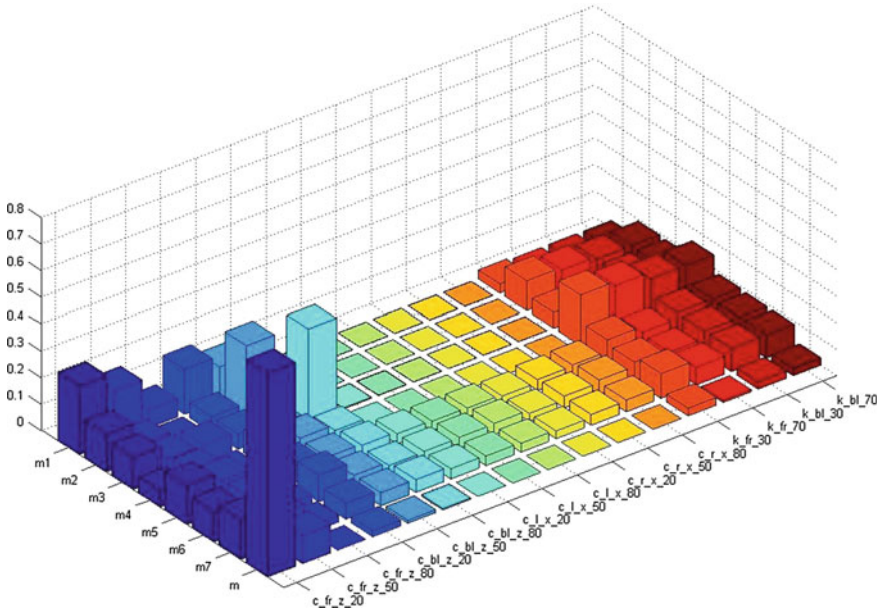


Fig. 6.6 The results of fault isolation for t_c_fr_z_20

Table 6.1 The success rate of fault isolation based on D–S evidence theory under excitation of track irregularity

Fault type	Component failure (%)	Failure location (%)	Fault level (%)
Success rate	100	75	68.75

6.5 Conclusion

This paper has studied the fault detection of the incentive suspension system under the track irregularity, using PLS method and combines with the actual simulation data,accomplishes the fault detection. The analysis of two kinds of fault detection methods, the methods based on PLS can effectively detect the fault of urban rail train suspension system. At the same time, for moderate and severe warning, the response of these two kinds of methods are efficient and quick, but the detection effect for a small fault is not good, sometimes the warning is delayed and the fault cannot be detected. Of course, considering that the small fault has little influence on the safety and stability of traffic, the two methods can be used in fault detection, fault warning, and ensure traffic safety during the running time.

The fault isolation method which bases on D–S evidence theory when combined with the actual simulation data, do the fault isolation for suspension components and identify the fault location, fault level, and component size. From the results of fault isolation it can be found that D–S evidence theory can completely separate the fault types and components. The isolation effect of fault location and fault level is very good.

Acknowledgments This work is partly supported by Chinese National Key Technologies R&D program (Contract No. 2015BAG13B01) and State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2014ZZ005). Shanghai Key Laboratory of Computer Software Evaluating and Testing (Shanghai Development Center of Computer Software Technology).

References

1. Zonghua H, Songzi W, Qiguang H (2007) The operation and repair of city rail transit vehicles. China Building Industry Press, Beijing (in Chinese)
2. Bruni S, Goodall R, Mei TX, Tsunashima H (2007) Control and monitoring for railway vehicle dynamics. Veh Syst Dyn 45(7–8):765–771
3. Goodall R, Roberts C (2006) Concepts and techniques for railway condition monitoring. In: IET international conference on railway condition monitoring, pp 90–95
4. Li P, Goodall R (2004) Model-based condition monitoring for railway vehicle systems. Master's thesis, Control, University of Bath, London, UK
5. Wei X, Jia L, Liu H (2013) A comparative study on fault detection methods of rail vehicle suspension systems based on acceleration measurements. Veh Sys Dyn Int J Veh Mech Mobility 51(5):700–720
6. Mei TX, Ding XJ (2008) A model-less technique for the fault detection of rail vehicle suspensions. Veh Syst Dyn 46(1):277–287

7. Tang Q, Tang J (2005) Partial least squares regression analysis in modeling analysis of uniform design experiment application. Appl Stat Manage 25(5):45–49 (in Chinese)
8. Chenglin W, Xiaobin X (2012) Multi-source uncertain information fusion theory and application. Science Press, Beijing (in Chinese)
9. Haina J (2009) The information fusion method of fault diagnosis based on Desert-Smarandache theory. Hangzhou Dianzi University, Hangzhou (in Chinese)
10. Francis T, Varma H, Fadaie K et al (2003) Confusion in data fusion. Int J Remote Sens 24 (4):627–636
11. Hui Y (2010) Research and application of multi sensor information fusion technology. Harbin Engineering University, Harbin 3 (in Chinese)

Chapter 7
Metro Station Safety Status Prediction Based on GA-SVR

Zhenyu Zhang, Yong Qin, Xiaoqing Cheng, Lei Zhu, Linlin Kou, Jian Li and Fang Sun

Abstract Metro station is one of the most important parts in the metro system. Once an accident occurs, it will cause a lot of casualties and property losses. To ensure the safety of the station operation, accurately predicting the safety status of metro station is of great significance. The paper analyzed the influencing factors of metro station's safety status, and established the metro station safety status prediction model based on GA-SVR. The results show that safety status prediction model based on GA-SVR can predict the trend of metro station safety status accurately, which makes the change from passive safety to active safety and has good practical value.

Keywords Metro station · Safety status prediction · Genetic algorithm · Support vector regression

7.1 Introduction

Metro station is an important part of metro system, for the passengers to alight, wait, and move to the places. Metro station has characteristics of strong operating professionals, complex technical equipment, huge traffic. Once accidents occurr, because of the difficult evacuation and rescue, it will cause a large number of casualties and property losses. Therefore, it is a great responsibility to protect the safety of metro station operation. Safety, is the lifeline of the metro station, had to place work safety in the first place.

Metro station system is a complex dynamic system. The general safety state evaluation method cannot meet the demands of the precise safety status information

Z. Zhang · Y. Qin (✉) · X. Cheng · L. Zhu · L. Kou
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
e-mail: qinyong2146@126.com

J. Li · F. Sun
Beijing Traffic Control Center, Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_7

analysis, dynamic security evaluation, and the system safety analysis. Therefore, we propose a quantitative system safety state evaluation of metro station security domain theory and method.

Safety region [1] is a quantitative model to describe the area that the whole system operate safely and stably from the viewpoint of the region. The relative relationship between the boundary of safety region and state point of the system can provide quantitative safety margin and optimal control information of the system in different conditions.

From the system perspective, the metro station security domain is a state in metro station's system characteristic variables (such as passenger flow, equipment, environment, and management, etc.) of variable space, which is used to determine whether the station operation state is a secure area. In metro station systems, security domain refers to the state of the station operation state space, determined by characteristic variables, and whether the object is used to identify the current operation state security area or not. As shown in Fig. 7.1, if the running state of the objects is in the security domain (figure in the green area), we can determine that the object operation state is normal, otherwise think that object does not run properly. If there are hidden dangers or external interference it may lead to the running state of the objects near security domain boundaries, that object may have hidden trouble, has a tendency to malfunction, needs the relevant prevention and control measures in time, the running state of the object is away from security domain boundaries (as shown in figure of dotted arrow). If the running state of the objects is in the approach of the security domain boundary position, without timely and correctly the prevention and control measures, then object running condition will worsen, evolution to the security domain (as shown in figure of fault trajectory), leading to safety accidents. If the running state of the objects is in the approach of the security domain boundary position, without timely and correctly prevention and control measures, the object running condition will worsen, evolution to the security domain (as shown in figure of fault trajectory), leading to safety accidents.

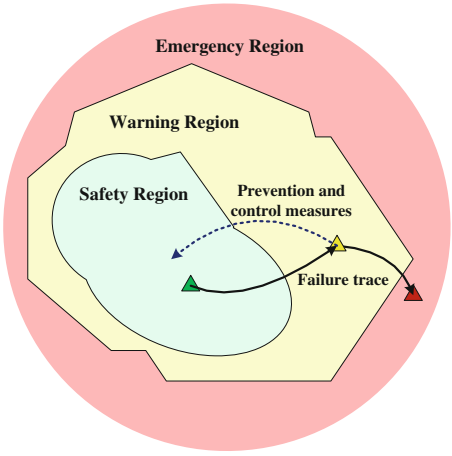


Fig. 7.1 Schematic of security region of metro station

Therefore, in order to prevent the station near position further deterioration in the security domain boundaries, accurately forecasting the metro station safety status is the inevitable trend to realize active safety research. Accurately predicting the state of safety can reduce the unsafe factors of the station, the security risks of accidents for passengers, implemented from "passive safety" to "active safety," to prevent accidents, reduce passengers' life and property losses, to ensure the safety of the station operation is of great significance.

At present, the metro station security state prediction method with time series prediction method, neural network prediction method, the gray model method, and support vector machine (SVM) such as [2, 3]. Support vector regression (SVR) in the aspect of safety state prediction has a fast convergence rate, absolute error is small, the fitting ability, and high accuracy etc. [4]. Therefore, the SVR model is chosen to forecast metro station safety status.

7.2 Safety Factor of Metro Station

The metro station is a huge and complex system. According to the former research [5], people, equipment, environment, management reflect the safety status of metro station and have influence with each other. But, we find that the number of accidents, the casualties, and economic loss during the accidents reflect the safety [6]. Therefore, the safety factors of metro station are considered from the five indexes such as passenger flow, equipment, environment, management, and accident. The calculation formulas of each factor are put forward to quantify the new factor. Table 7.1 gives the calculation formulas of each factor in detail.

7.3 Prediction Model Based on GA-SVR

7.3.1 SVM Model

SVM was first proposed in 1995 by Corinna Cortes, is mainly used in small sample, nonlinear and high-dimensional pattern recognition and function fitting, and operates in strict accordance with the structural risk minimization principle, make up for the deficiency of the neural network, has better generalization ability, is a research hotspot in recent years. At present the method is mainly divided into two application directions, support vector machine classification (SVC) and SVR.

The basic idea of SVM is explained as follows: As to the problem of R^n classification (regression) in n-dimension Euclidean space, we find real function $g(x)$ in the R^n, and use the decision function $f(x) = \mathrm{sgn}[g(x)]$ to calculate the corresponding output y of input x. The determination of $g(x)$ is through the construction and solving of the nonlinear planning problem dual to the original classification

Table 7.1 The explanation and calculation method of factors [6]

The explanation of factors	Calculation formula and parameters
$C_1/C_2/C_4/C_5/C_6$ is the load degree of entrance AFC/exit AFC/stairway/passageway/escalator at peak times of a cycle	$\begin{cases} C = \sum_{i=1}^{n} \lambda_i S_i \\ S_i = \frac{Q_i}{C_i} \end{cases}$ S_i—the load degrees of the ith entrance AFC/exit AFC/stairway/passageway/escalator, λ_i—the weight of the ith entrance AFC/exit AFC/stairway/passageway/escalator, Q_i—passenger numbers at the ith entrance AFC/exit AFC/stairway/passageway/escalator, C_i—the ith entrance AFC/exit AFC/stairway/passageway/escalator capacity, n—entrance AFC/exit AFC/stairway/passageway/escalator numbers
C_3 is the ratio of passenger numbers and the area of the waiting zone on the platform at peak times of a cycle	$C_3 = \frac{(Q_1 \cdot T_1 + Q_2 \cdot T_2)}{S} \cdot \Phi$ Q_1—passenger numbers enter the metro station, people/s, Q_2—passenger numbers exit the metro station, people/s, S—the area of the waiting zone on the platform, m^2, T_1—the interval time between trains arriving, s, T_2—The longest walking time between the platform and the hall, s, Φ—The uneven coefficient of passenger flow density on the platform
$C_7/C_8/C_9/C_{10}/C_{11}/C_{12}$ is the ratio of escalator/drainage/shielding door/FAS/lighting/air conditioning system failure numbers and escalator/drainage/FAS/lighting/air conditioning system numbers in a cycle	$C = \frac{M}{N}$ M—escalator/drainage/shielding door/FAS/lighting/air conditioning system failure numbers, N—escalator/drainage/shielding door/FAS/lighting/air conditioning system numbers
$C_{13}/C_{14}/C_{15}/C_{16}/C_{17}$ is the ratio of the measurement value of temperature/humidity/particles smaller than 2.5 μm (PM2.5)/particles smaller than 10 μm (PM10)/CO_2 and the standard value	$C = \frac{c}{S}$ c—the measurement values, S—the standard values
C_{18} is the comprehensive assessment of safety management in a cycle	$C_{18} = 1 - \frac{c}{1000}$ c—the score given by the experts
C_{19} is the accident evacuation time of the platform layer at peak times of a cycle	$C_{19} = \left[\frac{Q_1 + Q_2}{0.9[A_1(N-1) + A_2 B]} + 1\right]/6$ Q_1—passenger numbers on the train, Q_2—passenger numbers waiting for the train on the platform, A_1—escalator capacity, people/min·m, A_2—

(continued)

Table 7.1 (continued)

The explanation of factors	Calculation formula and parameters
	stairway capacity, people/min·m, N—escalator numbers, B—stairway width, m, 0.9 — reduction factor
C_{20} is the number of accidents in a cycle	$C_{20} = \sum_{i=1}^{5} w_i s_i$ w_i—the weight of ith accident, S_i—the number of ith accident, accidents include special major accidents, major accidents, accidents, accident insurance and general accident.
C_{21} is the casualty rate in a cycle	$C_{21} = \frac{n}{N}$ n—the number of casualties during statistical period, N—the passenger flow in the station
C_{22} is the economic loss caused by accidents in a cycle	$C_{22} = \sum_{i=1}^{5} w_i s_i$ w_i—the weight of ith accident, S_i—the economic loss of ith accident

(regression) problem. Use the nonlinear project $\phi(\cdot)$ to project the input x to the high-dimensional feature space (Hilbert space), and solve the linear planning problem in this space.

High-dimension project is realized specifically through kernel function $K(x_i, x) = \phi(x_i)^T \phi(x)$, where the kernel function satisfies the Mercer theory, namely it can be expanded by positive coefficient α_m as

$$K(u, v) = \sum_{m=1}^{\infty} \alpha_m \psi(u)\psi(v) \tag{7.1}$$

The sufficient and necessary condition is that formulas (7.2) and (7.3) are established.

$$\iint K(u, v)g(u)g(v)\mathrm{d}u\mathrm{d}v > 0 \tag{7.2}$$

$$\int g^2(u)\mathrm{d}u < \infty \tag{7.3}$$

Kernel function in SVM algorithm plays a key role; the commonly used function is shown as follows:

1. Polynomial kernel function

$$K(x_i, x_j) = \left[(x_i \cdot x_j) + 1\right]^q \tag{7.4}$$

   Equation (7.4) is a q-th order kernel function.
2. Gaussian radial basis kernel function (RBF)

$$K(x_i, x_j) = \exp\left[-\left|x_i - x_j\right|^2 / (2\sigma^2)\right] \tag{7.5}$$

   where σ is the kernel parameter. Gaussian RBF is the most used kernel function.
3. Sigmoid kernel function

$$K(x_i, x_j) = \tanh\left[v(x_i \cdot x_j) + c\right] \tag{7.6}$$

   Sigmoid kernel function is different from polynomial kernel function and Gaussian RBF, which needs specified v and c to satisfy the Mercer theory.

According to the theories of kernel function, the decision function in R^n space can be expressed as

$$f(x) = \mathrm{sgn}\left[\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right] \tag{7.7}$$

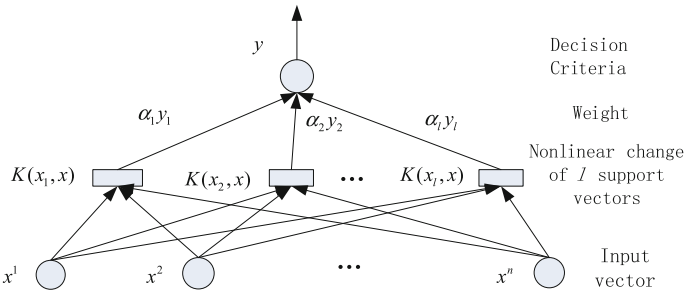


Fig. 7.2 Structure of support vector machine

where, l is the number of support vectors, $\alpha_i y_i$ is the weight, and b is the threshold.

The form of SVM is like a neural network, where output is the linear combination of intermediate nodes (each node correspond to a support vector), the structure of which is shown in Fig. 7.2.

7.3.2 SVR Model

The basic idea of SVR is, given the input and output sample $(x_1, y_1), (x_2, y_2) \ldots (x_t, y_t)$, where $x_t \in R^n$ and $y_t \in R$ (t = 1,2, …, N indicating the number of sample points), we use SVR-trained regression to fit the function $f(x)$, limiting the difference between the function output and the target output of sample points under the given threshold and guaranteeing the smoothness of $f(x)$ [8]. The corresponding function model of the jth output of the sample above is expressed as formula (7.8), where w is the project coefficient, $b \in R$, and $\phi(\cdot)$ is the nonlinear project of high-dimensional space.

$$f(x_i) = w \cdot \phi(x_i) + b \tag{7.8}$$

For a given sample (x_t, y_t), where x_t is input and y_t is output, SVR can be described as the following optimization problem.

$$\min \frac{\|w\|^2}{2} + C \sum_{i=1}^{t} \left(\xi_i + \xi_i^* \right) \tag{7.9}$$

$$\text{s.t.} \begin{cases} y_i - (w \cdot \phi(x_i)) - b \le \varepsilon + \xi_i \\ (w \cdot \phi(x_i)) + b - y_i \ge \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{7.10}$$

The loss function in formula (7.9) is Vapnik ε insensitive loss function, which equals 0 when $\left| y_i - \left(\phi(x_i)^T w + b \right) \right| \le \varepsilon$, and equals $\left| y_i - \left(\phi(x_i)^T w + b \right) \right| - \epsilon$ when

$\left|y_i - \left(\phi(x_i)^T w + b\right)\right| > \epsilon$. ξ_i, ξ_i^* are slack variables, and constant $C > 0$ is penalty factor.

To solve the optimization problem, we introduce Lagrange function.

$$\begin{aligned} L = \frac{\|w\|^2}{2} + C\sum_{i=1}^{t}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{t}\alpha_i(y_i - (w\cdot\phi(x_i)) - b + \epsilon + \xi_i) \\ - \sum_{i=1}^{t}\alpha_i^*\left(y_i - (w\cdot\phi(x_i)) - b + \epsilon + \xi_i^*\right) - \sum_{i=1}^{t}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right) \end{aligned} \tag{7.11}$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$. According to the optimization condition, the formula can be changed into the following formula:

$$\left.\begin{aligned} \frac{\partial L}{\partial b} &= \sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right) = 0 \\ \frac{\partial L}{\partial w} &= w - \sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right)x_i = 0 \\ \frac{\partial L}{\partial \xi_i^{(*)}} &= C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \end{aligned}\right\} \quad t = 1, 2, \ldots, N \tag{7.12}$$

By substituting the formula into L, dual parameters can be calculated.

$$\begin{aligned} W(\alpha, \alpha^*) = -\frac{1}{2}\sum_{i,j=1}^{t}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)K(x_i, x_j) + \sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right)y_i \\ - \sum_{i=1}^{t}\left(\alpha_i + \alpha_i^*\right)\epsilon \end{aligned} \tag{7.13}$$

Under the condition $\sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right) = 0$, solve for the maximum of $W(\alpha, \alpha^*)$, the parameters α_i, α_i^* are calculated, and finally figure out the regression function.

$$\begin{aligned} f(x) = w\cdot\phi(x) + b &= \sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right)\phi(x_i)\cdot\phi(x) + b \\ &= \sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right)K(x_i, x) + b \end{aligned} \tag{7.14}$$

7.3.3 Prediction Model Based on GA-SVR

Genetic algorithm (GA) is a reference to biological evolution evolved to a high efficient and parallel global optimization method, can get the optimal solution by less experience knowledge [9]. Genetic algorithm for optimization process involves

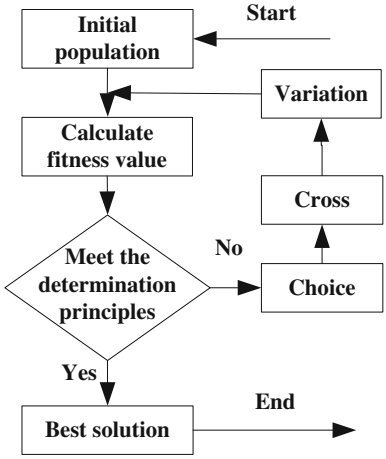


Fig. 7.3 Optimize solution process of genetic algorithm

chromosome coding, fitness function, selection operator, crossover operator, mutation operator, and an end to the principle. Genetic algorithm to optimize the basic process is shown in Fig. 7.3.

The SVR to nonlinear regression problems through nonlinear $\sigma\phi(\cdot)$ mapping train the input mapped to high-dimensional feature space, and then set up a linear regression function in the space. But generally $\phi(\cdot)$ consists of kernel function, so the selection of kernel function and parameters to a large extent affect learning and generalization ability of SVR. As stated earlier, the commonly used kernel function has a linear kernel function, sigmoid function, Gaussian radial basis function (RBF). As document [10] research shows that relative to other kernel functions when prior knowledge lack of RBF kernel function is the function with better effect, so the choice of RBF kernel function and its concrete form the formula (7.5).

When the kernel function is determined, parameter selection is an important step, such as penalty parameter C, epsilon insensitive factor ε, and kernel parameter σ. C is complexity and generalization ability of the compromise of regression model parameter. Its value is too large, said VC (Vapnik–Chervonenkis) rights [7] is too small, the model generalization ability is poor. Once its value is too small, it will cause the epsilon failure and cause the training error.

The size of the ε determines the number of SVM and influences the precision of the model. σ is the nuclear width, can reflect distribution features of training samples, the smaller the value, RBF function fitting performance is better, but it starts to cause model generalization ability.

GA-SVR forecasting model includes the following four steps, modeling flow diagram as shown in Fig. 7.4.

1. Obtaining experimental data samples, to filter the data samples, determine the sample training data, and testing data samples.
2. The GA algorithm is used to determine the SVR parameters. The SVR parameters contains the initialization of SVR parameters C, nuclear epsilon ε

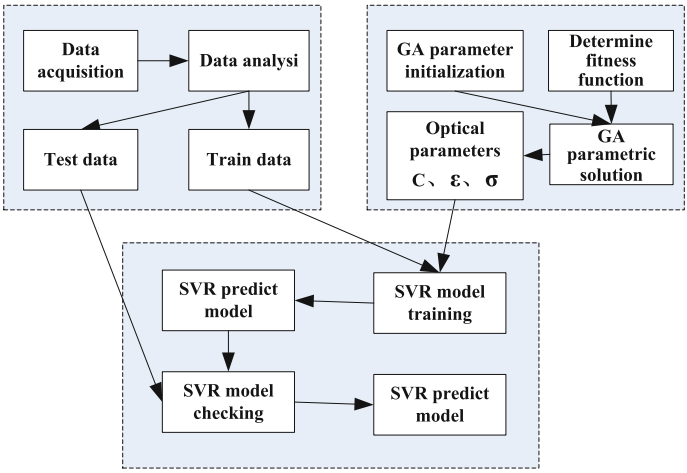


Fig. 7.4 Process of GA-SVR model

and upper and lower σ, the fitness function of GA, and the population size of genetic and iterative algebra, etc.

3. The parameters of C, ε, σ by GA training as well as the training sample data as input for the SVR model training.
4. Test data will be input to the prediction model, obtained from the training test prediction precision of the model, and we finally get a GA-SVR forecasting model.

To assess the performance of the neural network, this paper selects root-mean-squared error and the correlation coefficient R as model evaluation.

$$\mathrm{RMSE}(y, y_m) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y(i) - y_m(i))^2} \tag{7.15}$$

$$R(y, y_m) = \frac{\sum_{i=1}^{N}(y(i) - \overline{y})(y_m(i) - \overline{y_m})}{\sqrt{\sum_{i=1}^{N}(y(i) - \overline{y})^2 \cdot \sum_{i=1}^{N}(y_m(i) - \overline{y_m})^2}} \tag{7.16}$$

In Eqs. (7.15) and (7.16), y is the target value of test sample, y_m is model output value. N is number of data samples. $\overline{y}$ is test sample average, $\overline{y_m}$ is the average model output. The value R is more close to 1, model has higher prediction precision.

7.4 Results Analysis

Through investigation and research, the data of a metro station from January to September, will be preprocessed according to the formula in Table 7.1. The first 243 groups of data are used for training and after 30 groups of data are used for testing.

The SVR model prediction accuracy depends on the selection of kernel function and the related parameters. This paper uses the RBF kernel function and uses the root–mean-square error as shown in formula (7.15) model fitness function to find the best parameters C, ε, σ.

First GA training initial parameters should be set. The maximum evolution algebra is 100. The largest population number is 20. Roulette method is used to select the outstanding individuals. Arithmetic crossover should be used, crossover probability is 0.4, and the mutation probability is 0.01. Parameters C, ε, σ can have the values of [0.1, 1000], [0, 1], [0.001, 100]. Through genetic algorithm's optimal parameters, the best parameter for C is 11.3618, the best parameter for ε is 0.1022, and the best parameter for σ is 0.0020. The security state fitness training process curve is shown in Fig. 7.5.

Model output and the expected results are shown in Fig. 7.6, the figure shows that GA-SVR model for safety state prediction results track precision and small error. Test results can well track the desired output, the safety state of training RMSE is 0.0522, and testing RMSE is 0.0777. So it is obvious that error is smaller. Figure 7.7 shows the training and testing of the proceeds of the GA-SVR output and desired output correlation analysis diagram, state security training correlation coefficient is 0.9460, tests the correlation coefficient as 0.9306, correlation coefficients are all greater than 0.8. Thus, GA-SVR model output and the expected output correlation is higher, it shows that GA-SVR forecasting model can accurately predict the metro station safety status.

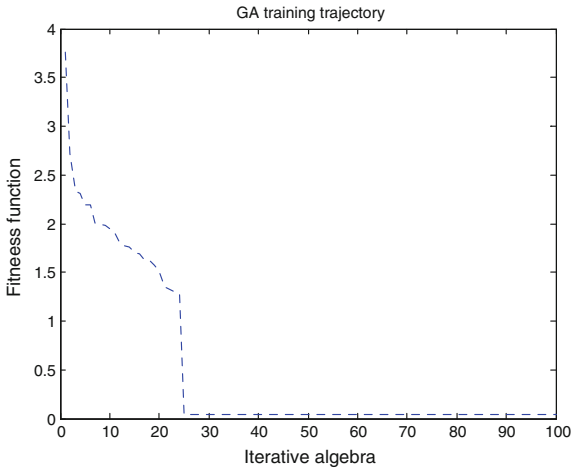


Fig. 7.5 The fitness curve of safety status

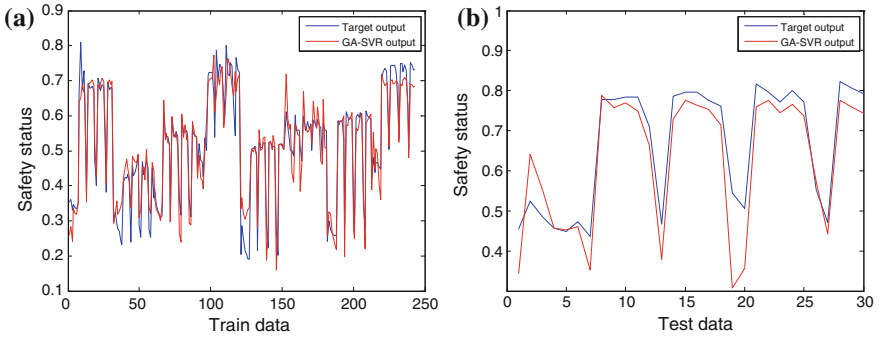


Fig. 7.6 Comparison chart of GA-SVM model outputs and expected results. **a** Training data (safety status). **b** Testing data (safety status)

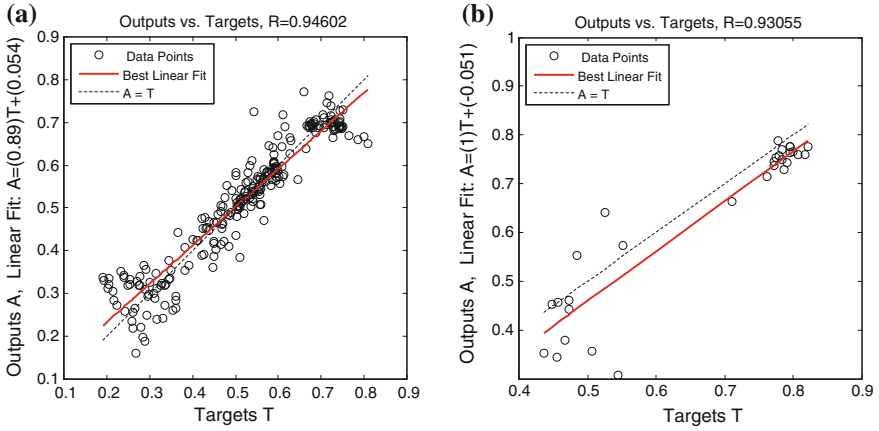


Fig. 7.7 Correlation coefficient of GA-SVM model outputs and expected results. **a** Training data (safety status). **b** Testing data (safety status)

7.5 Conclusions

The paper takes one metro station as an example, analyzes the influence factors of safety status for the metro station, extracts characteristic variables of metro station safety status, and establishes the metro station safety status prediction model based on GA-SVR. Through analyzing the example, the results show that the algorithm can accurately predict the trend of metro station safety status. Moreover, it has a good practical application value and can prevent the accident, even to ensure the safety of the station operation. The change from passive safety to active safety has very important significance.

Acknowledgments The authors would like to express our thanks to the editor and anonymous reviewers for their help in revising the manuscript. This research is sponsored by national natural science foundation of China (No. 61374157). The support is gratefully acknowledged.

References

1. Zhang Y et al (2013) Roller bearing safety region estimation and state identification based on LMD–PCA–LSSVM. Measurement 46(3): 1315–1324
2. Xi RR, Yun XC et al (2012) Research survey of network security situation awareness. J Comput Appl 32(1):1–4
3. Gao K, Jianming L et al (2011) A hybrid security situation prediction model for information network based on support vector machine and particle swarm optimization. Power Syst Technol 35(4):176–182
4. Endsley MR, Sollenberger R, Stein E (2000) Situation awareness: a comparison of measures. In: Proceedings of the human performance, situation awareness and automation: user-centered design for the new millennium
5. Qin Y, Zhang Z, Chen B, Xing Z, Liu J, Li J (2013) Research on the prediction model for the security situation of metro station based on PSO/SVM. J Intell Learn Syst Appl 5(04):237
6. Qin Y, Zhang Z, Liu X, Li M et al. (2015) Dynamic risk assessment of metro station with interval type-2 fuzzy set and TOPSIS method. J Intell Fuzzy Syst 1(29):93–106
7. Wang D, Wang X, Liu T et al (2012) Prediction of total viable counts on chilled pork using an electronic nose combined with support vector machine. Meat Sci 90(2):373–377
8. Scholkopf B, Smola A (1998) A tutorial on support vector regression. Technical Report Series NC2-TR-1998-030
9. Ran B, Boyce D (2012) Dynamic urban transportation network models: theory and implications for intelligent vehicle-highway system. Springer Science & Business Media
10. Ha M, Wang C, Chen J (2013) The support vector machine based on intuitionist fuzzy number and kernel function. Soft Comput 17(4):635–641

Chapter 8
Reliability Analysis of Metro Vehicles Bogie System Based on Fuzzy Fault Tree

Lei Zhu, Zhenyu Zhang, Yong Qin, Xiaoqing Cheng, Linlin Kou, Minzheng Yuan and Guangwu Liu

Abstract Metro vehicles bogie system is one of the important components of urban rail train. Its performance largely determines the vehicle running stability and safety. Accurately analyzing the reliability of urban rail train bogie system for urban rail train running stability and security is of great significance. This paper establishes fault tree model through the study found that all of the bottom events occur will cause bogie system failure. This paper is based on the analysis of the fuzzy importance degree according to the reliability analysis. The results show that the air spring breakages, brake failure, wheel tread strip, and flange abrasion of bogie had a greater influence on the performance of the system. This result is similar to actual operation.

Keywords Metro vehicles · Bogie system · Reliability analysis · Fuzzy fault tree

8.1 Introduction

As people living standard rises, and people for the level of public passenger transport service requirements rise, urban rail transit has the characteristics of energy saving, fast and large capacity which can arise more and more attention. Therefore, investment for the subway construction is very good. Developing rail transit as the main body of the urban conventional public traffic system will be an

L. Zhu · Z. Zhang · Y. Qin (✉) · X. Cheng · L. Kou
State Key Laboratory of Rail Traffic Control and Safety, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing, China
e-mail: qinyong2146@126.com

M. Yuan · G. Liu
Guangzhou Metro Corporation, Guangzhou, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_8

efficient way of solving urban traffic congestion. And in order to adapt to the growing orbit transportation capacity requirements and ensure safe and reliable running of the rail transit vehicles, the study is urgent.

The city railway train bogie system is one of the important components of urban rail train. It can be used to transfer various load, and use the stick to ensure to produce between the wheel traction. Its performance largely determines the vehicle running stability and safety. Therefore, how to accurately assess the reliability of urban rail train bogie system is of great significance for urban rail train running stability and security. Statistical model and mechanism model method can be used for research system reliability [1].

Statistical model method is mainly based on the analysis of system failure probability distribution function of time, and it is concluded that the reliability of the system related indexes, such as the average time between failures, failure rate, etc. Li et al. [2] apply the method of gray correlation degree and the fault tree to analyze the reliability of the bogie, and to find out the weak spot of freight car bogies. Gruden et al. [3] collected data based on wireless sensor network and uses the method of statistical system of Sweden railway bogie for reliability analysis. Lu et al. [4] analyzed the reliability of high-speed train bogie design optimization based on Monte Carlo simulation for railway bogie frame and the parameter sensitivity analysis.

Mechanism model method is mainly based on the structure and principles of system reliability analysis. Zhang [5] through the Beijing–Tianjin intercity train bogie frame for fatigue strength analysis, find out the key to structure fatigue location. Baek et al. [6] use the rain flow count method, and fitting of P-S-N curve, thus to predict the fatigue reliability of truck bogie, and error control in 2.7–3.1 %.

The research on the bogie system reliability analysis has made certain achievements, but there are still some problems. (1) At present, most of the researches are on railway passenger cars, vans, and less research on the reliability of the urban rail train bogie; (2) The method based on the monitoring and maintenance data and the failure probability of failure mode importance of fuzziness and uncertainty problems are lack of consideration.

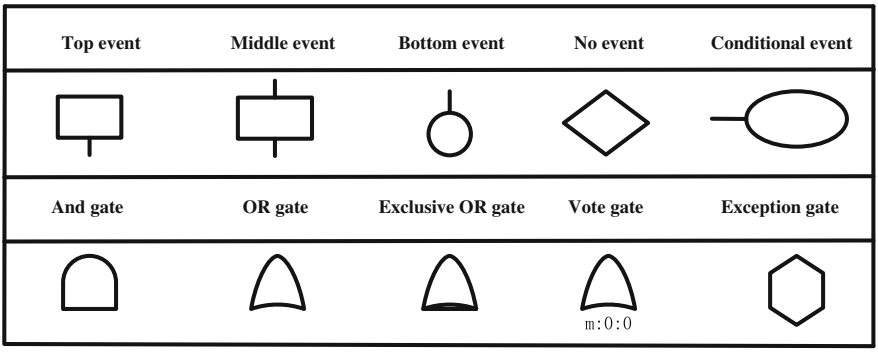
Therefore, in this paper, according to the data bogie maintenance the reliability of urban rail train bogie is analyzed by the fuzzy fault tree method.

8.2 Fault Tree Analyses

Fault tree analysis is through the establishment to determine system failure for the top events. Listed in Table 8.1 is a common fault tree event symbols and symbolic logic gate.

The fault tree analysis steps as shown in Fig. 8.1, generally can be divided into the following five stages [7]:

Table 8.1 Common event fault tree symbols and symbolic logic gate



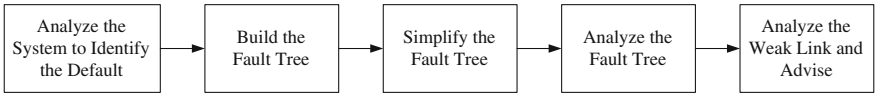


Fig. 8.1 The establishment of fault tree steps

8.3 Based on the Fuzzy Fault Tree of Metro Bogie System Reliability Analysis

8.3.1 The Fuzziness of Failure Analysis

System failure generally can be divided into affect security failure, the failure of influence the task and the need of maintenance types. The influence of the reliability of the system uncertainty generally has the following three:

1. Due to the causality uncertain events, finally forms the judgment result uncertainty, such as between the phenomenon and reason no definite corresponding relationship;
2. The extension of concept itself is not clearly, because of a malfunction of equipment "fault" and "no fault" there is no clear definition and evaluation standard. So it can make the fault judgment standard not clear;
3. Due to statistical data and the information is not enough, the judgment of the conclusion is uncertain. Boils down to in the complexity of objective things or incompleteness of information leads to people's subjective understanding is not clear.

Fuzzy reliability analysis is based on the deepening development of conventional reliability analysis is in system with fuzzy phenomenon when the useful supplement to conventional reliability analysis. Fuzzy reliability analysis is still a

hot research topic at present. Through the fuzzy set theory is introduced into the conventional reliability analysis, the results include the fuzzy reliability of series system and multiple objective fuzzy optimization decision problem, for the fuzzy fault tree analysis, and so on.

8.3.2 The Concept of Fuzzy Number

8.3.2.1 The Fuzzy Set and Membership Function

The concept of fuzzy set is first proposed in 1965 by American scholar L.A. Chad, which is a branch of mathematics, fuzzy mathematics. Different from the classic collection of "yes" or "no" two states, the subordination of one fuzzy set is usually represented by a membership function. Based on the fuzzy probability theory, mathematical statistics and the theory of fuzzy reliability research expand the processing system reliability theory of random and fuzzy phenomenon.

In the classic collection, $x \in A$ or $x \notin A$ represents the classic collection A's one set of ordered pairs $(x, 0)$ or $(x, 1)$. In the fuzzy set, closed interval $[0, 1]$ can represent the characteristic function $(x, 0)$ or $(x, 1)$.

Definition Set X is the collection of object x, x is any element of X. X on the fuzzy set is defined as A.

$$A = \{(x, \mu_A(x))|x \in X\} \tag{8.1}$$

In this equation, the membership function of the fuzzy set is $\mu_A(x)$. Through the membership function, each element of the set will be mapped to a value between 0 and 1.

8.3.2.2 $L-R$ Fuzzy Number

When event fault probability are fuzzy, usually with fuzzy number to represent the event, the event with a range of the failure probability distribution. and it usually consists of three parameters (m, α, β) which describe the left and right reference functions to describe fuzzy number. Set L and R are the reference function of fuzzy number, if the following equation is a true statement.

$$\mu_A(x) = \begin{cases} L((m-x)/\alpha), & x \leq m, \ \alpha > 0 \\ R((x-m)/\beta), & x \geq m, \ \beta > 0 \end{cases} \tag{8.2}$$

In Eq. (8.7), $L(x)$ and $R(x)$ are the monotone decreasing function between $(0, \infty)$. and $L(0) = R(0) = 1$, $\lim_{x\to\infty} L(x) = \lim_{x\to\infty} R(x) = 0.\mu_A(x)$ is a membership function of A. So the fuzzy number A is $L{-}R$ fuzzy number, and is recorded as $A = (m, \alpha, \beta)_{LR}$. When the values of α and β are smaller, the fuzzy number is more precise. When the value of α and β is zero, the fuzzy number is normal. When the values of α and β are bigger, the fuzzy number is more fuzzy.

Membership function is the most common triangular functions, normal type two kinds. Choose linear function, usually to a particular event is limited to a specific range, if can limit, the persons to be normal function. In this paper, the fault tree model of each event probability type membership function to choose normal function.

Normal type membership function of fuzzy number mathematical formula is as follows:

$$\mu_A(x) = \begin{cases} \exp(-((m-x)/\alpha)^2), & x < m, \ \alpha > 0 \\ 1, & x = m \\ \exp(-((m-x)/\beta)^2), & x > m, \ \beta < 0 \end{cases} \tag{8.3}$$

The normal fuzzy number $\lambda(0 \le \lambda \le 1)$ do cut set, the member function of the cut set A^λ can be obtained.

$$\mu_A^\lambda(x) = \begin{cases} \exp\left(-((m-x)/\alpha)^2\right), & x \in \left[m - \alpha\sqrt{-\ln\lambda}, m\right] \\ 1, & x = m \\ \exp\left(-((m-x)/\beta)^2\right), & x \in \left[m, m + \beta\sqrt{-\ln\lambda}\right] \\ 0, & \text{others} \end{cases} \tag{8.4}$$

So cut set interval of λ is $A_\lambda = [m - \alpha(-\ln\lambda)^{1/2}, m + \beta(-\ln\lambda)^{1/2}] = [L_A^\lambda, R_A^\lambda]$.

8.3.3 The Fuzzy Importance Degree Analysis Method

In the fault tree model, need to adopt the corresponding fuzzy importance degree analysis method, is used to assess the role of each of the system of minimum cut sets and effect size, and has improved according to the system according to the results of the analysis of the key parts.

In the process of the fault tree analysis of reliability, the failure probability of events is defined as a triangular fuzzy number (or fuzzy set and sets), conventional importance also correspondingly becomes fuzzy importance degree, use more important degree of fuzzy analysis method, etc. This paper mainly introduces Hideo Tanaka method of fuzzy importance degree.

Based on the reliability of the fault tree model, usually with triangular fuzzy number said the probability of occurrence of events, at this point, become fuzzy importance to the importance of the event. Analysis methods mainly include the median method of fuzzy importance [8], Hideo Tanaka method [9], etc., this paper adopts Hideo Tanaka method.

Assuming that it has the fault tree model, the probability and fuzzy number of top event T is as follows:

$$F_{\mathrm{T}} = F(q_1, q_2, \ldots, q_n) = (a_{\mathrm{T}}, b_{\mathrm{T}}, c_{\mathrm{T}}) \tag{8.5}$$

Defining the fault tree model of the influence degree of the end of each event as follows, F_{T_i} is top event. And F_{T} does not have the first i bottom event failure probability. The difference V_{T_i} of F_{T_i} and F_{T} is event x_i's fuzzy importance. When bottom T_i has the more influence to top event T, V_{T_i} will be bigger. Calculation formula is shown below.

$$V(F_{\mathrm{T}}, F_{\mathrm{T}_i}) = (a_{\mathrm{T}} - a_{\mathrm{T}_i}) + (b_{\mathrm{T}} - b_{\mathrm{T}_i}) + (c_{\mathrm{T}} - c_{\mathrm{T}_i}) \tag{8.6}$$

8.4 The Example Analysis

In bogie system failure as the research object, and connecting with the fault history and failure mode analysis, bogie system fault tree model is set up, as shown in Fig. 8.2. In Fig. 8.2, the bogie system is made up of each subsystem in series system, any subsystem faults in the system, can lead to the development of bogie system failure. The data collected from vehicle department of Guangzhou metro.

To solve the minimum cut sets, starting from the basic event, represented by the fault tree structure.

$$T = \bigcup_{i=1}^{26} x_i \tag{8.7}$$

So all the minimal cut set bogie system failure are $\{x_1\}, \{x_2\}, \ldots, \{x_{27}\}$. According to the minimum cut sets, all of the above first-order minimum cut sets, the corresponding bottom events will lead to bogie system failure (Table 8.2).

1. The traditional fault tree reliability index calculation

According to the historical records of citizen rail train somewhere, average probability can be concluded that the bottom of the bogie system event occurs, as shown in Table 8.3.

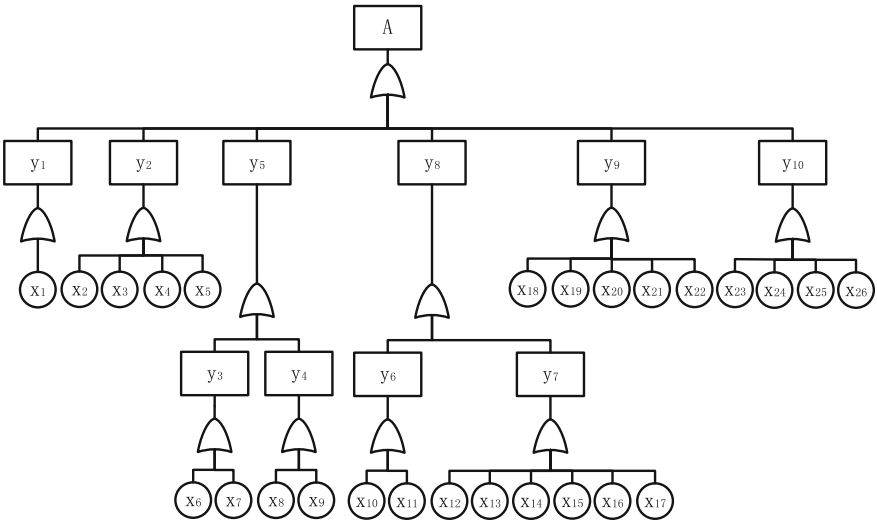


Fig. 8.2 Bogie system fault tree model

Table 8.2 Bogie system events in fault tree and the corresponding meaning

Event	Meaning	Event	Meaning
X1	Shaft end capping failure	X19	Traction motor oil
X2	Transverse shock absorber oil spill	X20	Traction converter failure
X3	Air spring breakage	X21	Blower different sound
X4	Resist sinusoidal vibration absorber function failure	X22	Auxiliary power failure
X5	The rubber section cracking	X23	The brake clamp leak
X6	Water retaining plate breakage	X24	Brake failure
X7	The bolt fracture	X25	Shaft plate clamp brake piece burst
X8	Wounded axle box boom node	X26	Shaft disk brake clamp gap is too small
X9	Lateral rubber node breakage	y1	Frame fault
X10	The wheel tread of stripping	y2	The spring mechanism failure
X11	Rim abrasion	y3	Coupling fault
X12	The bottom wounded	y4	Traction rod failure
X13	Before the axle box cover crack	y5	Connection failure
X14	Shaft bruise	y6	Wheel of the fault
X15	Outside the windshield rubber tearing holes	y7	Axle box fault
X16	The deformation of axle box body end cover	y8	Wheel axle box of failure
X17	The hollow axle with sundry	y9	Driving mechanism of failure
X18	Crack of dust cover plate	y10	The foundation brake failure

Table 8.3 Bogie system event table probability

Bottom event	Failure rate/10^{-3}	Bottom event	Failure rate/10^{-3}	Bottom event	Failure rate/10^{-3}	Bottom event	Failure rate/10^{-3}
x1	0.6471	x8	1.0732	x15	2.0134	x21	1.6037
x2	0.5103	x9	0.8963	x16	1.6037	x22	0.2368
x3	0.1253	x10	0.1356	x17	0.7543	x23	1.4671
x4	1.3304	x11	0.1634	x18	1.0735	x24	0.1354
x5	0.3368	x12	0.2413	x19	0.9315	x25	2.8323
x6	1.0571	x13	0.6471	x20	0.2843	x26	1.1938
x7	0.2052	x14	0.2704				

In Table 8.3, bottom event failure rate is the unit of time/million miles of bogie system for failure rate.

$$\lambda = \bigcup_{i=1}^{26} \lambda_i = 0.0206 \tag{8.8}$$

Therefore, bogie system reliability is 0.9794, namely the bogie running to 1 million kilometers, the reliability of 0.9794.

2. Based on the fuzzy importance degree analysis of reliability index is calculated.

System reliability analysis method of fault tree, usually to a certain value to represent the failure probability of bottom events, and in engineering application, the probability of bottom events is often fuzzy uncertainty. Moreover, based on the reliability of the fault tree model, the bogie system is affected by various events, cannot treat as the same, so, the method based on fuzzy importance in theory is more suitable for the analysis of the bogie system reliability.

Failure the bottom events of the influence degree of the bogie system is also different, so, this article mainly uses the method based on fuzzy importance degree analysis the reliability of the bogie system.

Fuzzy probability interval of each bottom event indeed is concrete as shown in Table 8.4.

In Table 8.4, the first column is the bottom event, the second column is the average failure rate of the basic events, the unit of time/millions of kilometers, the third and the fourth row failure probability and fuzzy number L and R, respectively. For example, 0.6471 is happening probability, using the normal fuzzy number is expressed as $f_T = (0.6469 \times 10^{-2}, 0.6471 \times 10^{-2}, 0.6473 \times 10^{-2})$. That is to say, when confidence $\lambda = 0.95$, x_1 probability is between 0.6469×10^{-2} and 0.6473×10^{-2}, the probability 0.6471×10^{-2} is the most likely.

In this paper, the calculation Hideo Tanaka fuzzy importance degree method to calculate the fuzzy importance, according to the formula (12), modeling of the fault tree analysis of important degree of minimum cut sets of fuzzy analysis, get the influence extent of each event fault as shown in Table 8.5. Determine which of

Table 8.4 Bogie system event table and probability and fuzzy number

Bottom event	Failure rate/10^{-3}	Probability and fuzzy number L	Probability and fuzzy number R	Bottom event	Failure rate/10^{-3}	Probability and fuzzy number L	Probability and fuzzy number R
x1	0.6471	0.6469	0.6473	x14	0.2704	0.2702	0.2706
x2	0.5103	0.5101	0.5105	x15	2.0134	2.0132	2.0136
x3	0.1253	0.1251	0.1255	x16	1.6037	1.6035	1.6039
x4	1.3304	1.3302	1.3306	x17	0.7543	0.7541	0.7545
x5	0.3368	0.3366	0.3370	x18	1.0735	1.0733	1.0737
x6	1.0571	1.0569	1.0573	x19	0.9315	0.9313	0.9317
x7	0.2052	0.2050	0.2054	x20	0.2843	0.2841	0.2845
x8	1.0732	1.0730	1.0734	x21	1.6037	1.6035	1.6039
x9	0.8963	0.8961	0.8965	x22	0.2368	0.2366	0.2370
x10	0.1356	0.1354	0.1358	x23	1.4671	1.4669	1.4673
x11	0.1634	0.1632	0.1636	x24	0.1354	0.1352	0.1356
x12	0.2413	0.2411	0.2415	x25	2.8323	2.8321	2.8325
x13	0.6471	0.6469	0.6473	x26	1.1938	1.1936	1.1940

bogie had a greater influence on the performance of the system are air spring breakage, brake failure, wheel tread stripping, rim abrade, conform to the results, and the actual engineering application situation.

In the national standard draft, according to different levels, various components have different influence degree. Probability is 0, said no effect; Damage may occur rarely in 0.1; Damage may occur in 0.5; 1 said that damage. According to the fault history and experience of field workers, sorting for 1–8, function of loss probability to 1; Sort of 9–17, loss probability function of 0.5; Sorting for 18–26, loss probability function of 0.1.

Table 8.5 Bogie system to minimum cut set of events and sorting

Bottom event	Important degree	Sorting	Bottom event	Important degree	Sorting
x1	0.0599	13	x14	0.0610	8
x2	0.0603	11	x15	0.0558	25
x3	0.0615	1	x16	0.0570	23
x4	0.0578	21	x17	0.0596	14
x5	0.0608	10	x18	0.0586	19
x6	0.0587	17	x19	0.0590	16
x7	0.0612	5	x20	0.0610	9
x8	0.0586	18	x21	0.0570	24
x9	0.0591	15	x22	0.0611	6
x10	0.0614	3	x23	0.0574	22
x11	0.0613	4	x24	0.0614	2
x12	0.0611	7	x25	0.0533	26
x13	0.0599	12	x26	0.0583	20

$$\lambda_{57} = \sum \omega_i \lambda_i = 0.005965 \tag{8.9}$$

Bogie system with the normal fuzzy number is expressed as failure rate $f_T = (0.5963 \times 10^{-2}, 0.5965 \times 10^{-2}, 0.5967 \times 10^{-2})$, so the reliability of the top event is located in between $[0.99403, 0.99404]$, the credibility of 0.95. Traditional fault tree of the bogie system reliability is 0.9808, the results without taking the fuzziness possessed by probability of bottom events, also did not give full consideration to the importance degree of each event and therefore has a certain discrepancy with actual situation. Based on experience and expertise in the field, based on the fuzzy importance of bogie system reliability more accord with the engineering practice.

8.5 Conclusions

Based on a subway bogie system as the research object, and combining the fault history and failure mode analysis, a fault tree model of the bogie system, through the study of all find out bogie system of minimum cut sets, and found that all of the bottom events occur will cause bogie system failure. In reliability analysis, this article adopts the method of based on fuzzy importance degree analysis for reliability calculation. Citizen somewhere rail train bogie survey data analysis, the results show that the air spring breakage, brake failure, wheel tread strip, flange abrasion of bogie had a greater influence on the performance of the system, conform to the results and the actual operation situation. Based on fault tree analysis in reliability, considering the fuzzy uncertainty of fault probability and influence degree of bogie system, event failure probability with fuzzy numbers, said and fuzzy reliability of the results. Based on experience and expertise in the field, based on the fuzzy importance of bogie system reliability analysis result is closer to the engineering practice.

Acknowledgments The authors would like to express our thanks to the editor and anonymous reviewers for their help in revising the manuscript. This research is sponsored by the Research Fund for the Doctoral Program of Higher Education of China under grand 20120009110035 and the State Key Laboratory Program under Grant RCS2014ZT24. The support is gratefully acknowledged.

References

1. Qin Y, Zhang ZY, Shi JX (2014) Reliability analysis and prediction of metro vehicles' bogie frame. Appl Mech Mater 590
2. Li Y, Liu HT, Gao Y et al. (2012) Fuzzy grey fault tree analysis on wagon bogie reliability. In: Proceedings of the 1st international workshop on high-speed and intercity railways. Springer, Berlin, Heidelberg, pp 117–123

3. Grudén M, Westman A, Platbardis J et al. (2009) Reliability experiments for wireless sensor networks in train environment. In: Wireless technology conference, 2009. EuWIT 2009, European. IEEE, pp 37–40
4. Lu Y, Zeng J, Wu P et al (2010) Reliability and parametric sensitivity analysis of railway vehicle bogie frame based on monte-carlo numerical simulation. High Performance Computing and Applications. Springer, Berlin, Heidelberg, pp 280–287
5. Zhang SG (2008) Study on testing and establishment method for the load spectrum of bogie frame for high-speed trains. Sci China Ser E: Technol Sci 51(12):2142–2151
6. Baek SH, Cho SS, Joo WS (2008) Fatigue life prediction based on the rain flow cycle counting method for the end beam of a freight car bogie. Int J Automot Technol 9(1):95–101
7. Kumar M, Yadav SP (2012) The weakest t-norm based intuitionist fuzzy fault-tree analysis to evaluate system reliability. ISA Trans 51(4):531–538
8. Wang J, Nie R, Zhang H et al (2013) New operators on triangular intuitionist fuzzy numbers and their applications in system fault analysis. Inf Sci 251:79–95
9. Tanaka H, Fan LT, Lai FS, Toguchi K (1983) Fault-tree analysis by fuzzy probability. IEEE Trans Reliab 32(5):453–457

Chapter 9
Analyzing and Predicting Forwarding Behaviors in Sina Weibo

Yazhou Hu, Bo Deng, Bing Wang and Bin Hong

Abstract With the microblog becoming more and more popular, people spend much time on the microblogging site, such as Sina Weibo. Users can push a new microblog by themselves or forward a microblog which is interesting and valuable from their followers and followees, and forwarding behaviors have important influences in information spread. Studying forwarding behaviors enable researchers to understand which kind of information will be propagated in social network. In this paper, we experiment with a corpus of Sina Weibo messages and investigate five important features to find how these features affect forwarding behaviors. Then, we train a prediction model to forecast whether a microblog will be forwarded by a given user. After getting the parameters learned from the model, our model can achieve a precision and recall of 20.13 and 26.27 % on average.

Keywords Forwarding behavior · Microblog · Model · Analyzing · Predicting

9.1 Introduction

With the development of mobile Internet, online social network [1] plays a more and more important role in people's life. People want to communicate conveniently with friends, express their feelings, and record their daily life in social networks, so microblogging services becoming are more and more popular, such as Twitter and Sina Weibo. For example, Sina Weibo has more than 530 million registered users which send one hundred million microblogs every day. But the number of original tweets is only little, and most of microblogs are forwarding from others.

Forwarding is a very common and important behavior in Sina Weibo. If a user finds an interesting microblog, he can forward it by using "//@". When his followers also like this microblog, they can also easily forward it. In this way, a message will quickly spread.

Y. Hu (✉) · B. Deng · B. Wang · B. Hong
Beijing Institute of System Engineering, Beijing, China
e-mail: hyz1011088@gmail.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_9

The reason why a microblog is forwarded by a given user has frequently been addressed, but mainly in a scenario of forwarding prediction for a given user by analyzing the structure of the social network. While only focus on the structure of the social network, we cannot find a given user's interest and personality. Therefore, in this paper, we analyze some features which affect user's forwarding behaviors and train a model based on user's forwarding history to forecast whether a microblog will be forwarded by a given user.

For this purpose, we analyze the user's forwarding history content to find the user's interest, followees, close friends, the frequency, number of microblogs, etc. We collect 568,425 tweets from 2380 users through the Sina API. The analysis show that big V users (people who have a large group of active followers and important influence) are different from common users in forwarding behaviors, so we divide all users into two parts to find different features related with forwarding behaviors. Then, we train a model and get the parameters learned by the model to predict forwarding behaviors.

The main contributions of our work can be summarized as follows:

- We train the model based on user's forwarding history. This model analyzes not only the structure of the social network, but also can analyze the user's interest, personality, and habit which are important features for predicting.
- We use natural language processing methods to analyze the correlation between two microblog. This method based on a sentence segmentation processing analyzes the content of microblog in Sina Weibi.
- We divide users into two parts: big V users and common users. The result shows big V users who a given user follows have important impact on forwarding behaviors of the user.

This paper is organized as follows. First of all, we introduce other researchers' works. Second, we analyze some important features to find how these features affect forwarding behaviors. Third, we introduce a model to predicate the user's retweeting behaviors. Fourth, we do experiments to analyze the performance of the model. At last, we make conclusions to summarize our work.

9.2 Related Work

Sina Weibo has attracted much research in recent years. As forwarding behaviors represent the characteristics of users, many researchers want to find the reason why a given user will forward a microblog.

Boyd et al. [2] analyze the way of retweeting, the purpose of retweeting behaviors, and information items of a tweet. Suh et al. [3] select several contextual features, such as URLs, hashtags, and mentions, to analyze which features will affect retweeting behaviors using PCA. Webberley et al. [4] pay attention to the

dynamics and patterns of retweets, and analyze some features which affected retweeting behaviors, such as length of retweeting chains, size of retweeting groups, and retweet time delay. However, this research is just a statistical analysis of retweeting behaviors, which lacks of prediction the users' retweeting behaviors.

Zaman et al. [5] train Matchbox models to predicate the probability of retweeting behaviors based on analyzing username, followers, words of the microblog, etc. Xie et al. [6] proposed two algorithms to predicate user's forwarding behaviors. The first predication algorithm is based on user's microblogs content, and another one is based on user's characterization. Cao et al. [7] analyzed some possible features whether affected forwarding behaviors and put forward three comprehensive features to predicate forwarding behaviors.

9.3 Features Analysis

In this section, we provide an overview of different features. Then we analyze those features to estimate the relevance of a feature to forwarding behaviors.

9.3.1 Forwarding Time

Forwarding time is an important feature we should focus on. If a microblog is pushed at a time when the given user do not on line, this user may ignore it instead of forwarding it.

The rate of forwarding microblogs is different in different time periods of a day, p_t is the rate of forwarding microblogs when time is t. We divide 24 h of a day into 12 parts, and get p_t by counting the number of microblogs in each period of time and the number of all microblogs of a user. n_{ti} is the number of microblogs in ith period of time, and n_{all} is the number of all microblogs. p_t is defined as follows.

$$p_t = \frac{n_{ti}}{n_{\text{all}}} \tag{9.1}$$

We find different users have obvious difference on forwarding time. But most of the users are fit to the performance as shown in Fig. 9.1.

Figure 9.1 shows the rate of microblogs in different time periods is different. The high rate often appears at noon and nightfall when users get off works or rest. So we can statistically study the rate of forwarding microblogs at different time periods of a user to evaluate the probability of forwarding microblogs.

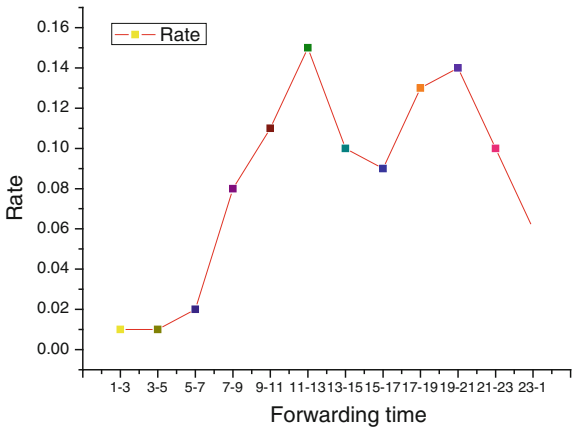


Fig. 9.1 The rate of forwarding microblogs in different time periods

9.3.2 Forwarding Frequency

The frequency of forwarding microblogs is also an important aspect, we should focus on. p_f the value of frequency. n_t is the number of microblogs that the user forwards. p_f is defined as follows.

$$p_f = \frac{n_t}{n_{\text{all}}} \tag{9.2}$$

From the definition of p_f above, we can get the frequency of forwarding a microblog of a user. But different users have different results, so we analyze all forwarding microblogs and the result is in Fig. 9.2.

Figure 9.2 shows the rate of forwarding microblogs in all microblogs of a given user is very high, and it is the reason why we should focus on forwarding

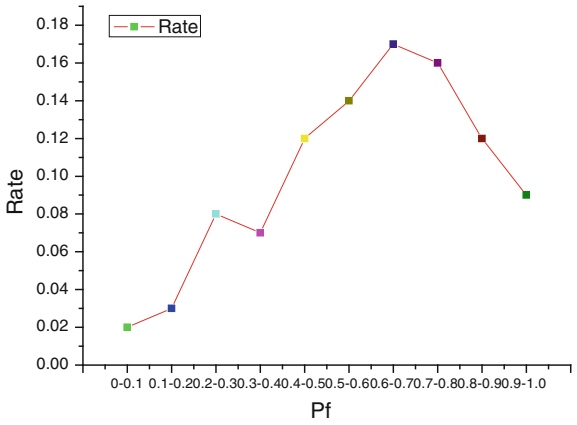


Fig. 9.2 The rate of forwarding microblogs in different p_f

microblogs. The microblogs of most of users are forwarding from others. After analyzing users' forwarding history, we find about 59 % users forward more than half microblogs in their all microblogs.

9.3.3 *Forwarding from Followees*

In Sect. 9.3.2, we find most of the users' microblogs is forwarded from others, but forwarding from which followee is also an interesting feature. p_{f_i} is the rate of forwarding a microblog from the ith user. n_{fti} is the number of microblogs from the ith user. p_{f_i} is defined as follows.

$$p_{f\_i} = \frac{n_{tfi}}{n_{\text{all}}} \tag{9.3}$$

We find common users and big V users are different in the rate of forwarding microblogs from common users and big V users, and we analyze it in Fig. 9.3.

Figure 9.3 shows big V users often forward microblogs from other big V users, and the rate arrives 98 % when the number of followers arrives 10 million. Common users also often forward microblogs from big V users, but the rate of forwarding from common users is about 21.7 % which is significantly higher than the big V users' rate.

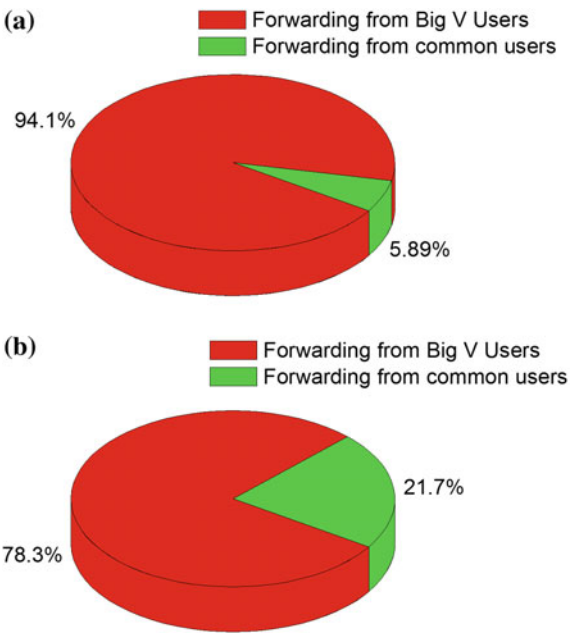


Fig. 9.3 The rate of forwarding microblogs from (**a**) big V users and (**b**)common users

9.3.4 Forwarding Chains Length

Forwarding chain length is the number of forwarding steps from the original microblog to current microblog. If forwarding chains is very long, we may not forward it sometimes. So, we focus on it and define it as l_{rc}. Then, we define p_{lrc} which is the rate of forwarding a microblog which forwarding chains length is l_{rc}.

We find the length of forwarding chains will affect forwarding behaviors, and analyze it in Fig. 9.4.

Figure 9.4 shows users often forwarding a microblog when the length of forwarding chains is below 3. The rate of users forwarding a microblog from the original author is about 33.73 %, and the rate is 36.32 % of forwarding from the original author's followers.

9.3.5 Microblogs Contents

The microblog content is very important for us to predicate users' forwarding behaviors. We use VSM model [8] to analyze the relationship of two microblogs. The content of the microblog that we should predicate is defined as c_{p}, and the content of history microblogs is defined as c_{h}. We use ICTCLAS to segment the content of c_{p} and c_{h}. Then, we get the dictionary, $d = \{w_i, i = 1, 2, \ldots, n\}$, and w_i is keyword of the content of c_{p} and c_{h}, and n is the number of all keywords. We calculate the $tf - idf$ [9] value of each word, and it is defined as t_i.

$$t_i = tf - idf = \frac{n_i}{\sum_1^k n_k} \times \log \frac{\left|c_{\mathrm{p}} \cup c_{\mathrm{h}}\right|}{\left|\left\{c : w_i \in c_{\mathrm{p}} \cup c_{\mathrm{h}}\right\}\right|} \tag{9.4}$$

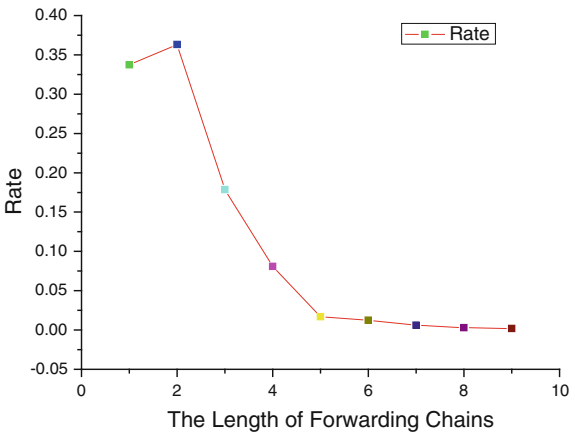


Fig. 9.4 The rate of different forwarding chains length

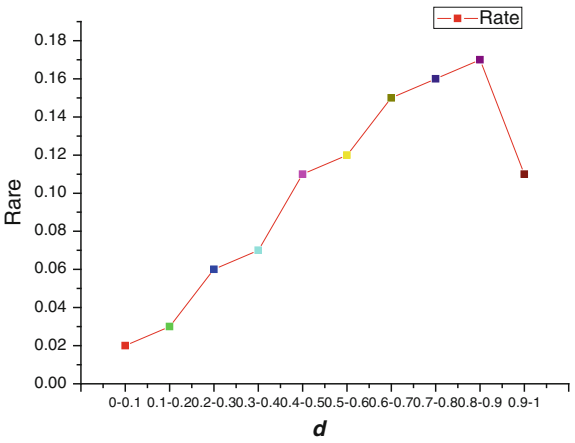


Fig. 9.5 The rate of forwarding microblogs of the different distance of two microblogs

We get two vectors, $T_c = \{t_i, i = 1, 2, \ldots, n\}$, $T_h = \{t'_i, i = 1, 2, \ldots, n\}$. Then we calculate the distance of two vectors, and define it as d.

$$d = \frac{T_c \cdot T_h}{T_c \times T_h} \tag{9.5}$$

If the distance of two microblogs is small, such as more than 0.30, then we think two microblogs are similar. So the distance of two microblogs is important feature we should focus on. Then, we statistically study how the distance of two microblogs affects forwarding behaviors in Fig. 9.5.

With the increasing of d, the rate of forwarding microblogs will increasing. The rate reaches 17.0 % when the value of d is between 0.8 and 0.9. But when the value of d is between 0.9 and 1.0, the rate decrease. The reason may be that users have forwarded similar microblogs already.

9.4 Modeling and Experiments

In this section, we build a prediction model which is based on users' forwarding history. This model gets the parameter value of each feature by training the model. Then we set the data into different datasets and define the evaluation methodology. At last, we conduct experiments using those parameters values on the dataset for forwarding prediction of a given user.

9.4.1 Definition of Model

In this section, we define a model using the basic features which is defined in the Sect. 9.3. For a microblog we have to analyze P is the rate of forwarding it by a given user, such as i. It is defined as follows.

$$P = \frac{1}{Z_k} \times p_t \times p_f \times \mathrm{e}^{(t_i + p_{fi} - t_i \times p_{fi} + \alpha \times p_{lrc})} \tag{9.6}$$

where Z_k is the normalization factor, α is a factor that can get by train the data.

9.4.2 Dataset and Evaluation Methodology

In this section, in order to predicate forwarding behaviors, we divide dataset into training set and testing set. We select 5 % of microblogs of a user as this user's testing set. All users' testing set combine testing set and other microblogs combine training set.

We can get the parameter value by training set, and we can use the model to predicate forwarding behaviors. Then we use testing set to evaluate the performance of the model.

We use R_p and R_r to quantify the accuracy of the prediction model. R_p (also called Precision Rate) measures the number of microblogs of correctly predicted in all microblogs of predicted by the model. R_r (also called Recall Rate) measures the number of microblogs correctly predicted in all microblogs forwarded by the given user. We also calculate F-measure to obtain a harmonic mean of prediction and recall. N_tp, N_fn, N_fp and N_tn denote the true positive number, false negative number, false positive number and true negative number. We define R_p, R_r and F-measure in a standard way as follows.

$$R_\mathrm{p} = \frac{N_\mathrm{tp}}{N_\mathrm{tp} + N_\mathrm{fp}} \tag{9.7}$$

$$R_\mathrm{r} = \frac{N_\mathrm{tp}}{N_\mathrm{tp} + N_\mathrm{tn}} \tag{9.8}$$

$$F\mathrm{-measure} = \frac{2 \times R_\mathrm{p} \times R_\mathrm{r}}{R_\mathrm{p} + R_\mathrm{r}} \tag{9.9}$$

Table 9.1 Results of prediction

	Precision rate	Recall rate	*F*-measure
Average	0.2013	0.2627	0.2279
Max	0.8625	0.9276	0.8939
Min	0	0	–

9.4.3 Results of Prediction and Discussion

Using the model proposed in Sect. 9.2, the result is showed in Table 9.1.

In Table 9.1, the maximum and minimum prediction performances of this model are 0.8625 and 0, respectively, and the performance is 0.2013 on average. In other words, if a given user finds a microblog, he forwards it with precisions rate of 0.2013 on average.

Then, we analyze the reason why the accuracy of prediction is not high. If a user has fixed interests and followees, the accuracy of prediction will be high. While, if a user has unstable and variable interests and followees, the accuracy of prediction will be low. In our dataset, some users forward a little microblog even the number is zero. This leads to the prediction accuracy for some users is 0. Meanwhile, we find the calculation of the distance between two microblogs' contents is not accurate, because microblogs often have ambiguous words in Chinese. Sometimes, microblogs just contain music, pictures, and video, and not text messages, and we cannot extract keywords to calculate the distance between two microblogs' contents.

But our model is efficient to predicate the forwarding behavior by a given user on a given microblogs. Meanwhile, our model can reach a high precision of 0.8625 and recall of 0.9276 for a special user, which is more valuable and attractive than other model [6, 7].

9.5 Conclusions

Forwarding is a very common and important behavior in Sina Weibo, which has important influences in information spread. We believe studying forwarding behaviors enable researchers to understand which kind of information will be propagated in social network.

In this paper, we analyze forwarding behaviors on Sina Weibo, and find some valuable phenomena, and we analyze some features to find which feature affects forwarding behaviors. Then we build a model to predict forwarding behaviors by a given user, based on statistical analysis and natural language processing methods. The result of prediction model on the dataset shows that the model can achieve a precision of 20.13 % and recall of 22.79 % on average.

References

1. Ellison NB (2007) Social network sites: definition, history, and scholarship. J Comput Mediated Commun 13(1):210–230
2. Boyd D, Golder S, Lotan G (2010) Microblog, microblog, forward: conversational aspects of forwarding on twitter. In: System sciences (HICSS), 2010 43rd Hawaii international conference on. IEEE, pp 1–10
3. Suh B, Hong L, Pirolli P et al. (2010) Want to be forwarded? large scale analytics on factors impacting forward in twitter network. In: IEEE second international conference on social computing (socialcom). IEEE, pp 177–184
4. Webberley W, Allen S, Whitaker R (2011) Forwarding: a study of message-forwarding in twitter. In: workshop on mobile and online social networks (MOSN). IEEE, pp 13–18
5. Zaman TR, Herbrich R, Van Gael J et al. (2010) Predicting information spreading in twitter. In: Workshop on computational social science and the wisdom of crowds, nips, vol 104(45), pp 17599–17601
6. Xie J, Liu G, Su B et al (2013) Prediction of user's forward behavior in social network. J Shanghai Jiaotong Uni 4:014 (in Chinese)
7. Jiu-Xin C, Jiang-Lin W, Wei S et al. (2014) Sina Microblog information diffusion analysis and prediction. Chin J Comput (in Chinese)
8. Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18(10):613–620
9. TF-IDF, available: http://zh.wikipedia.org/wiki/TF-IDF

Chapter 10
Modeling and Implementation for Dynamic Curves of Onboard Automatic Train Protection System Based on SCADE

Qingyun Wu, Teng Pan, Xiaoping Xue, Fang Zhang and Jianhua Jia

Abstract Software of railway system has high safety requirement. Formal method based on SCADE (safety-critical application development environment) was applied to the software development of onboard automatic train protection system to guarantee safety of the software. We built up a mathematical model of dynamic curves and transformed it into SCADE model in the form of dataflow and safe state machine. Simulation test, coverage test, and formal verification were performed. The results of the test and verification prove the model safe. Compared to the traditional software development method of onboard automatic train protection system, formal development method based on SCADE can improve reliability of software effectively.

Keywords Onboard ATP · Dynamic curves · SCADE · Formal method

10.1 Introduction

Automatic train protection (ATP) measures speed and position of the train, and computes the movement authority, which is a key equipment of CBTC. It is a key task for onboard ATP to compute dynamic curves to guarantee the safe intervals between trains. If there are errors in dynamic curves, serious accidents may happen, such as traffic collision. The software in onboard ATP has a high safety requirement. This is a challenge in onboard ATP software development.

Q. Wu · X. Xue (✉) · F. Zhang · J. Jia
Department of Information and Communication Engineering,
Tongji University, No. 4800 Cao'an Road, Shanghai, China
e-mail: xuexp@tongji.edu.cn

T. Pan
School of Michigan, Shanghai Jiaotong University,
No.800 Dongchuan Road, Shanghai, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_10

Traditional software development method of onboard ATP is coding manually, which produces bugs easily, and cannot satisfy safety requirement of ATP. Formal method, such as VDM, B method and Z, is increasingly used in software development and can guarantee correctness of software [1–4]. The formal method based on developing tool SCADE (safety-critical application development environment) is applied in several safety-critical applications, such as aerospace system [5] and railway systems [6–11]. In [6–11], SCADE is used in modeling train behaviors, failure handling function, and interlocking system, but the dynamic curves of ATP are not modeled and implemented by formal method yet.

In this paper, we build a mathematical model to calculate dynamic curves. Formal method based on SCADE is used to satisfy high safety requirement of software. Dynamic curve model was established by dataflow and safe state machine in SCADE. We performed a variety of verification including simulation test, coverage test, and formal verification.

The rest of this paper is structured as follows. Section 10.2 introduces the definition and mathematical model of the dynamic curve. Section 10.3 models the dynamic curve in SCADE by dataflow and safe state machine. Section 10.4 verifies the designed model by simulation, model test coverage, and formal verification. Section 10.5 concludes the paper.

10.2 Mathematical Model of Dynamic Curves

10.2.1 Calculation of Dynamic Curves

Usually, there are several trains moving on one track. The trailing train should guarantee that it can stop before the rear end of the leading train. The leading train defines a target point, i.e., the end of the leading train abstracting safety protection distance, and it changes with the position of the leading train. The onboard ATP of the trailing train calculates the speed limit according to the target point dynamically as showed in Fig. 10.1.

Dynamic curves include emergency brake deceleration (EBD) curve and emergency brake intervention (EBI) curve. EBD curve is calculated based on the assumption that full brake force can be built up immediately when emergency brake is triggered. But usually time is needed to build up full brake force. EBI is calculated

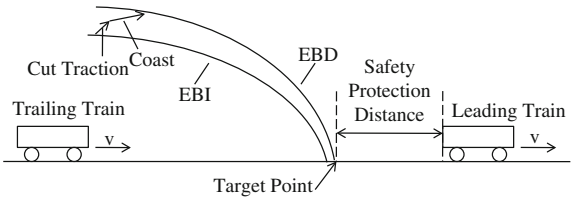


Fig. 10.1 Dynamic curves

based on EBD in consideration of the delay time. Emergency brake is activated if train speed is over EBI. Emergency brake can be released only when train stops.

The deceleration during emergency brake is determined by brake force and resistance. The deceleration due to brake force *A_brake* can be regarded as a safe constant, i.e., guaranteed emergency brake rate (GEBR). The deceleration due to gravity generates when the train is on ramp. For a particle in a ramp with gradient i, the deceleration due to gravity is

$$A\_grad(D) = i(D) * M\_train * A\_GRAVITY/(M\_train + M\_rotat) \quad (10.1)$$

A_grad	deceleration due to gravity (m/s^2),
M_train	mass of the train (kg),
M_rotat	nominal rotating mass (kg),
A_GRAVITY	acceleration of gravity (m/s^2),
i(*D*)	gradient, which changes with position of track *D*.

A train has a certain length, therefore track data should be revised by train length. The deceleration due to gravity is calculated using the lowest gradient value between the position of front end and rear end of the train.

Resistances can be ignored except for resistance due to grade, because other resistances have small values and the directions must be opposite to the train velocity. The deceleration after full brake force builds up is

$$A\_eb(D) = A\_brake + A\_grad(D) \quad (10.2)$$

Time is needed to build up full brake force after the onboard ATP commands the brake. The time delay can be divided into two segments. In the segment of cutting traction train moves with the initial traction acceleration. In the next segment, train coasts until full brake force is built up. EBI can be calculated from EBD in consideration of the delay before full brake force builds up.

10.2.2 Safety Properties

Safety properties are the properties that the model must satisfy. Three safety properties can be defined for speed supervision as follows.

Property 1. If the train overspeeds, it must in the emergency braking state;
Property 2. If the train is in the emergency braking state and not stops, the emergency brake cannot be released;
Property 3. If the train stops and the emergency brake is released, the train can move normally.

The properties are described in natural language, which has ambiguity. Therefore, the properties should be described by mathematical language. SCADE formal verification tool is based on the theory of Boolean satisfiability problem (SAT). The safety properties can be described as tautologies in the form of propositional logic or predicate logic.

Assume that *V_train* indicates the train speed, and *V_ebi* indicates the speed limit. *B_release* indicates whether the emergency brake is released or not, and *B_braking* indicates whether the train is braking or not. Then the safety properties can be described as follows.

Property 1. $(V\_train \geq V\_ebi) \wedge (V\_ebi > 0.0) \Rightarrow B\_braking$;
Property 2. $B\_braking \wedge (V\_train > 0.0) \wedge (V\_ebi > 0.0) \wedge B\_release \Rightarrow B\_braking$;
Property 3. $(V\_train = 0.0) \wedge B\_release \wedge (V\_ebi > 0.0) \Rightarrow \neg\, B\_braking$.

10.3 Formal Modeling of Dynamic Curves

10.3.1 Modeling Process

SCADE is a popular formal software development tool. The code generators of SCADE are certified according to various international safety standards, such as EN-50128 and DO-178B. In this paper, we use SCADE to model dynamic curves.

The process of SCADE Modeling is displayed in Fig. 10.2. Since the mathematical model has been built up, the input data and the output data can be determined. The input data and the output data can be defined formally as input interfaces and output interfaces in SCADE with explicit types. Then the top-level dataflow diagram can be determined. The logic relationship between input interfaces and output interfaces should be built up according to the mathematical model. It is the main task in SCADE modeling since the qualified C code can be generated automatically according to the SCADE model. The SCADE model should be tested and verified to guarantee the correctness. The simulation test can be used to

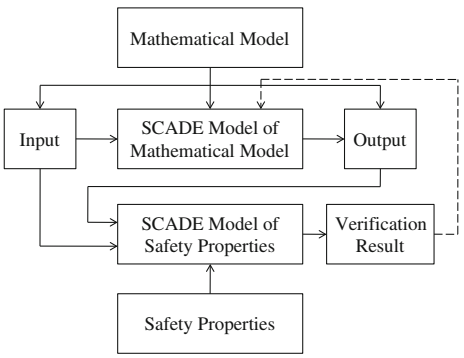


Fig. 10.2 Process of SCADE modeling

discover obvious errors, while the formal verification can prove the model correct strictly. Safety properties are determined according to the requirement of the software and should be transformed into SCADE model. The SCADE model of safety properties can be used to perform formal verification and the verification result can help developers to discover the errors in the SCADE model of the mathematical model.

10.3.2 Modeling of Dynamic Curves

SCADE provides two modeling method, i.e., dataflow and safe state machine (SSM). Dataflow is used to model continuous systems. Dataflow shows the process how data changes in the form of block diagram. The calculation modules of dynamic curves process continuous data and should be modeled by dataflow.

Input interfaces and output interfaces should be determined first to draw the top-level dataflow diagram. According to the mathematical model the input data include train speed, target point position, train position, train parameters, track data, and a signal indicating that emergency brake is released by hand, and the output data include a signal indicating whether emergency brake should be triggered. Train position, target point position, train parameters, and track parameters are used to compute speed limit. The onboard ATP determines whether emergency brake should be triggered by comparing train speed and speed limit and determines whether emergency brake should be released when train stops.

The top-level dataflow diagram can be built up as Fig. 10.3 shows. *D_target* is the target point position. *trainParameters* indicates the parameters of the train, which includes GEBR, length, mass, nominal rotating mass, delay time of cutting traction and equivalent coast time of the train. *trackData* records the data of the track, which includes gradients and static speed restrictions of every point in the track. *D_train* indicates the position of the train, which is used to obtain the speed limit of EBI. *V_train* is the speed of the train. The speed of the train is used to compare against the speed limit of EBI to determine whether the emergency brake should be triggered. *B_release* indicates whether to release the emergency brake with data type *bool*. Only when train stops emergency brake can be released. *B_braking* indicates whether the train is braking with data type *bool*.

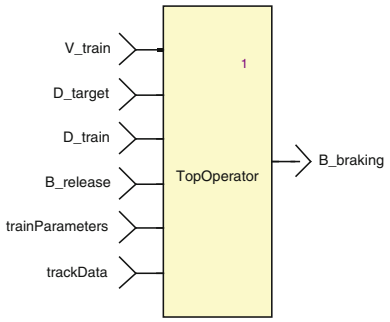


Fig. 10.3 Top-level dataflow diagram

The top-level dataflow diagram should be decomposed into several lower level dataflow diagrams. The decomposition is continued until every module is small and completes a simple and specific task.

The speed limits of EBD curve and EBI curve are continuous variables, which should be modeled by dataflow. Formulas (10.1) and (10.2) can be modeled as Fig. 10.4 shows with predefined mathematical operators. The operator 'grad' revises track data by the length of train *L_train* and outputs the deceleration due to gravity.

Speed supervision function determines whether the emergency brake should be triggered. The train running status is a discrete variable and should be modeled by SSM. A SSM consists of several independent states, and states are connected by transitions. A transition can be fired if its trigger condition is true.

SSM modeling should define states, transitions, and activities. The SSM model of speed supervision is displayed in Fig. 10.5. It includes two states, i.e., *Moving* and *Braking*. The state *Moving* is drew by the bold line, which indicates the state is the initial state. The system enters in the initial state first after it is started, and determines whether to enter in braking state by comparing the train speed and the speed limit. There is a condition on every transition. The condition of the transition from *Moving* to *Braking* is *V_train* >= *V_ebi*, which indicates the train is overspeed. The condition of the transition from *Braking* to *Moving* is *V_train* <= *0.0 and*

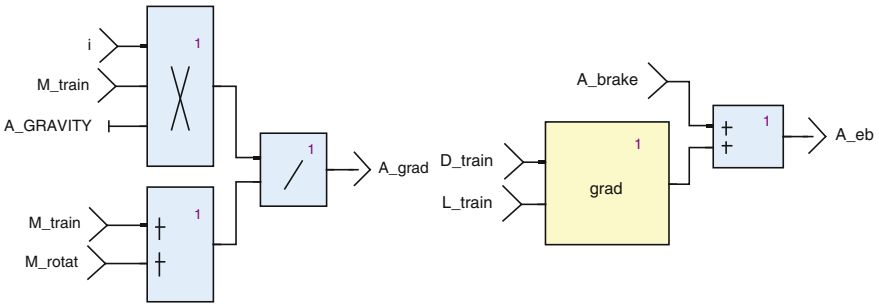
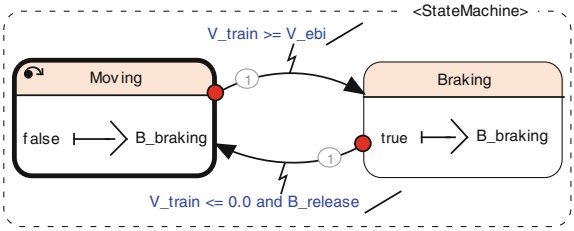


Fig. 10.4 SCADE models of the formulas

Fig. 10.5 Model of speed supervision



B_release, which indicates the train stops and the emergency brake is released manually. The activities of the states are defined. The braking signal is outputted if the train is braking, i.e., *B_braking* = *true*.

10.4 Safety Verification

10.4.1 Simulation Test

SCADE provides a simulator to perform simulation tests. Simulation test is a traditional verification method and is used in early phases to discover obvious errors.

The test cases of the speed supervision operator are displayed in Table 10.1, which records the sequence and values of the inputs and the outputs expected. The test cases are recorded in the form of the test case language used by SCADE simulator. For example, *SSM::set supervise/B_release* indicates the location of the input interface *B_release*, *false* indicates the value of *B_release* is false. Through the test the test results are identical with the outputs expected, and the operator passes the test.

The disadvantage of simulation testing is that the method can only discover errors, but cannot prove the model error-free. Only every paths are covered by test cases can model be checked completely, but in complex systems, test cases cannot achieve 100 % coverage. Therefore, even if every test case has been performed and passed, errors may still exist.

Table 10.1 Test cases of speed supervision operator

Cycle	Inputs	Outputs expected
1	SSM::set supervise/B_release false SSM::set supervise/V_train 1.0 SSM::set supervise/V_ebi 2.0	B_braking = false
2	SSM::set supervise/B_release false SSM::set supervise/V_train 1.99 SSM::set supervise/V_ebi 2.0	B_braking = false
3	SSM::set supervise/B_release false SSM::set supervise/V_train 2.0 SSM::set supervise/V_ebi 2.0	B_braking = true
4	SSM::set supervise/B_release true SSM::set supervise/V_train 1.0 SSM::set supervise/V_ebi 2.0	B_braking = true
5	SSM::set supervise/B_release false SSM::set supervise/V_train 0.0 SSM::set supervise/V_ebi 2.0	B_braking = true
6	SSM::set supervise/B_release true SSM::set supervise/V_train 0.0 SSM::set supervise/V_ebi 2.0	B_braking = false

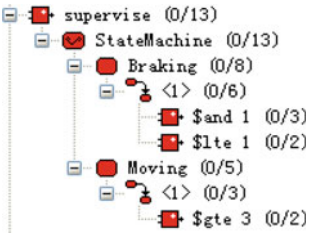


Fig. 10.6 Architecture of speed supervision model

10.4.2 Coverage Test

Model test coverage (MTC) is provided by SCADE to test coverage of test cases. For safety-critical software, strict coverage criteria should be used and coverage should reach 100 %.

MTC supports two kinds of coverage criteria for operators, i.e., decision coverage (DC) and modified condition/decision coverage (MC/DC). DC means test cases should have each decision evaluated both to true and false. MC/DC means test cases should make each decision take on every possible outcome, make each condition in a decision take on every possible outcome and make each condition in a decision independently affect the outcome of the decision.

MC/DC is stricter than DC, therefore we choose MC/DC. SCADE can analyze the architecture of the model. The architecture of the speed supervision model is displayed in Fig. 10.6. The model includes a state machine, which includes state *Braking* and state *Moving*. The states include a transition separately. In the figure, *$and* indicates 'and', *$lte* indicates 'less than or equal', and *$gte* indicates 'equal or greater than'. Brackets indicates the branches covered/all branched.

Every branch in the instrumented model is in red, which indicates the branch is not covered yet. MTC shows coverage of every cycle of test cases by performing the test cases. The coverage achieves 100 % after the test cases in Table 10.1.

10.4.3 Formal Verification

SCADE integrates a formal verification tool, which can perform formal verification for models. Formal verification plays a significant role in verification since test cases cannot achieve 100 % coverage in complex systems. Formal verification in SCADE includes the steps that defining safety properties, modeling safety properties, and performing verification usually.

1. Model Safety Properties

In Sect. 2.3, three safety properties are defined in the form of tautologies. The tautologies should be transformed into SCADE models as Fig. 10.7 shows. The

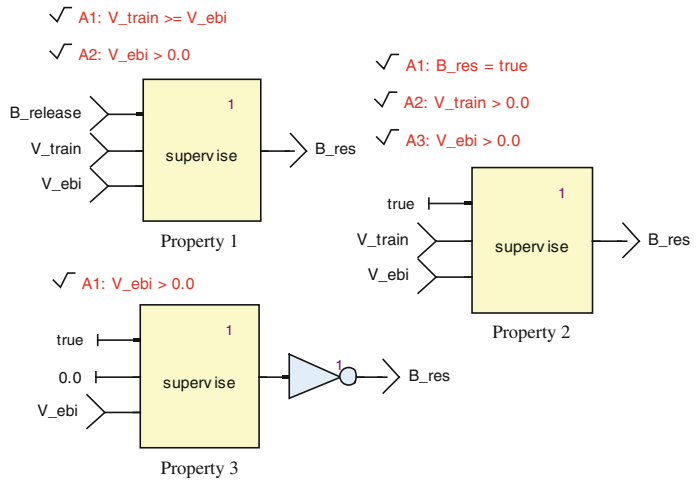


Fig. 10.7 SCADE models of safety properties

formulas above the operators are assumptions. The values of inputs must satisfy the assumptions. An interface can have any values if it is not limited by assumptions.

2. States Traversal

The *B_res* in Fig. 10.7 is equivalent to the corresponding tautology. Formal verification is to discover a group of input values to make *B_res* become *false*. If *B_res* is *true* under any input values, the formal verification is passed. Formal verification can be performed automatically. If the safety property is satisfied, a report of passing the verification is generated after the formal verification is performed. The designed SCADE models satisfy the safety properties by the formal verification.

Hidden errors can be discovered by formal verification. If the condition *V_train* >= *V_ebi* of the transition from state *Moving* to state *Braking* is written wrongly by the implementer and it becomes *V_train* > *V_ebi* as Fig. 10.8 shows, the error may not be discovered by traditional tests, but this hidden error can be discovered through formal verification.

The verification will not be passed when the wrong model is verified by the property 1. The counter example is generated in the form of test case as Fig. 10.9 shows. The property is not satisfied when the values of *V_ebi* and *V_train* is 1.0, which means the train speed is equivalent to the speed limit but the emergency brake is not triggered. The counter example can be performed by the simulator to discover the error.

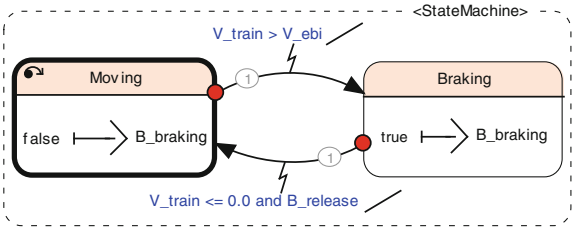


Fig. 10.8 Wrong model of speed supervision

```
SSM::set Formal::verif_01/V_train 1.
SSM::set Formal::verif_01/V_ebi 1.
SSM::set Formal::verif_01/B_release f

SSM::cycle
```

Fig. 10.9 Counter example

There is state explosion phenomenon in model checking, which leads to the difficulty to perform formal verification. Formal method should be applied in safety-critical functions, while other non-safety functions can use traditional development method. Completeness of safety properties is another issue in model checking. Model checking can prove that a designed model satisfies a safety property, but it is impossible to determine whether the given safety properties covers all the properties that the system should satisfy. Formal verification should be combined with traditional simulation test to improve reliability of the system.

10.5 Conclusion

Software in rail transit field needs high reliability, which cannot be achieved by traditional software development method. In this paper, we apply the formal method based on SCADE to the software development of onboard ATP. SCADE is used to model dynamic curve calculation and speed supervision. The model is tested and verified. The result suggests that formal method can improve reliability of software.

There are still some problems to be solved in formal method, such as state explosion problem in model checking and completeness of safety properties. The method also needs to be tested and verified in real systems.

Acknowledgment This work is partially supported by the Signal System Safety Method and Demonstration Research Project of Shanghai Shentong Metro Technology Center.

References

1. Xie G, Asano A, Hei X et al (2011) Formal verification of Communication based train control system. In: International conference on quality, reliability, risk, maintenance, and safety engineering. IEEE Press, Xi'an, China, pp 394–399
2. Xie G, Hei X, Mochizuki H et al (2012) Model based specification validation for automatic train protection and block system. In: 7th international conference on computing and convergence technology. IEEE Press, Seoul, South Korea, pp 485–488
3. Yan F (2011) Studying formal methods applications in CBTC. In: International conference on management and service science. IEEE Press, Wuhan, China, pp 1–3
4. Wu T, Dong Y, Hu N (2014) Formal specification and transformation method of system requirements from B method to AADL model. In: 17th international conference on computational science and engineering. IEEE Press, Chengdu, China, pp 1621–1628
5. Martin L, Schatalov M, Hagner M et al (2013) A methodology for model-based development and automated verification of software for aerospace systems. In: Aerospace conference. IEEE Press, Big Sky, MT, USA, pp 1–19
6. Wang H, Gao C, Liu S (2009) Model-based software development for automatic train protection system. In: Asia-Pacific conference on computational intelligence and industrial applications. IEEE Press, Wuhan, China, pp 463–466
7. Wang H, Liu S, Gao C(2009) Study on model-based safety verification of automatic train protection system. In: Asia-Pacific conference on computational intelligence and industrial applications. IEEE Press, Wuhan, China, pp 467–470
8. Wang H, Liu S (2010) Modeling communications-based train control system a case study. In: 2nd international conference on industrial mechatronics and automation. IEEE Press, Wuhan, China, pp 453–456
9. Quan Z, Choi S, Choi D et al (2011) Modeling for CBTC car-borne ATP/ATO functions and its applications. In: Proceedings of SICE annual conference. IEEE Press, Tokyo, Japan, pp 1929–1932
10. Cho C, Choi D, Quan Z (2011) Modeling of CBTC carborne ATO functions using SCADE. In: 11th international conference on control, automation and systems (ICCAS). IEEE Press, Gyeonggi-do, South Korea, pp 1089–1093
11. Wang X, Tang T, Liu S (2009) Study on modeling and verification of CBTC interlocking system. In: 5th IET international conference on wireless, mobile and multimedia networks. IET Press, Beijing, China, pp 467–470

Chapter 11
The Component Importance Evaluation Based on System Network and Fuzzy Partial Ordering Relation

Lifeng Bi, Yanhui Wang, Hao Shi and Shuai Lin

Abstract The reliability and safety of a system, notwithstanding it intended function, can be significantly influenced by the characteristics of components that define the system, especially the critical components. Importance evaluations are component related indices that allow to assessing how a component in the system affects one or more components. This paper assesses the component importance through two aspects. The first perspective is based on the system network modeled, which is the assessment of the structure position effect. The other aspect is the component reliability taken into account which reflects the function status on the basis of the fuzzy partial ordering relation model. Finally, a numerical instance of the bogie system is presented to show the rationality of the results compared with a previous method.

Keywords Network · Structure importance · Fuzzy partial ordering · Attribute importance · Bogie

11.1 Introduction

With the rapid development of science and technology, the scale and structure of the engineering system is becoming more and larger, such system is also referred to as complex system, like the power system, communication network, traffic system, and even large-scale circuit are all regarded as complex systems [1]. Importance evaluation is a key point in tracing bottlenecks in systems and in identifying the most important component.

In the course of the last decade complex network theory has become of great interest and has been widely applied in different domains including biology [2], communication [3], social science [4], economy, and climatology [5, 6]. Nodes

L. Bi (✉) · Y. Wang · H. Shi · S. Lin
State Key Laboratory of Rail Traffic Control and Safety, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China
e-mail: 1028314098@qq.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_11

importance is a vague concept, although lots of academic papers used it to describe the properties of networks. Researchers have reached little consensus on the concept, but fortunately most concede that it can indeed be described by some rules or the intuitive ideas [6]. Therefore, an improved method on the basis of the node contraction method in Ref. [7] is proposed to compute the structure importance of a node.

In this paper, the importance of components is associated with the following two factors. One is the structure location of the component in the network and the other is the failure properties of the component [8].

Evaluation of the failure properties of the component importance is part of the problem of multiple attributes decision [9]. Popular methods at present such as technique for order preference by similarity to an idea solution, fuzzy comprehensive evaluation, set-pair analysis, and information entropy were widely used in the modern industry. These methods combined the subjective experience and the objective data. In fact, the objective weight is difficult to accept. Based on the weakness, Refs. [10, 11] use an improved or more methods to research the property importance and try to avoid the influence of the subjective factors. In this paper, the model of fuzzy partial ordering is used to evaluate the attribute importance of components.

The rest of the manuscript is organized as follows: Sect. 11.2 summarizes symbols and hypothesis that are involved in the entire manuscript. Section 11.3 provides an improved method based on the node contraction method to assess the structure importance and the attribute importance on the basis of the order relation of rough set is prevented. In Sect. 11.4 the method with an example of the bogie system is illustrated. Finally, some conclusions drown are presented in Sect. 11.5.

11.2 Symbols and Hypothesis

The network is modeled as a graph, $S = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is a finite set of nodes and $E = \{e_1, e_2, \ldots, e_m\}$ is a finite set of directed edges among the corresponding nodes.

Definition 1 In the network, set the length of an edge is 1, because the edge only represents the connection relationship between nodes. The adjacency matrix can be expressed as follows:

$$h_{ij} = \begin{cases} 1, & i \text{ and } j \text{ directly connected} \\ 0, & \text{otherwise} \end{cases} \tag{11.2.1}$$

The graph representation is used to abstract the properties of the system, while retaining its structural properties.

11.3 The Model of Component Importance

The structure importance (SI) of the system indicates the structure position of the node in the network. Therefore, it is critical to assess the SI on the basis of the network. The importance of nodes primarily depends on two factors. One is the connection degree of the node, and the other one is the position of the node.

11.3.1 The Structure Importance

Definition 2 The node correlation degree is used to measure the degree of connection between nodes. There are more nodes connected to the node i demonstrate that i is more important. Based on this idea, the node correlation degree $C(i)$ of node i is defined as

$$C(i) = \frac{m(i)}{n} \tag{11.3.1}$$

where m means the number of nodes that connect with node i and n represents the total number of nodes.

Definition 3 The node society degree is adapted to assess the position of the node in the network. For instance, if somebody has high status in society, then his (her) influence is more important than others. The society degree $B(i)$ of node i is given by

$$B(i) = \frac{1}{\sum_{j=1}^{n} d(i,j)}, j \neq i \tag{11.3.2}$$

where $d(i,j)$ is the shortest path length from node i to j. Note that the smaller of the value $d(i,j)$, the society is more higher and indicates the node i is more important.

Definition 4 In order to integrate node correlation degree with society degree, the topologic property $T(i)$ of node i is defined as the sum of these two degrees, i.e.,

$$T(i) = C(i) + B(i) \tag{11.3.3}$$

To limit the value in between 0 and 1, the structure importance $S(i)$ of node i which is normalized is

$$S(i) = \frac{T(i)}{\sum_{j=1}^{n} T(j)} \tag{11.3.4}$$

11.3.2 The Function Properties Importance

The function relationships of components in the systems are related and even interacted. Therefore, it is necessary for us to study the function properties of components in the system, such as reliability, failure rate, and rams. To research the importance of function properties, the evaluation model is established based on the fuzzy partial ordering relation [12].

11.3.2.1 The Fuzzy Rough Set Theories

The relationships among reliability attributes can be described with the rough sets theory. Generally, the decision of multiple attributes can be expressed as following.

Definition 5 Consider (U, A, F), it represents the model of assessment value [12–14], where $U = \{x_1, x_2, \ldots x_n\}$ is a nonempty set which is called the assessment object set. x_i represents the object *i*. $A = \{a_1, a_2, \ldots, a_m\}$ is a set of assessment attribute, where a_l represents the attribute *l*. $F = \{f_l : U \rightarrow V_l(l \leq m)\}$ is a relation set between the object and attribute, where $f_l(x_i)$ is the relation value between the object x_i.

Definition 6 Consider (U, R) is the model of assessment relation [13], where R is the relation set between the objects. R is given by

$$R = \begin{bmatrix} R(x_1, x_1) & R(x_1, x_2) & \cdots & R(x_1, x_n) \\ R(x_2, x_1) & R(x_2, x_2) & \cdots & R(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ R(x_n, x_1) & R(x_n, x_2) & \cdots & R(x_n, x_n) \end{bmatrix} \tag{11.3.5}$$

11.3.2.2 The Fuzzy Partial Order Relation Model

According to the Ref. [13], the fuzzy partial order relation model is defined as

$$R(x_i, x_j) = \frac{\sum_{l=1}^{m} (f_l(x_i) \wedge f_l(x_j))}{\sum_{l=1}^{m} f_l(x_j)} \tag{11.3.6}$$

where "$\wedge$" means taking the minimum.

Table 11.1 The reliability ($r(i)$) of components

Node	$r(i)$	Node	$r(i)$	Node	$r(i)$	Node	$r(i)$	Node	$r(i)$	Node	$r(i)$	Node	$r(i)$
1	0.9922	5	0.9880	9	0.9856	13	0.9863	17	0.9872	21	0.9888	25	0.9858
2	0.9834	6	0.9908	10	0.9921	14	0.9863	18	0.9894	22	0.9894	26	0.9858
3	0.9865	7	0.9852	11	0.9912	15	0.9856	19	0.9910	23	0.9883	27	0.9858
4	0.9906	8	0.9884	12	0.9853	16	0.9856	20	0.9874	24	0.9705		

The function property importance $P(i)$ is defined as [11, 13]

$$P(i) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} R(v_i, v_j) \tag{11.3.7}$$

where n is the total number of objects.

11.3.3 The Component Importance Model

The coefficient θ is the proportion of the function property and represents the percentage of the component failures which are caused by the function. The synthesize importance $G(i)$ of component i is given by

$$G(i) = \theta S(i) + (1 - \theta) P(i) \tag{11.3.8}$$

11.4 An Illustrate Example

To show the efficiency and feasibility of the proposed approach, the component evaluation method is applied in a high-speed train bogie system [14]. Table 11.1 shows the components reliability $r(i)$ (Fig. 11.1).

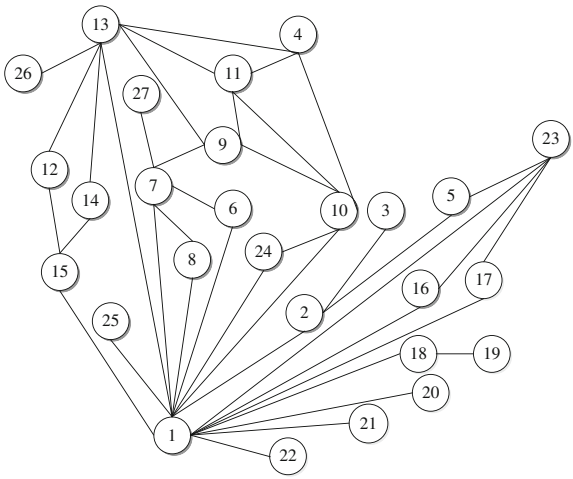


Fig. 11.1 The undirected network of bogie system

Table 11.2 The structure importance of nodes

Node	$s(i)$	Node	$s(i)$	Node	$s(i)$	Node	$s(i)$	Node	$s(i)$	Node	$s(i)$	Node	$s(i)$
1	0.6204	5	0.0864	9	0.1638	13	0.2797	17	0.0910	21	0.0534	25	0.0534
2	0.1293	6	0.0913	10	0.2037	14	0.0880	18	0.0910	22	0.0534	26	0.0506
3	0.0494	7	0.2037	11	0.1631	15	0.1287	19	0.0491	23	0.1657	27	0.0497
4	0.1250	8	0.0913	12	0.0880	16	0.0910	20	0.0532	24	0.0916		

Table 11.3 The attribute importance of components

Node	AI	Node	AI	Node	AI	Node	AI	Node	AI	Node	AI	Node	AI
1	1.0000	5	0.9992	9	0.9979	13	0.9984	17	0.9642	21	0.9992	25	0.9981
2	0.9960	6	0.9999	10	1.0000	14	0.9984	18	0.9997	22	0.9997	26	0.9981
3	0.9829	7	0.9976	11	0.9946	15	0.9946	19	0.9999	23	0.9993	27	0.9981
4	0.9999	8	0.9994	12	0.9977	16	0.9946	20	0.9990	24	0.9985		

11.4.1 The Structure Importance of Components

Table 11.2 shows the results of structure importance of nodes with the proposed method. The structure importance of node 1 is higher than others because the frame is the carrier of lots of components which connecting with it.

11.4.2 The Attribute Importance of Components

Based on the failure data, the component reliability can be obtained (the data comes from the CSR Qingdao Sifang Co., Ltd). Assuming the component obeys exponential distribution, and then we can get the reliability attribute.

From Eqs. (11.3.5) and (11.3.6), the fuzzy partial order relation set can be obtained and then using Eq. (11.3.7) we can get the attribute importance of components. Table 11.3 shows the attribute importance (AI). It is known that the node 1 and 11 have the highest AI which represents the importance of these two components in the view of function. The main functions of frame and wheels are bearing, steering, and driving, which determine the operation and steadiness of the high-speed train. If the wheels failed, it may cause an accident likes derailing and collision.

11.4.3 The Component Importance of the Bogie System

Finally component importance can be obtained by Eq. (11.3.8) when the coefficient is determined. Figure 11.2 shows the components importance in different θ. From the image, we can get the order of component importance. The components in the

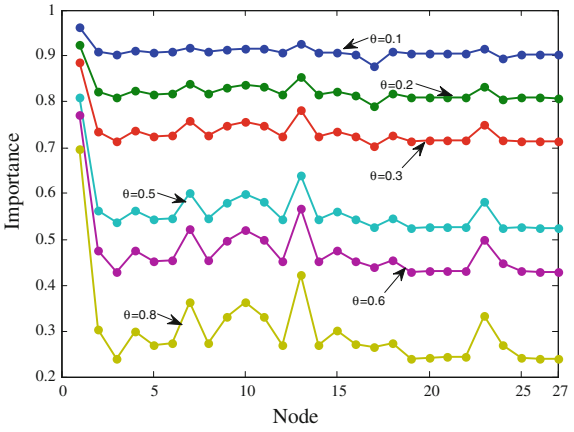


Fig. 11.2 Nodes importance in different θ

top six are 1, 7, 10, 11, 13, and 23. No matter from the view of structure and function the components 1, 7, 10, 11, and 13 are critical to the high-speed train. While component 23 provides the power for multiple components such as leveling value, air spring, and pressure-charging cylinder, and ensures their normal operation. Therefore, the method is reasonably presented in this paper.

11.5 Conclusions

The component importance of the high-speed train bogie system is researched based on the system network is modeled. We can get the results through calculating the structure importance and attribute importance. The structure importance is studied with an improved node contraction method. And the attribute importance is discussed by the fuzzy partial ordering relation model. The results of the example show that the method can well judge the component importance and indicates that the method is rational and feasible. Meanwhile, it provides the basis for making repair strategies.

Acknowledgments The project is supported by the independent subject (Grant No. I14K00451); National High Technology Research and Development Program of China (863 Program) (Grant No. T13B200011).

References

1. Yuan H (2011) Network topology for the application research of electrical control system fault propagation. Proc Eng 15:1748–1752
2. Zhou C, Zemanová L, Zamora G et al (2006) Hierarchical organization unveiled by functional connectivity in complex brain networks. Phys Rev Lett 97(23):238103
3. Capocci A, Servedio VDP, Colaiori F et al (2006) Preferential attachment in the growth of social networks: the internet encyclopedia Wikipedia. Phys Rev E 74(3):036116
4. Newman MEJ (2003) Properties of highly clustered networks. Phys Rev E 68(2):026121
5. Baskaran T, Blöchl F, Brück T et al (2011) The Heckscher-Ohlin model and the network structure of international trade. Int Rev Econ Financ 20(2):135–145
6. Zheng B, Li D, Chen G et al (2012) Ranking the importance of nodes of complex networks by the equivalence classes approach. arXiv preprint arXiv:1211.5484
7. Tan Y, Wu J, Deng H (2006) Evaluation method for node importance based on node contraction in complex networks. Syst Eng Theory Pract 11(11):79–83 (in Chinese)
8. Chybowski L (2011) A note on modifications to the methodology for components in the complex technical systems reliability structure importance evaluation. J Pol CIMAC 6(2):59–64
9. Chen Y, Hu AQ, Hu J et al (2004) A method for finding the most vital node in communication networks. High Technol Lett 1(2):573–575
10. Zhang Y, Guo J, Shan N (2010) Comprehensive evaluation of risk severity level of railway signal system. J Southwest Jiaotong Univ 5:016 (in Chinese)
11. Li GZ, Tan NL, Zhang JB (2012) Criticality analysis of subway train equipment based on improved analytical hierarchy process. J Electron Measur Instrum 26(6):503–507 (in Chinese)

12. Jia Z, Hou X (2013) Assessment of combat unit importance in networked antiaircraft system. Mod Defence Technol 41(5):12–16 (in Chinese)
13. Zhang WX, Qiu GF (2005) Uncertain decision making based on rough set. Tsinghua University Press, Beijing (in Chinese)
14. Zhang SG (2010) CRH2 EMU. China Railway Publishing House Press, Beijing (in Chinese)

Chapter 12
Reliability Study of Bogie System of High-Speed Train Based on Complex Networks Theory

Shuai Lin, Limin Jia, Yanhui Wang, Yong Qin and Man Li

Abstract Bogie system of high-speed train is described as a network, and its components represent nodes and connections represent edges (electrical connection, mechanical connection, and information connection). Several measures are discussed to investigate the topological characteristics of bogie system based on network. In order to analysis of system reliability, network agglomeration is introduced based on complex networks theory. Finally, reliability of bogie system is discussed according to network agglomeration.

Keywords Reliability · Bogie system · Networks · High-speed train · Agglomeration

12.1 Introduction

Bogie system is a typical complex mechatronic systems as well as the important part of high-speed train system. If a small failure of components occurs, it may lead to a loss of production and investments in components or casualties and damage of the system. Consequently, the reliability analysis of bogie system is a very significant problem and has become one of the most active research fields in high-speed train system.

S. Lin (✉) · M. Li
School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China
e-mail: 13114253@bjtu.edu.cn; linshuai2013@126.com

L. Jia · Y. Wang · Y. Qin
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

L. Jia · Y. Wang · Y. Qin
Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_12

Currently, a great number of evaluation methods have been used for reliability analysis and assessment, such as fault tree analysis (FTA), Markov analysis, Petri net, reliability block diagrams (RBD), and failure mode and effect analysis and derivatives (FMEA). Now a great number of methods have been proposed in the literature and used successfully sometimes. However, the applications of these methods are very limited, especially in the bogie system. This is because all these methods need relation information such as failure effects of components and systems and failure probability, which are often too difficult to obtain.

Fortunately, over the years, in complex network science [1–5], graph–theoretical network measures have proven to be a suitable tool to analyze the characteristics of different types of natural or man-made networks [6, 7]. One of the many network characteristics that researchers in this field have focused on in the past is network safety [8]. It has recently been suggested to apply such network measures to classical complex electromechanical system problems [9].

Recently, many networked models for complex system are provided. Different types of nodes (or edges) exist in the network, and nodes or edges may have a variety of properties associated with them [10]. However, the existing networks are not able to reflect the whole topological structure of system, such as mechanical structure. Therefore, a new method to construct network of system is proposed, and analyze reliability of system based on network.

In this paper, a new method for reliability analysis of bogie system based on complex networks theory is introduced. The remainder of the paper is organized as follows. Section 12.2 constructs the network of the bogie system. Section 12.3 introduces the definition of traditional measures in complex networks theory and then computes this measures in bogie system. Section 12.4 describes the proposed agglomeration measures based on the bogie system network and then analyzes reliability of bogie system by agglomeration measures. Lastly, Sect. 12.5 concludes the paper.

12.2 The Bogie System Network

12.2.1 Assumption

In order to facilitate further research, the simplifying assumptions employed here are described as follows: (i) Components are regarded as the minimum units which are consisted of bogie system. (ii) Physical connections between components include the mechanical connection, electrical connection, and information connection. (iii) The multi-connections between any pair of components can be ignored.

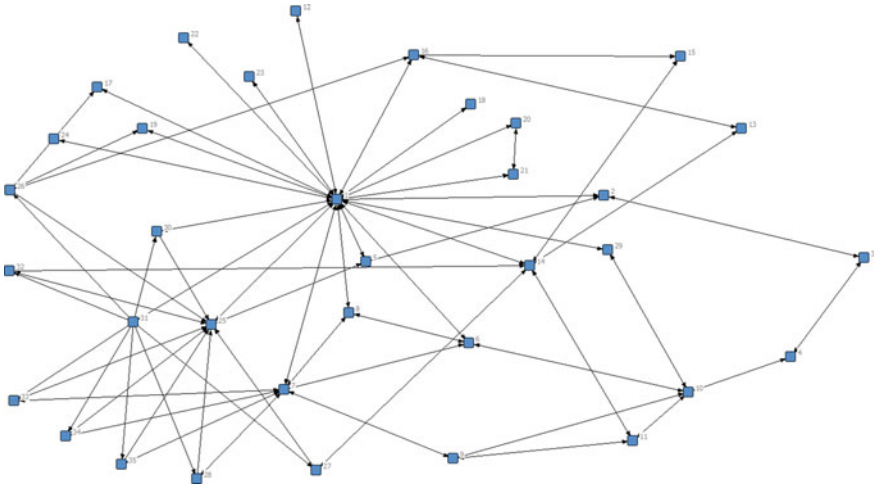


Fig. 12.1 The bogie system network

Table 12.1 The components that vertices represent in bogie system

No.	Name	No.	Name
1	Bogie frame	19	Air spring
2	Brake caliper	20	Center pin bush
3	Brake lining	21	Traction rod
4	Brake disks	22	Transverse shock absorber
5	Booster cylinder	23	Transverse backstop
6	Spring	24	Anti-side-rolling torsion bar
7	Axle box body	25	Control valve
8	Vertical shock absorber	26	Speed sensor 1
9	Bearing	27	Speed sensor 2
10	Wheel	28	LKJ2000
11	Axle	29	Device for cleaning the tread band of vehicle wheels
12	Secondary vertical shock absorber	30	Acceleration sensor
13	Railway coupling	31	Junction box
14	Gearbox	32	Temperature sensor bearing
15	Grounding device	33	Axle temperature sensor
16	Traction motor	34	AG37
17	Height adjusting device	35	AG43
18	Anti-hunting damper		

12.2.2 Network of Bogie System

In the typical bogie system, there are a lot of components, such as bogie frame, and brake caliper, while brake lining, traction tod, anti-side-rolling torsion bar and other sensors. All components are connected into a holistic system, and the connections among components may be mechanical connection, electrical connection, and information connection. The scheme is shown in Fig. 12.1. The bogie system was abstracted as a network model, which the components were abstracted as the vertices, and the connections between two components can be represented by the edges as shown in Fig. 12.1 and Table 12.1 In fact, the network model is a relatively complex system.

12.3 The Topological Characteristics

12.3.1 Degree Centrality

Degree k_i is expressed as the number of edges connected to a node v_i, given by

$$k_i = \sum_{j}^{N} x_{ij} \tag{12.1}$$

where v_i is the focal node, v_j represents all other nodes, N is the total number of vertices, and x_{ij} represents the connection between node v_i and v_j. The value of x_{ij} is defined as 1 if node v_i is connected to node v_j, and 0 otherwise.

K is the average degree of nodes.

From Fig. 12.2, the largest degree is node 1. The out-degree and in-degree are not always same. And the average degree of nodes is $K = 3.714$ in bogie system

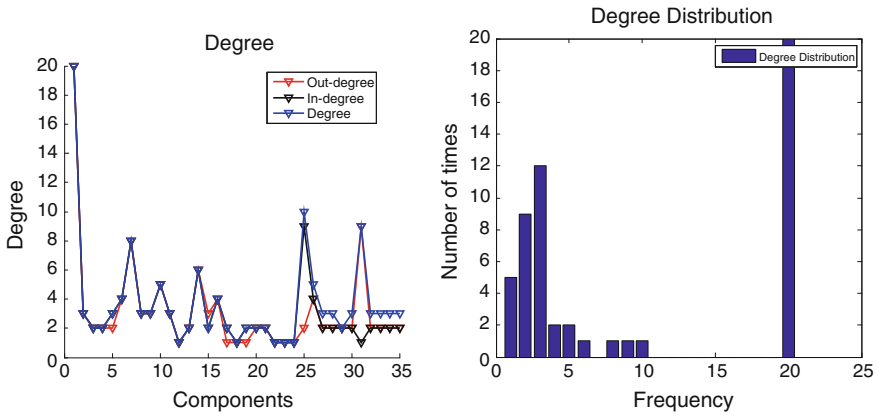


Fig. 12.2 Degree and degree distribution of each node in bogie system network

network. Figure 12.1 shows the node degree in bogie system network, obviously, most nodes have very few edges, however, and a few nodes have a great deal of connections. So it can be seen that the bogie system network is an inhomogeneous network.

12.3.2 Average Shortest Length

The average shortest length L (the average path length) is discussed as the average over the network of the shortest path length among all nodes in the network:

$$L = \frac{2}{N(N-1)} \sum_{1 \le i,j \le N} d_{ij} \tag{12.2}$$

where N is the total number of nodes, d_{ij} is the number of edges along the average path connecting node v_i and v_j.

For the bogie system network, the shortest path of all nodes is shown in Fig. 12.4. And the average shortest length is 2.494. In fact, this network has a low diameter and a small average shortest length.

12.3.3 Clustering Coefficient

The clustering coefficient C indicates a local property of node. It is defined as the average fraction of neighbors of a node-pair that are also neighbors of each other.

$$C = \frac{1}{N} \sum_{i=1}^{N} \frac{2t_i}{k_i(k_i - 1)} \tag{12.3}$$

where k_i is the number of edges which connected to a node i, t_i is number edges among the node neighbors i.

The clustering coefficient of the bogie system network is 0.605. Figure 12.3 shows that the clustering coefficient of all nodes in the bogie system network. It indicates that the bogie system network a high clustering coefficient, like regular network.

12.3.4 Node Betweenness

The node betweenness B_i is described as the total number of average paths between nodes in the network, which pass through a certain node v_i. The exact definition of the betweenness centrality B_i of a node v_i is the sum over the number of shortest

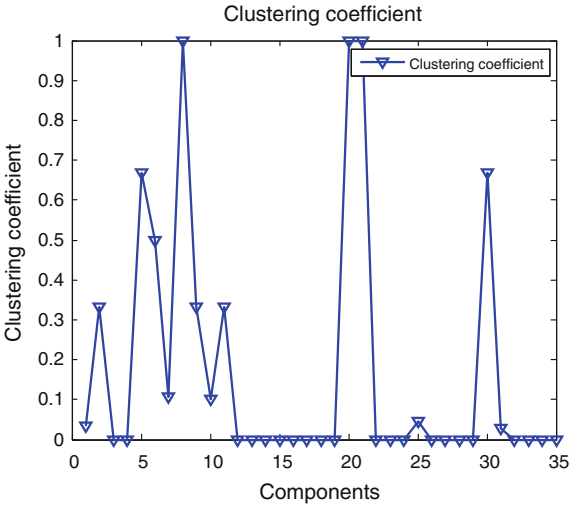


Fig. 12.3 The clustering coefficient of all nodes in bogie system network

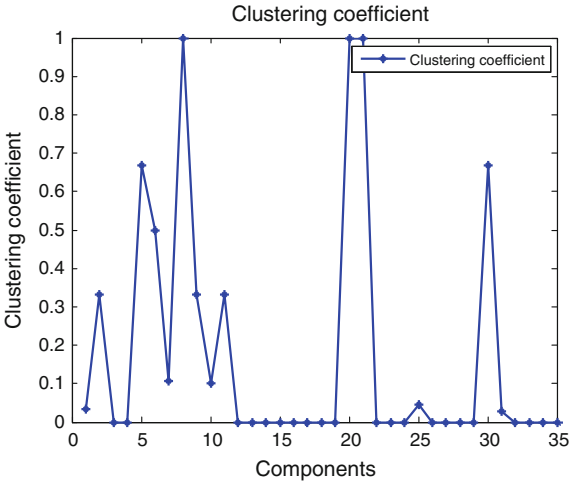


Fig. 12.4 The node betweenness in bogie system network

path d between all pairs of nodes v_i and v_j that pass through node v_v, normalized by the total number of average paths between v_i and v_j [10], see (12.4):

$$B_i = \sum_{ij} \frac{d_{i,j}(v)}{d_{i,j}} \tag{12.4}$$

Figure 12.4 shows the betweenness of all vertices in bogie system network. In fact, node betweenness is always used to identify critical component. Therefore, the highest betweenness of node should be paid more attention.

12.4 Reliability Analysis Based on Bogie System Network

12.4.1 Network Agglomeration

Network agglomeration is described that the connectivity of localized network G_i, which is constructed by node v_i and its adjacent nodes. Let $R(G_i)$ denote the network agglomeration, then

$$R = (k_i L_i)^{-1} = \left(k_i \frac{\sum_{s,t \in G} d_{s,t}(i)}{\frac{k_i(k_i-1)}{2}} \right) = \frac{(k_i - 1)}{2 \sum_{s,t \in G} d_{s,t}(i)} \tag{12.5}$$

where k_i is degree centrality of node v_i; l_i is average path length in localized network G_i. $d_{s,t}(i)$ is the shortest path between all pairs of nodes v_s and v_t that pass through node v_i. Obviously, $0 < R(G_i) \leq 1$, and $N = 1$, we defined the network agglomeration as 1.

Following from the definition of network reliability, we recognize that the network agglomeration lies on two factors: (i) The degree centrality k_i. If the degree of v_i is higher, the number of adjacent nodes is more, and the network agglomeration is higher. (ii) The position of v_i. If v_i is located in a critical position, many average paths of nodes pairs pass v_i, so we can get the higher the connectivity of localized network G_i. For bogie system, $R(G_i)$ the higher $R(G_i)$ indicates that effective redundancy of v_i is better, and system reliability is higher.

12.4.2 Simulation and Analysis

For bogie system network, we can calculate network agglomeration as shown in Table 12.2. Table 12.2 shows that the largest and smallest network agglomeration is

Table 12.2 Network agglomeration

Node	$R(G_i)$	Node	$R(G_i)$	Node	$R(G_i)$
1	0.0008	13	0.0032	25	0.0231
2	0.0124	14	0.0142	26	0.0218
3	0.0164	15	0.0069	27	0.0136
4	0.0324	16	0.0168	28	0.0269
5	0.0246	17	0.0499	29	0.0825
6	0.0127	18	0.0319	30	0.0248
7	0.0034	19	0.0376	31	0.0216
8	0.0163	20	0.0380	32	0.0214
9	0.0035	21	0.0654	33	0.0242
10	0.0039	22	0.0825	34	0.0231
11	0.0075	23	0.0807	35	0.0218
12	0.0520	24	0.0566		

node 29 (Device for cleaning the tread band of vehicle wheels) and node 1 (Bogie Frame), respectively. It indicates that bogie frame should be paid more attention in maintenance process.

12.5 Conclusion

In fact, we present a new measure—network agglomeration—for reliability analysis based on networks theory. First, the network model of the bogie system is constructed and then we analyze the topological properties based on above network. It can be seen that the network of the bogie system has a small average path length and a high clustering coefficient. But it is neither completely random network nor completely regular network. The each node degree is different from each other, which we called heterogeneous. Then we introduce network agglomeration measure to study system reliability in network. The network agglomeration reflects the connectivity of localized network, i.e., system reliability. If the node degree is greater or its position is pivotal, the network agglomeration after contraction is higher, and then system reliability is also higher.

Acknowledgments The project is supported by the independent subject (Grant No. I14K00451); National High Technology Research and Development Program of China (863 Program) (Grant No. T13B200011).

References

1. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393 (6684):440–442
2. Amaral LAN, Scala A, Barthelemy M et al (2000) Classes of small-world networks. Proc Natl Acad Sci 97(21):11149–11152
3. Kleinberg J (2000) The small-world phenomenon: an algorithmic perspective. In: Proceedings of the thirty-second annual ACM symposium on Theory of computing. ACM, pp 163–170
4. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Phys Rev Lett 86(14):3200
5. Barabási AL, Albert R, Jeong H (1999) Mean-field theory for scale-free random networks. Phys A 272(1):173–187
6. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74 (1):47
7. Boccaletti S, Latora V, Moreno Y et al (2006) Complex networks: structure and dynamics. Phys Rep 424(4):175–308
8. Husejnagić D, Butala P (2012) Discovering autonomous structures within complex networks of work systems. CIRP Ann Manufact Technol 61(1):423–426
9. Becker T, Meyer M, Windt K (2013) A network theory approach for robustness measurement in dynamic manufacturing systems. In: 17th Annual Cambridge international manufacturing symposium, Cambridge, Sept 2013
10. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74 (1):47

Chapter 13
Fault Identification Method Based on Fuzzy Fault Petri Net

Yanhui Wang, Yujing Lu, Man Li and Lifeng Bi

Abstract The technology of intelligence fault recognition develops quickly, which makes a new way for the reliability and safety of equipment-integrated system. With the development of complex system, the existing method is difficult to meet the needs of fault diagnosis, the fault recognition method based on the fuzzy fault Petri nets (FFPN) was put forward. In order to make the process of fault propagation simple and convenient, based on relationship matrix and state equation, fault propagation flow function and the set of occurred transition are introduced . The path of fault recognition is gained through forward and backward reasoning algorithm utilizing associated failure between equipment. Finally, the validity and practicability of this approach were proved by a case study in the associated failure model of a high-speed train traction system.

Keywords Fault identification · Fuzzy fault petri net · Forward reasoning

13.1 Introduction

Fault diagnosis is of great importance to all kinds of industries in the competitive global market today [1]. Now, in the area of fault recognition, more and more people prefer to adopt the method of computing intelligence and knowledge engineering. Such as, based on expert system theory, Rusheng et al. [2] put forward a new way to express the expert's knowledge in this field which is to divide the knowledge into failure phenomena, failure causes, and failure rules. Tayarani-Bathaie et al. [3] put forward a method that applies dynamic neural networks in the fault recognition of the dual spool turbo fan engine. By optimizing the

Y. Wang · Y. Lu (✉) · M. Li · L. Bi
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing, China
e-mail: 14120861@bjtu.edu.cn

Y. Wang · Y. Lu · M. Li · L. Bi
School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_13

parameters of SVM by chaos particle swarm, Rusheng et al. [4] put forward the fault diagnosis of sensor by chaos particle swarm optimization algorithm and support vector machine. Jiang et al. [5] present a new methodology for predicting failures of a gear shaft system, using the EM algorithm and an optimal Bayesian fault prediction scheme to estimate the model parameters. Pamuk et al. [6] apply Fuzzy Petri Nets technology in fault diagnosis of power system, which breaks a new path to improve fault recognition technology. However, those fault diagnosis methods are more and more difficult to meet requirements.

In this paper, we put forward the fault recognition method based on fuzzy fault Petri net. First, in order to make the process of fault propagation more simple, the fault propagation flow function and the set of occurred transition are introduced . Second, in order to adapt to the multicomponent and multifunction characteristics of complex equipment system, improve the propagation algorithm of traditional Petri net: according to the fuzzy reasoning rule and the characteristics of failure propagation, describe the uncertainty of fault information accurately. Third, utilizing associated failure between equipment, get the path of fault recognition by forward and backward reasoning algorithm. Finally, the validity and practicability of this approach were proved by a case study in the associated failure model of a high-speed train traction system.

13.2 Fuzzy Fault Petri Net

A Petri net is a one of several mathematical and graphical representations of discrete distributed systems [7, 8]. In this section, we introduce the formal description of fuzzy fault Petri net (FFPN).

Definition 1 A fuzzy fault Petri net (FFPN) is defined as an 11-tuple. FFPN = $(P, T, D, I, O, W, \alpha, F, \lambda, U, M)$, where

$P = \{p_1, p_2, \ldots, p_n\}, T = \{t_1, t_2, \ldots, t_n\}, D = \{d_1, d_2, \ldots, d_n\}$ are finite sets of places, transitions, and propositions, respectively;

$I : (P \times T)_{m \times n}$ is an input incidence matrix, representing the directed arcs from places to transitions;

$O : (T \times P)_{n \times m}$ is an output incidence matrix, representing the directed arcs from transitions to places;

$W = (w_1, w_2, \ldots, w_n)^T, \theta = \{\theta_1, \theta_2, \theta_n\}, \lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ are weighted set of arc, probability set of the existence of token, and threshold set of transition, respectively. $w_i \in \{0, 1\}, i = 0, 1, \ldots, n$. Make $\theta_i (i = 1, 2, \ldots, n) = 1$;

$M = \{m_1, m_2, \ldots, m_n\}$ denotes the set of the marking of FFPN, $m_i \in \{0, 1\}$;

F denotes the set of the relationship of fault transmission flow;

T' denotes the set of transition which has occurred, the set is empty in initial, when t_j happened, $T' = T + \{t_j\}$, finally, $T = \emptyset$.

FFPN rules are the basis of simulating fault propagation process of system. First, we give the transition firing rules and the fuzzy production rule [9]. According to

the numerical description of the statistical analysis of historical failure data, by the normalized processing, determine confidence level C, reduce the effects subjectivity of expert experience effectively.

Definition 2 Transition firing rule: under the current marking M, if the transition t_1 met the conditional (1): $w_i \times \theta_i \geq \lambda_i$, and the conditional (2): $\lambda_k \notin T'$, copy the token which is in input place to output place, according to the corresponding rule change θ_i.

13.3 The Analysis of Failure Propagation Based on FFPN

Forward and reverse reasoning algorithms are the key to determine the fault propagation path. First, deriving the result by forward reasoning propagation; Then searching the reason by reverse reasoning propagation and fuzzy rules. If the forward and reverse reasoning propagation paths are the same, the fault identification process is effective.

13.3.1 Forward Reasoning Algorithm

Step1: According to the associated failure model of system, build the model of FFPN.
Step2: Determine the corresponding place of initial fault.
Step3: Determine the initial marking M_0 in the model and the initial credibility of each place, as well as the threshold value and weight of transition through collecting failure data and other information of system.
Step4: Under the initial marking M_0, according to the fault propagation mode, calculate possible failure phenomenon caused by the occurred transition, namely the phenomenon of place which caused by transition and get a new set T'.
Step5: Repeat Step 4, reach the final state until no transition can happen.
Step6: Observation T', find the connecting transition, and then find the failure propagation paths, the process of reasoning is end.

13.3.2 Reverse Reasoning Algorithm

A complex system has the properties of multicomponent, multifunction, according to the fuzzy inference rules, and the probability of occurred transition, achieve reverse the fuzzy reasoning by selecting the end nodes from big to small. For facilitating, introduce the concept in the reasoning process of FFPN.

Definition 3 Set t_x for a transition, p_i, p_j, and p_m are three places associated with t_x, if $p_i \in {}^*t_x$, at this time, p_i demote the directly backflow associate place of p_j, the set of directly backflow associated place of p_j, down as $\mathrm{DRC}(p_i)$. If p_i is the directly backflow associated place of p_i, p_j is the directly backflow associated place of p_m, it is called a backflow associated place of p_m, the set including the backflow associated place, down as $\mathrm{RC}(p_i)$.

Step 1: The initial object place $(p_i, \mathrm{DRC}(p_i), \theta(p_i))$ is an end node, where p_i is the target place, ${}^*p_i = t_i, \mathrm{DRC}(p_i)$, *is the set of directly backflow associated place of* p_i, $\theta(p_i) = 1$ denotes that the equipment is failed.

Step 2: Reverse inference by the four patterns: one cause and one effect, several causes and one effect, one cause and several effects, competition pattern.

Definition 4 Fuzzy production rule:

(1) One cause and one effect pattern
If p_i Then p_j $(C = \mu_R)$ $i, j = 1, 2, \ldots, n$. It can be modeled as shown in Fig. 13.1. The transition t_k has only one input arc $p_i \rightarrow t_k$, if $\theta_i \times w_i \geq \lambda_k$ and $\lambda_k \notin T'$, then transition t_k is ignited in Fig. 13.1 and the value of θ_{i+1} can be update.

$$\theta_2 = \begin{cases} \min\left\{\frac{w_1 \times \theta_1 \times w_2}{\lambda}, 1\right\}, & p \in t_i^* \\ \theta_1, & p \notin t_i^* \end{cases} \tag{13.1}$$

(2) Several cause and one effect pattern
If $p_1 \wedge p_2 \wedge \cdots \wedge p_{n-1}$ Then p_n $(C = \mu_R)$. It can be modeled as shown in Fig. 13.2. The transition t_k has several input arcs, $p_i \rightarrow t_k (i = 1, \ldots, n)$, if $\sum_{i=1}^{n-1} \theta_i \times w_i \geq \lambda_k$ and $\lambda_k \notin T'$, then transition t_k is ignited in Fig. 13.2 and the value of θ_{n+1} can be update.

$$\theta_{n+1} = \begin{cases} \min\left\{\frac{w_{11} \times \min(\theta_1 \times w_1, \theta_2 \times w_2, \ldots, \theta_n \times w_n)}{\lambda}, 1\right\}, & P_{n+1} \in t_1^* \\ \theta_i (i = 1, 2, \ldots, n) & P_{n+1} \notin t_1^* \end{cases}, \tag{13.2}$$

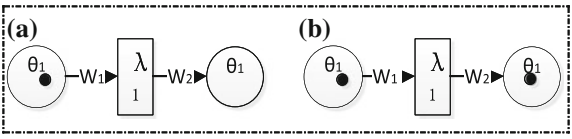


Fig. 13.1 One cause and one effect pattern (**a**) the state before the tranmission trigger, (**b**) the state after the tranmission trigger

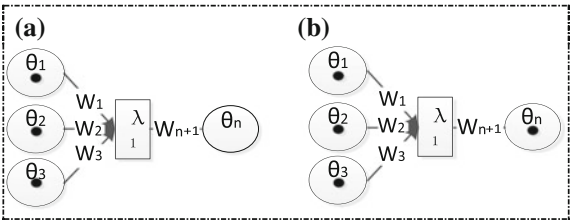


Fig. 13.2 Several cause ad one effect pattern (**a**) the state before the tranmission trigger, (**b**) the state after the tranmission trigger

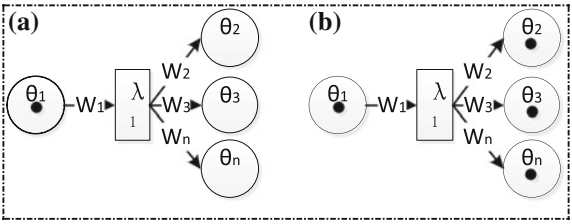


Fig. 13.3 One cause and several effect pattern (**a**) the state before the tranmission trigger, (**b**) the state after the tranmission trigger

(3) One cause and several effects pattern
If p_1 then $p_2 \wedge P_3 \wedge \cdots p_n(C = \mu_R)$. It can be modeled as shown in Fig. 13.3. The transition t_k has only one input arc, $p_1 \rightarrow t_k$, if $\theta_i \times w_i \geq \lambda_k$ and $\lambda_k \notin T'$, then transition t_k is ignited in Fig. 13.3 and the value of θ_i can be update.

$$\theta_i(i = 1, 2, \ldots, n) = \begin{cases} \min\left\{\frac{w_1 \times \theta_1 \times w_{11}}{\lambda_1}, 1\right\}, & p_2 \in t_1^* \\ \min\left\{\frac{w_1 \times \theta_1 \times w_{12}}{\lambda_1}, 1\right\}, & p_3 \in t_1^* \\ \vdots & \vdots \\ \min\left\{\frac{w_1 \times \theta_1 \times w_{1n}}{\lambda_1}, 1\right\}, & p_n \in t_1^* \\ \theta_1, & p_i \notin t_1^*(i = 2, 3, \ldots, n) \end{cases} \tag{13.3}$$

(4) Competition pattern
If $p_1 \vee p_2 \vee \cdots \vee p_{n-1}$ Then p_n $(C = \mu_R)$. It can be modeled as shown in Fig. 13.4. The transition t_k has several input arcs, if $p_i \rightarrow t_k(i = 1, \ldots, n)$ and $\lambda_k \notin T'$, then transition $t_i(i = 1, \ldots, n)$ is ignited in Fig. 13.4 and the value of $t_i(i = 1, \ldots, n)$ can be update.

$$\theta_{n+1} = \begin{cases} min\left\{\max\left(\frac{w_1 \times \theta_1 \times w_{11}}{\lambda_1}, \frac{w_2 \times \theta_2 \times w_{21}}{\lambda_2}, \ldots, \frac{w_n \times \theta_n \times w_{n1}}{\lambda_n}\right), 1\right\}, & p_{n+1} \in t_i^*(i = 1, 2, \ldots, n) \\ \theta_i, & p_{n+1} \notin t_i^*(i = 1, 2, \ldots, n) \end{cases} \tag{13.4}$$

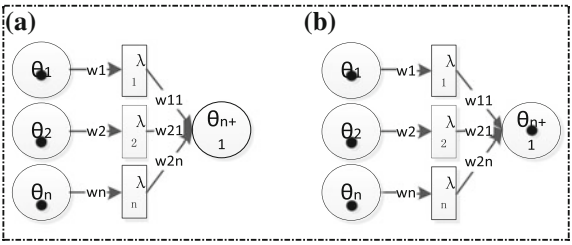


Fig. 13.4 Competition pattern (**a**) the state before the tranmission trigger, (**b**) the state after the tranmission trigger

In the above fuzzy production rules, $p_1, p_2, \ldots, p_n$ are fuzzy formulas, C is confidence level, μ_R is the confidence level of fuzzy rule.

13.4 Case Studies

High-speed train traction system is the main power source of the train, so it will always break down. According to the analysis of the historical fault data, and failure analysis of associated equipment, get the FFPN model of traction system. And based on the historical failure data of each equipment, get initial credibility of place, the meaning of transition, threshold value, and corresponding weights, as shown in Figs. 13.5 and 13.6 (Tables 13.1 and 13.2).

Forward reasoning: when p_{17} failed, $p_{17} \rightarrow t_4$, because $\theta_{17} \times w_{p_{17}t_4} \geq \lambda_4$, t_4 can happen, by the reasoning rule, $\theta_{14} = (\theta_{17} \times w_{p_{17}t_4} \times w_{t_4p_{14}})/\lambda_4, \theta_{14} = 0.057$, $T' = \{t_4\}$, the next step, reasoning according to the forward reasoning algorithm like this, finally $\theta_{13} = 1$ and $T' = \{t_{22}, t_1, t_5, t_6\}$.

Reverse reasoning: Assume p_{13} fails, then $\theta_{13} = 1$. According to reverse reasoning rule, $\theta_{10} = 0.874$. At this time, there are two reverse path sets, respectively, $\{(p_{10}, t_5, t_4, p_{17})\}$, and $\{(p_{10}, t_{10}, p_9, t_{12}, p_7, t_{15}, p_1, t_{20}, p_2, t_{21}, p_3)\}$. The next step, reasoning according to the reverse reasoning like this, finally $\theta_{17} > \theta_3$, so, the reverse path is $p_{10} \rightarrow t_5 \rightarrow p_{14} \rightarrow t_4 \rightarrow p_{17}$.

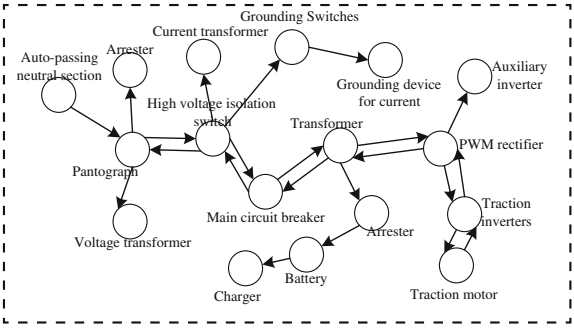


Fig. 13.5 Associated failure model

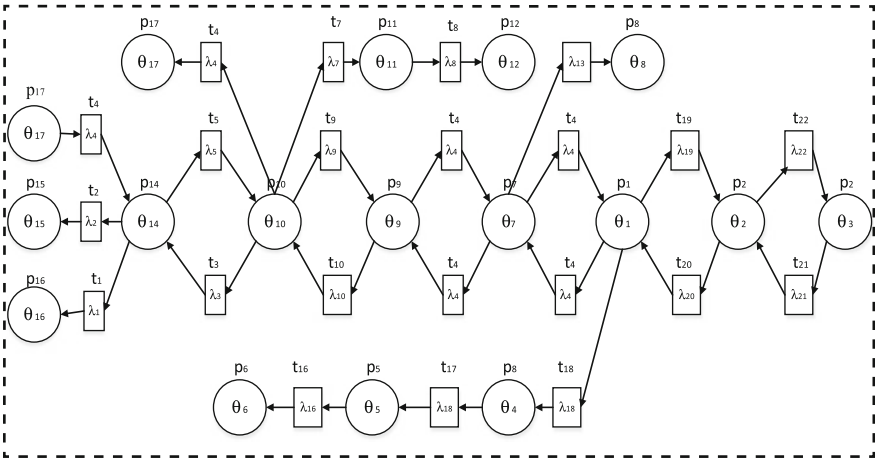


Fig. 13.6 Improve petri network models

Table 13.1 The meaning of the place and its initial credibility

Place name	Place meaning (failure parts)	Initial credibility	Place name	Place meaning (failure parts)	Initial credibility
p_1	PWM rectifier	*	p_{10}	High-voltage isolation switch	*
p_2	Traction inverters	*	p_{11}	Grounding switches	*
p_3	Traction motor failure	1	p_{12}	Grounding device for current	*
p_4	Auxiliary inverter	*	p_{13}	Current transformer	1
p_5	Charger	*	p_{14}	Pantograph	*
p_6	Battery	1	p_{15}	Auto-passing neutral section	1
p_7	Transformer	*	p_{16}	Voltage transformer	*
p_8	Arrester	1	p_{17}	Surge arrester failure	1
p_9	Main circuit breaker	*			

*Is unknown parameter

Assume that the traction system works correctly, according to the history data, we troubleshoot the equipment. By the forward reasoning: from the analysis of the whole process when p_{17} failed, the set of propagation path most likely to be $\{(p_{17}, t_4, p_{14}, t_5, p_{10}, t_6, p_{13})\}$, we should also pay attention to t_1 at the same time. By the reverse reasoning, according to the reverse reasoning process and result, when p_{13} fails, the most likely reason is p_{17}, which is consistent with forward reasoning. Meantime, Auto-passing neutral section failure is easy to cause surge arrester fails. The forward and reverse reasoning propagation paths are the same, the fault identification process is effective.

Table 13.2 Transition meaning, and its corresponding threshold and weight

Transition name	Transition meaning	Threshold		Weight
t_1	Pantograph and voltage transformer fail	0.0026	0.0490	0.5690
t_2	Pantograph and auto-passing neutral section fail	0.0140	0.0490	0.0960
t_3	High-voltage isolation switch and pantograph fail	0.0350	0.6950	0.0490
t_4	Pantograph and auto-passing neutral section fail	0.0590	0.0690	0.6480
t_5	High-voltage isolation switch and pantograph fail	0.0021	0.0490	0.6950
t_6	High-voltage isolation switch and current transformer fail	0.5940	0.6950	0.9780
t_7	High-voltage isolation switch and grounding switches fail	0.0350	0.6950	0.0790
t_8	Grounding device for current and grounding switches fail	0.0590	0.0790	0.0750
t_9	High-voltage isolation switch and main circuit breaker fail	0.0540	0.6950	0.7690
t_{10}	High-voltage isolation switch and main circuit breaker fail	0.0540	0.2690	0.6950
t_{11}	Transformer and main circuit breaker fail	0.0560	0.2690	0.6870
t_{12}	Transformer and main circuit breaker fail	0.0560	0.6870	0.0690
t_{13}	Transformer and arrester fail	0.1100	0.6870	0.1250
t_{14}	Transformer and PWM rectifier fail	0.0480	0.6870	0.0540
t_{15}	Transformer and PWM rectifier fail	0.0480	0.0540	0.6870
t_{16}	Battery and charger fail	0.0540	0.0590	0.0690
t_{17}	Auxiliary inverter and charge both fail	0.0120	0.6890	0.0231
t_{18}	Auxiliary inverter and PWM rectifier fail	0.0480	0.0540	0.6870
t_{19}	Traction inverters and PWM rectifier fail	0.0510	0.0540	0.7890
t_{20}	Traction inverters and PWM rectifier failure fail	0.0510	0.7890	0.0540
t_{21}	Traction inverters and traction motor fail	0.6230	0.6480	0.7890
t_{22}	Traction inverters and traction motor fail	0.6230	0.7890	0.6480

13.5 Conclusion

In this paper, we redefine the fuzzy fault Petri nets (FFPN), introduce the fault propagation flow function and the set of transition which has occurred, which makes the fault propagation process more simple and convenient; improve the traditional Petri net propagation algorithm, according to the fuzzy reasoning rule and the failure propagation characteristics to describe the uncertainty fault information accurately; utilizing associated failure between equipment, by forward and backward reasoning algorithm, get the path of fault recognition. Finally, the validity and practicability of this approach were proved by a case study in the associated failure model of a high-speed train traction system.

Acknowledgments The project is supported by the independent subject (Grant No. I14K00451); National High Technology Research and Development Program of China (863 Program) (Grant No. T13B200011)

References

1. Liu HC, Lin QL, Ren ML (2013) Fault diagnosis and cause analysis using fuzzy evidential reasoning approach and dynamic adaptive fuzzy petri nets. Comput Ind Eng 66(4):899–908
2. Rusheng Z, Zongxia J, Shaoping W et al (2008) Fault diagnosis expert system for hydraulic system of missile launcher. Acta Aeronaut Et Astronaut Sin Ser A B 29(1):197
3. Tayarani-Bathaie SS, Vanini ZNS, Khorasani K (2014) Dynamic neural network-based fault diagnosis of gas turbine engines. Neurocomputing 125:153–165
4. Chenglin Z, Xuebin S, Songlin S et al (2011) Fault diagnosis of sensor by chaos particle swarm optimization algorithm and support vector machine. Expert Syst Appl 38(8):9908–9912
5. Jiang R, Yu J, Makis V (2012) Optimal Bayesian estimation and control scheme for gear shaft fault detection. Comput Ind Eng 63(4):754–762
6. Pamuk N, Uyaroglu Y (2012) The fault diagnosis for power system using fuzzy petri nets. Przeglad Elektrotechniczny 88(7):231–236
7. Zurawski R, Zhou M (1994) Petri nets and industrial application: atutorial. IEEE Trans Ind Electron 41(6):567–583
8. Desel J, Esparza J (1995) Free choice petri nets. Cambridge University Press, UK
9. Chen S, Ke J, Chang J (1990) Knowledge representation using fuzzy petri nets. IEEE Trans Knowl Data Eng 2(3):311–319

Chapter 14
Correlation Failure Analysis Based on the Improved FP-Growth Algorithm

Yanhui Wang, Shujun Wang and Shuai Lin

Abstract In consequence of the characteristics of interdependence, inter-constraint, and highly coupling for complex equipment systems, there are complex correlations between equipments. Then a large number of diversified fault records are produced from the operating process of systems, and these fault records become reflected forms about correlation failures. Determining how to use the fault information feature words—mined from failure records—of equipment effectively to obtain the correlation failure relationship between the equipment becomes the focus of this paper. Given this, we propose the improved FP-Growth algorithm which is obtained through optimizing the classical correlation rule mining algorithm, with experimental results proving it to be actually efficient on running time for mining frequent patterns; based on the improved FP-Growth, correlation analysis is carried out based on the fault information feature words of equipment, then correlation failure rules are extracted, and a correlation failure model is established. At last, taking the CRH2 EMU traction system as example, the correlation failure rules are obtained and a correlation failure model for this system is established, thereby verifying the validity and practicability of the improved algorithm.

Keywords Fault information · FP-Growth algorithm · Correlation rules · Correlation failure model

Y. Wang
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing, China

Y. Wang · S. Wang (✉) · S. Lin
School of Traffic and Transportation, Beijing Jiaotong University,
Beijing, China
e-mail: 1156470358@qq.com

Y. Wang · S. Wang · S. Lin
Beijing Research Center of Urban Traffic Information Sensing
and Service Technologies, Beijing Jiaotong University, Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_14

14.1 Introduction

The equipment systems become more and more complicated with the rapid development of modern industry and science and technology, thereby strong correlations exist between different parts of systems, resulting in that single fault lead to abnormal running even functional failures. Meanwhile, a variety of fault records are produced inevitably, reflecting correlational relationships of failures indirectly. To do research on correlation failure relationships of complex equipment systems, we first should start from the mining of correlation rules about fault information. Correlation rules are a kind of descriptions of the interdependence and the relationships between objects, depicting their correlations in the database [1, 2]. Generating frequent itemsets is the key step of correlation rules mining [3].

Apriori [4] algorithm and FP-Growth [5] algorithm are two kinds of classic frequent itemsets mining algorithm. FP-Growth is about an order of magnitude faster than Apriori [6], but some defects still exist. So some researchers have proposed many improved algorithms based on FP-Growth [7–10]. Although they have some merits, the conditional pattern base still requires traversing for twice, making the execution efficiency greatly reduced and resulting in burden to database server.

Therefore, taking the fault information feature words, we improve the FP-Growth which used to mine frequent itemsets, and establish correlation failure models eventually. Paper's outline is as follows. Section 14.2 provides basic concepts of the correlation analysis. Section 14.3 describes the improved FP-Growth and discusses it. In Sect. 14.4, we construct the correlation failure model and in Sect. 14.5 a case study is proposed. Finally in Sect. 14.6 conclusions are exposed.

14.2 Basic Definitions

Based on the basic principles of the classical correlation analysis, the basic concepts and related knowledge can be described as follows [2, 11].

Definition 2.1 Let $I=\{\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, \ldots, \mathrm{I}_m\}$ be the universal item set and $S=\{\mathrm{T}_1, \mathrm{T}_2, \ldots, \mathrm{T}_n\}$ be a transaction database, where each $T_k\ (1 \leq k \leq n)$ is a transaction which is a set of items such that $T_k \subseteq I$. Every transaction has a number as TID. X is called an itemset if X is a set of items. Let X be an itemset. A transaction T is said to contain X if $X \subseteq T$. X is (k-itemset) if the number of elements in the X *is* k.

Definition 2.2 If itemset $X \subseteq I, Y \subseteq I$ and $X \cap Y = \Phi$, the formula like $X \Rightarrow Y$ is correlation rules. X is called the antecedent of the rules; Y is called the consequent of the rules, and it is a set of transactions including the item X, also the item Y. The formulas are:

$$\text{Support}(X \Rightarrow Y) = P(X \cup Y) \tag{14.1}$$

$$\text{Confidence}(X \Rightarrow Y) = P(Y/X) \tag{14.2}$$

Given a transaction database, strong correlations are correlations rules that satisfy the appointed minimum support threshold min Sup and minimum confidence threshold min Conf.

Definition 2.3 Given a min Sup, if the support of item X is equal or greater than min Sup, item X is a frequent itemset.

Definition 2.4 For every nonvoid subset s of itemset Z, it is indicated that $s \Rightarrow l - s$ is a strong correlation if meets $\text{Confidence}(s \Rightarrow l - s) \geq \min \text{Conf}$.

Definition 2.5 Confirming the affected items and affecting items in the correlation failure relationships: for the extracted correlation rules, there is a connection in the preceding and consequent item. The former one is affected item, and the latter is affecting item. Consequent failure can lead to failure of the preceding item.

14.3 The Improved FP-Growth Algorithm

14.3.1 The Improvement Ideas of FP-Growth Algorithm

Based on the classic FP-Growth, we add relational tables to avoid traversing the conditional pattern base for the first time. Improvement ideas are as follows:

(1) The transaction set S is traversed to obtain the frequency of each item. The items whose frequencies are below supports are eliminated. Then frequent itemsets L, an item table, and FP-tree are obtained. Then establish a two-dimensional table, and record support counts which any two items appear at the same time.
(2) The relational table records all of any two items' frequencies F that appear at the same time in the transaction database. The compounding items which F is below the given min Sup are deleted, and the simplified relational table is obtained.
(3) A FP-subtree in the condition of different suffix itemsets is established according to the above simplified relational table. Then all of the frequent itemsets in the condition of different suffix itemsets β is further obtained.

14.3.2 The Implementation of the Improved FP-Growth

In the process of frequent itemsets mining algorithms, we introduce the simplified relational table that will reduce the times of multiple scans. Specific implementation process is as follows:

Step 1: Establish item relational table. Make a count about the frequency that any two items appear at the same time in the transaction database and establish an item relational table that records the above frequency—it is called combined frequency. For a transaction: $\{ABCD\}$, we can obtain all the A two-dimensional vector: $(A, B), (A, C), (A, D), (B, C), (B, D), (C, D)$, and $(B, C) \neq (C, B)$.

Step 2: Establish the item simplified relational table. We need to delete some item relationships that two items' combined frequency is less than min Sup, and the new item relational table is obtained which is called the simplified relational table.

Step 3: Consider the item of minimum frequency. Item A's frequency is the smallest as it is in the end of the table. When the suffix itemset meets $\beta = \{A\}$, and the item's supports value about A: $\{(A, B) : 1 + 0; (A, C) : 2 + 0; (A, D) : 1 + 0; (A, E) : 1 + 0\}$; the supports set of related items is noted L_A, and $L\_A = \{B : 1 + 0; C : 2 + 0; D : 1 + 0; E : 1 + 0\}$; delete the items whose supports are less than 2, and the conditional item base is $\{(C : 2)\}$, so the frequency itemset is $\{\{C, A\} : 2\}$.

Step 4: When the suffix itemset meets $\beta = \{E\}$, and we can obtain all the item's support information about E in the first column of the simplified relational table: $\{(E, A) : 0 + 1; (E, B) : 0 + 3; (E, C) : 0 + 2; (E + D) : 0 + 0\}$; the support set of related items is $L\_E = \{A : 0 + 1, B : 0 + 3, C : 0 + 2\}$; delete the items whose supports are less than 2, we can obtain that conditional item base is $\{(B : 2), (C : 2)\}$, so the frequency itemset is $\{\{B, E\} : 2, \{C, E\} : 2\}$. It is similar when the suffix itemset meets.

14.3.3 Experimental Evaluation

In this section, we report the experiment results in which running time of the improved FP-Growth algorithm are compared with two state-of-the-art algorithms —Apriori and FP-Growth. Note that all these algorithms discover the same frequent itemsets under the same support, and the running time is the time that starts from reading data to generating all the frequent itemsets, not including the output time.

To test the algorithm performance on different environments with various data distributions, we used two datasets—T10I4D100K and Retail—in the experiments. All the tested algorithms are implemented in Java. All the experiments are performed on a computer with 4G memory and Intel Core @2.0 GHZ processor. Operating system is Windows 7 standard x64 Edition. Note that, the Apriori that we choose has proven to be the best algorithm among Apriori-like method.

We can conclude from Figs. 14.1 and 14.2 that the more the number of transaction, the higher the efficiency of algorithm. In Figs. 14.3 and 14.4, the efficiency of algorithms is higher with the increase of min Sup and the running

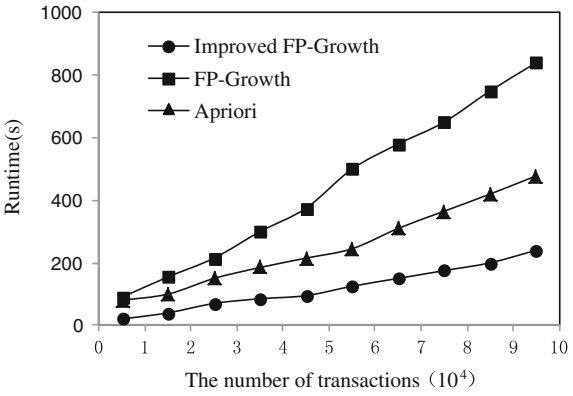


Fig. 14.1 Runtime on T10I4D100K for different number of transactions

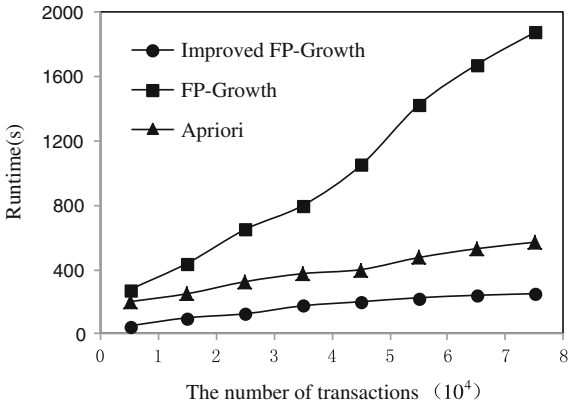


Fig. 14.2 Runtime on retail for different number of transactions

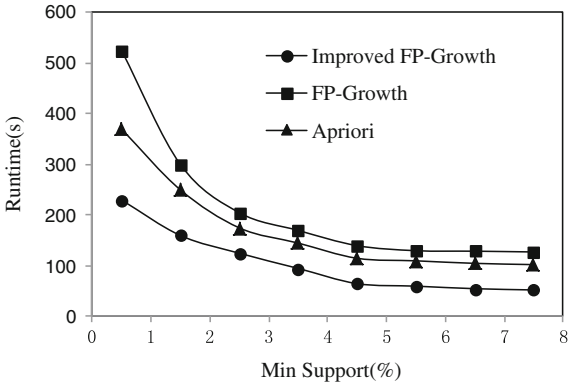


Fig. 14.3 Runtime on T10I4D100K for different min Sup

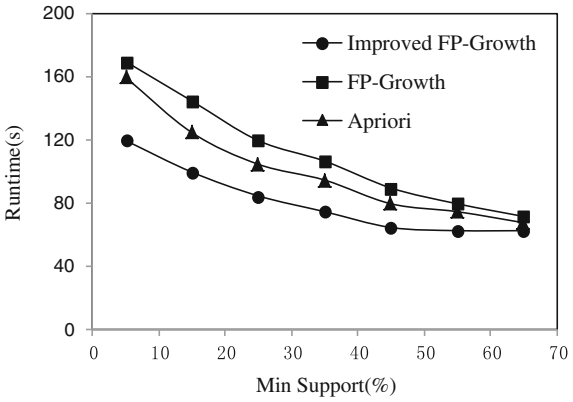


Fig. 14.4 Runtime on retail for different min Sup

time tends to stable while min Sup increases to a certain degree. As a result, the experimental results show the effectiveness and feasibility of the improved FP-Growth.

14.4 The Correlation Failure Modeling

14.4.1 The Extraction and Analysis of Correlation Failure Rules

In this part, we will mine for equipment correlation failure rules based on the improved FP-Growth algorithm. In Sect. 14.3.2, we have obtained all the frequent itemsets of every item in the transaction database S. Specific results are shown in Table 14.1. Here are steps that generate the correlation failure relationships.

Step 1: For any frequent itemset l, all the nonvoid subsets of l are produced;

Step 2: For every nonvoid subset s of l, there is a correlation failure relationship between s and $l - s$ if it meets $\text{Confidence}(s \Rightarrow l - s) \geq \text{minConf}$. And the failure of $l - s$ will cause failure of s.

According to Table 14.1, we can conclude the correlation failure relationships between items. Specific results are in Table 14.2.

Table 14.1 The frequent itemsets in the transaction database

Frequent itemsets	Support	*n*-frequent itemsets	Frequent itemsets	Support	*n*-frequent itemsets
A	2	1-frequent itemsets	*A*, *C*	2	2-frequent itemsets
B	3	1-frequent itemsets	*B*, *C*	2	2-frequent itemsets
C	3	1-frequent itemsets	*B*, *E*	3	2-frequent itemsets
E	3	1-frequent itemsets	*C*, *E*	2	2-frequent itemsets
			B, *C*, *E*	2	3-frequent itemsets

Table 14.2 Failure correlation table

Affected items	Affecting items
B	*E*, *C*
C	*A*, *E*, *B*
E	*B*, *C*
A	*C*
D	Φ

14.4.2 Correlation Failure Modeling

We carry out the correlation failure analysis of equipment items database to obtain the frequent itemsets by the improved FP-Growth algorithm, find the correlation failure relationships between items next and set up the correlation failure model according to these correlation failure relationships. Here are the specific steps:

Step 1: Establish a transaction database. A transaction database consists of the sets of equipment fault information feature words.

Step 2: Establish a library of frequent itemsets. We can find *i*-frequent itemset $(i = 1, 2, \ldots, n)$ by the improved FP-Growth algorithm, and in turn establish a library of frequent itemsets.

Step 3: Extract the correlation failure relationships of each item. The correlation failure relationships between the items are obtained according to the affected items and the affecting items.

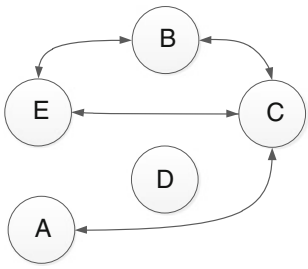


Fig. 14.5 Correlation failure model

Step 4: Establish a correlation failure model. We can obtain the correlation failure relationships of all the items according to the correlation failure relationships of each item, and in turn establish a correlation failure model. Here is the correlation failure model in Fig. 14.5.

14.5 Case Study

14.5.1 Failure Data Processing of the CRH2 EMUs Traction System

In this section, we analyze the traction system of CRH2 EMUs by the improved FP-Growth algorithm. After the preprocessing of fault data, we have found fault information feature words in the traction system and extracted more than 6000 fault data, then established more than 2000 transactions, further obtained a transaction database that consists of more than 2000 itemsets.

In order to simplify the correlation analysis for the transaction database, we need to encode these devices by letters. Encoding rules are shown in Table 14.3, and the

Table 14.3 Equipment codified definition

Device's name	Code	Device's name	Code
Four-quadrant rectifier	I_1	Main circuit breaker	I_9
Traction inverter	I_2	High-voltage isolator	I_{10}
Traction motor	I_3	Grounding switch	I_{11}
Auxiliary inverter	I_4	Grounding device for current	I_{12}
Battery charger	I_5	Current transformer	I_{13}
Storage battery	I_6	Pantograph	I_{14}
Main transformer	I_7	Passing over phase-splitter equipment	I_{15}
Arrester	I_8	Voltage transformer	I_{16}

Fig. 14.6 Part of the encoded transaction database

TID	Item sets
001	$I_{15}, I_{14}, I_{10}, I_9, I_7, I_8$
002	$I_{15}, I_{14}, I_{10}, I_9, I_7, I_1, I_3, I_2$
003	$I_{15}, I_{14}, I_{10}, I_9, I_7, I_8$
004	I_{15}, I_{14}, I_8
005	$I_{15}, I_{14}, I_{10}, I_{11}, I_{12}$
006	$I_{15}, I_{14}, I_{10}, I_9, I_7, I_1, I_4, I_5, I_6$
007	$I_{15}, I_{14}, I_{10}, I_9, I_7, I_1, I_2, I_3$
008	$I_{15}, I_{14}, I_{10}, I_9, I_7, I_8$
009	I_{14}, I_{10}, I_{13}
010	I_{10}, I_{13}
011	$I_{15}, I_{14}, I_{10}, I_9, I_7, I_1, I_4, I_5, I_6$

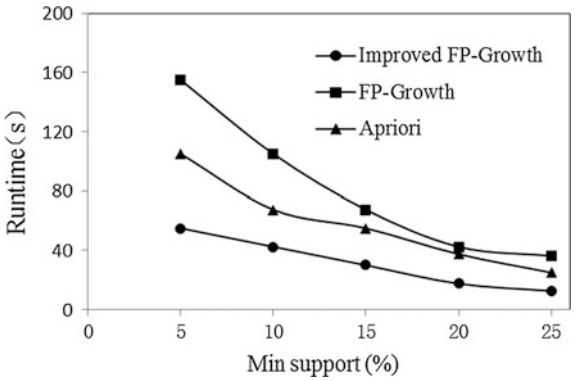


Fig. 14.7 Running time on different min Sup

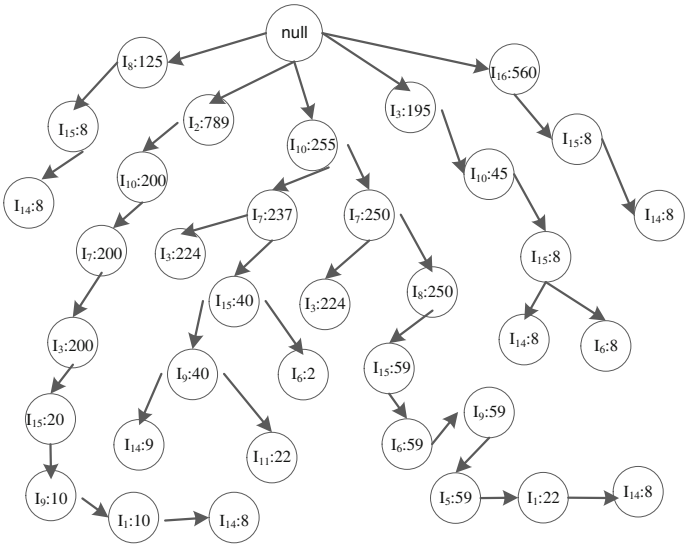


Fig. 14.8 FP-tree of traction system

new transaction database is obtained. Part results of the encoded transaction database are shown in Fig. 14.6.

We also adopt an experiment to evaluate the performance of the improved FP-Growth when feature words of fault information of CRH2 EMUs traction system is used as a dataset. Figure 14.7 shows the running time of all algorithms for different minimum support. The improved FP-Growth is obviously faster than Apriori and FP-Growth and has better performance in mining frequent itemsets.

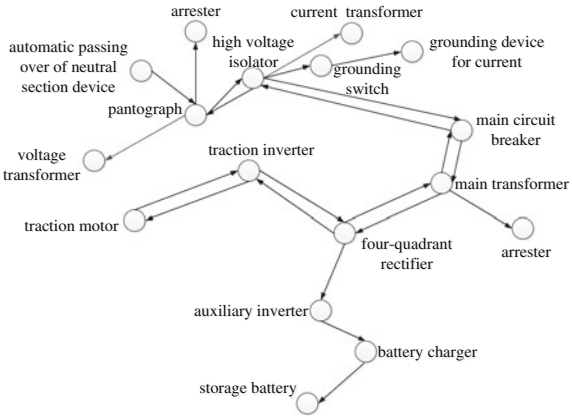


Fig. 14.9 The correlation failure model of traction system

14.5.2 Extracting Correlation Failure Rules of Traction System

We carry out the correlation analysis for the encoded transaction database of this traction system based on the improved FP-Growth algorithm. First, establish frequent pattern tree: FP-tree in Fig. 14.8. Second, traverse the FP-tree and excavate the frequent itemsets. We can find the affected and affecting items, and the correlation failure model of traction system is established. Results are shown in Fig. 14.9.

14.6 Conclusions

In this paper, the improved FP-growth algorithm is proposed by adding support count of two-dimensional tables—simplified relational tables. The experiments show that it is actually efficient for mining frequent patterns. In this way we conduct the correlation analysis of equipment fault information feature words, a correlation failure model is established and it provides a theoretical basis for further research. However, complex high-speed train system has complicated structures and has a large amount of fault information feature words, and this leads to the decrease of the efficiency for extracting correlation failure rules. Determining how to improve the extraction efficiency of correlation failure rules for large systems becomes the focus of further research.

Acknowledgments The project is supported by the independent subject (Grant No. I14K00451) and National High Technology Research and Development Program of China (863 Program) (Grant No. T13B200011).

References

1. Ordonez C, Zhao K (2011) Evaluating association rules and decision trees to predict multiple target attributes. Intell Data Anal 15(2):173–192
2. Chi X, Fang ZW (2011) Review of association rule mining algorithm in data mining. In: IEEE 3rd international conference on communication software and networks (ICCSN). IEEE, pp 512–516
3. Shen B (2011) In: Research on the technologies of association rules. Zhejiang University Publisher, Zhejiang, pp 3–12 (in Chinese)
4. Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. ACM sigmod record. ACM 22(2):207–216
5. Han J (2000) Mining frequent patterns without candidate generation. In: Proceedings of the international conference on management of data, dallas, 2000 ACM-SIGMOD. ACM, pp 1–12
6. Singh AK, Kumar A, Maurya AK (2014) An empirical analysis and comparison of apriori and FP-growth algorithm for frequent pattern mining. In: International conference on advanced communication control and computing technologies (ICACCCT). IEEE, pp 1599–1602
7. Li YB, Tang H, Zhang C et al (2011) Frequent pattern mining algorithm based on improved FP-tree. J Comput Appl 31(1):101–105 in Chinese
8. Li L, Zhai D, Jin F (2003) A parallel algorithm for frequent itemsets mining. J China Railway Soc 25(6):71–75 (in Chinese)
9. Chen XS, Zhang S, Tong H et al (2014) FP-Growth algorithm based on the boolean matrix and mapreduce. J South China Univ Technol, vol 1
10. Taktak W, Slimani Y (2014) MS-FP-growth: a multi-support version of FP-growth algorithm. Int J Hybrid Inf Technol 7(3):155–156
11. Shao F, Yu Z, Wang J et al (2009) Data mining principle and algorithm, 2nd edn. Science Press, Beijing, pp 8–23 (in Chinese)

Chapter 15
The Research and Realization of Contactless Collision Early Warning Assistant System

Yue Qu, Baohua Wang, Ruichang Qiu and Ce Zheng

Abstract With the progressive achievement of China's modernization and the urban rail transport development, rail transport has entered a rapid developing period, which makes the safety of trains more and more important. Because the traditional way, such as the telephone blocking mode, is difficult to complete the ATP automatic protection system. The aim of this paper is using wireless transmission technology based on the chirp spread spectrum signal (CSS) and the symmetric bilateral ranging method (SDS-TWR) to achieve the contactless train collision warning system that removes ATP. There is also an explanation about the hardware and software of the system.

Keywords Anti-collision warning · Ranging · Rail transportation

15.1 Introduction

With the development of the national economy and the growing urban population, the urban traffic faced more and more challenge. Subway has many advantages such as large capacity, high safety factor, and low pollution, which make the state to grow urban rail transit and metro has become a popular term in rail transport sector [1]. There are 22 cities which have been built the subway, about 101 urban rail lines totally 3155 km by December 31, 2014 in China [2]. The total length of Shanghai rail transportation network operator is 548 km. With the great development of transportation, there are lots of serious accidents results from artificial operation or locomotive trouble, which makes the locomotive collision a major issue to hinder the development of subway, thus collision warning system has become a hot topic [3].

Y. Qu (✉) · B. Wang · R. Qiu · C. Zheng
Beijing Jiaotong University (BJTU), Beijing, China
e-mail: 14121452@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_15

15.2 The Communication Design of Contactless Collision Early Warning Assistant System

Through the investigation of domestic and foreign literature, this paper design a system based on contactless wireless ranging collision warning system, whose principle relies on the wireless transmission between the two measuring points. The function of this system is to calculate the distance above two points and collect useful signal. When the host module receives the useful signal, this system read it and protects the safety of trains.

The system selected 2.4G antenna technology and its frequency is between 2.405 GHz and 2.485 GHz that this band is a general international wireless spectrum for industrial, scientific, and medical [4]. The second reason of choosing the band is the antenna interference is small.

When using linear frequency modulation techniques base on spread-spectrum signal, these signals are sinusoidal wave in a certain range, the frequency of which signal is shown as follows:

$$S(t) = W(t)\sin\left[2\pi\left(f_0 + \frac{k}{2}t\right)t\right] \tag{15.1.1}$$

The system produces a CSS-based wireless technology ranging signals rely on linear chirp based on spread spectrum signal and using symmetrical bilateral two-way ranging method. This technology is based on the principle of setting spread spectrum frequency modulation signal technology as the carrier, receiving a communication by two points and calculating the distance between two points by SDS-TWR algorithm [5].

Symmetric double-sided two-way ranging (SDS-TWR) distance measurement method is a symmetrical bilateral two-way ranging method and proposes a radio ranging method base on IEEE8021.15.4a standard [5]. The basic principle is using wireless communication propagation delay to measure the distance between two points. The measurement process is shown in Fig. 15.1.

On the map, the data sender sends a ranging packet, and the data receiver receives the packet when it is T_2, and then the data receiver return the packet. After T_1, the data sender receives the packet. In order to increase the accuracy and to get the value of T_2, try this measurement again and get T_3 and T_4. The electromagnetic wave propagation speed is c, therefore the formula for the distance between two points is shown as follows:

$$d = c[(T_1 + T_2) + (T_3 + T_4)]/4 \tag{15.1.2}$$

After the above analysis, the system chooses Germany nanotron public production nanoPAN 5375 to achieve precise ranging capabilities, and uses a master chip to initialize nanoPAN 5375, process SDS-TWR packet information and pass ranging information between the system [6].

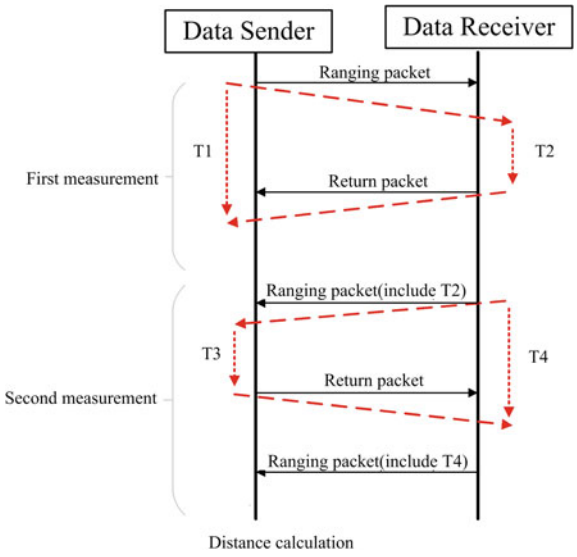


Fig. 15.1 The process of SDS-TWR ranging

15.3 The Structural Design of Contactless Collision Early Warning Assistant System

The contactless collision warning system model is shown in Fig. 15.2.

The basic principle of the early warning system is shown as follows: The wireless location technology uses wireless ranging method to measure the distance of the train, and then determines whether there is a train collision risk. When ATP is cut, the train judge the running direction and send upward, downward, forward, and backward information to the system; the active ranging module in locomotive sends a packet ranging; on the other hand, the passive ranging in tail of the train maintains

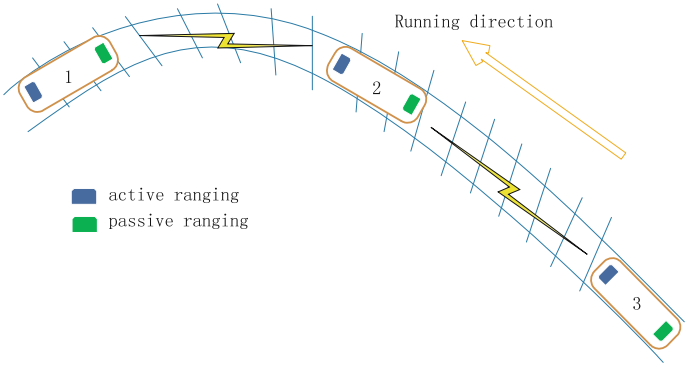


Fig. 15.2 System working schematic

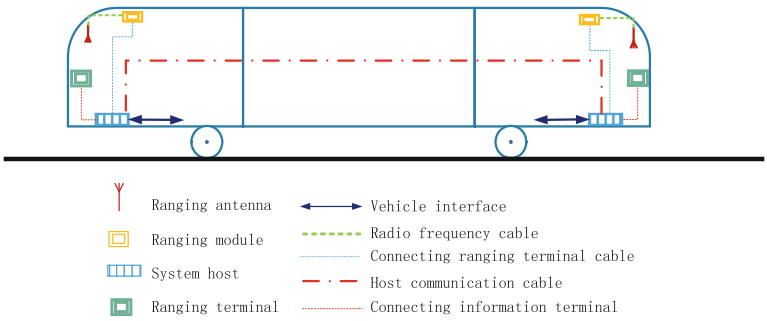


Fig. 15.3 System block diagram

monitoring of a packet ranging that is sent by other train. Once the passive ranging receives the packet ranging, the passive ranging module sends information to the host module. The host module receives the information and judges the situation of the train, and then makes a lot of actions to protect the train. The whole system structure is shown in Fig. 15.3.

Because the train can change its locomotive and tail, the ranging module installed in locomotive needs to send information to the ranging module installed in tail of train. So the host module can receive those information and distinguish which side is locomotive or tail. Only in this way, the system can realize a change between the send module and the receive module in order to achieve the conversion of the front and the rear of trains.

Trains were installed in the host module at the front and rear, and the host connects the ranging terminal that connects the antenna ranging 2.4G. All about form the entire car facilities.

15.4 The Hardware Design of Contactless Collision Early Warning Assistant System

The train collision avoidance system is divided into three parts: the distance module, information terminal module, and host module.

The host module contains power the conversion circuit, input signal module, and core control module. The main chip uses STM32 as core chip, which is a high-performance, low-power, and low-cost core chip. Its internal resources include: 28 kB FLASH program memory, 32 kB of SRAM memory, two SPI, 51 GPIO port, 80 % of the general purpose IO, and so on [7, 8]. The main logic diagram of the host module is shown in Fig. 15.4.

The host module controls circuit by STM32 main chip control, which can control the input signal of train, emergency braking function. The ranging module uses the German nanoPAN 5375 and 2.4G antenna RF module. The information

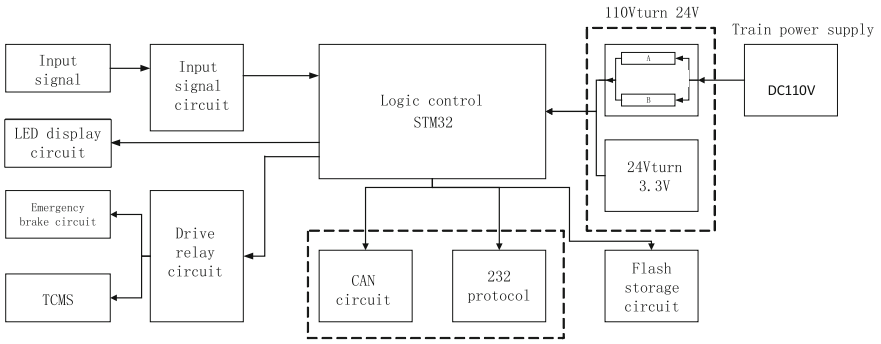


Fig. 15.4 Overall circuit diagram of the system host

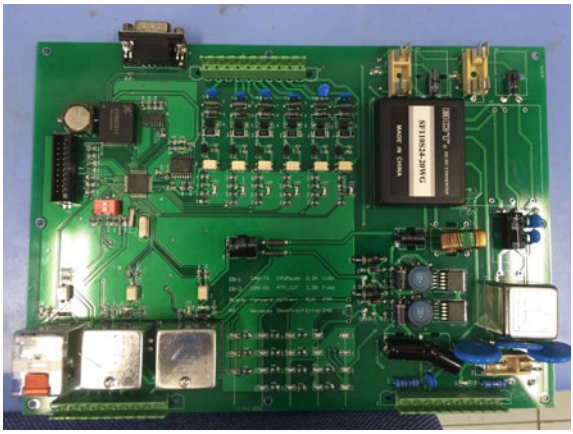


Fig. 15.5 The host board of the collision assistant system for train

terminal module receives information from the host module and the host module displays the information in the LED screen and makes audible alarm when two trains are too close. The PCB of the module is shown in Fig. 15.5.

15.5 The Software Design of Contactless Collision Early Warning Assistant System

The contactless collision warning system software design includes above three modules, and these three parts use ARM chips, so we select C language to embed. The host module not only carries the hardware initialization, but also reacts to the corresponding information. According to the hardware condition, the structure of the system is shown in Fig. 15.6.

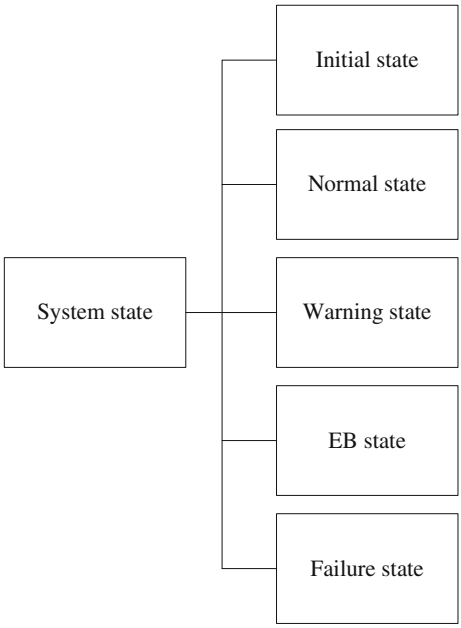


Fig. 15.6 System logic state

The initial state of the system state is to initialize the system, and warning state is working when the distance between trains is approaching 300 m. When the distance of two trains is less than 300 m, EB state works and the train sends an emergency brake. When Failure state is working, the train disconnects all EB contact and triggers the emergency brake.

The ranging module and the information terminal module finish the communication through serial port 232 protocol. The information terminal module will receive the data and sent them to the host, and the host judges the state of the train making the corresponding response to finish the system design.

15.6 Function Test

The system is complete and makes functional tests. The functional tests are divided into static test and dynamic test. The static test is repeatedly measuring the distance between two stationary points. The actual distance of two stationary points is 58 m, and there is the comparison of the measured data with the actual data as is shown in Fig. 15.7.

Additionally, the static test measures 166.9 and 210.5 m of two stationary points, the data is shown in Fig. 15.8.

Dynamic test simulates a straight choice open lane, whose length is 678 m. The speed is 50 and 40 km/h, the measurement results are shown in Table 15.1.

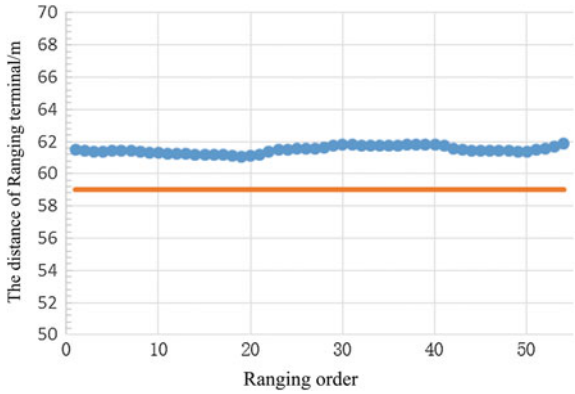


Fig. 15.7 Ranging results at fixed 58.8 m

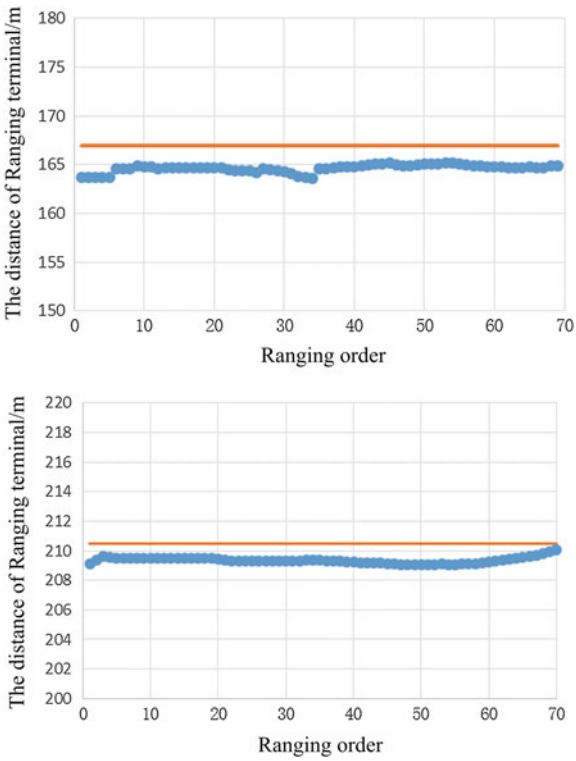


Fig. 15.8 Ranging results at fixed 166.9 and 210.5 m

From the static and dynamic test can be drawn: The ranging system can measure actual distance in the urban rail train, and the ranging distance not less than 300 m in less than 10 % ranging error. Those data show that the train can emergency brake early through this system when facing emergency situation. Thus can satisfy design requirements and effectively reduce the occurrence of accidents.

Table 15.1 Dynamic functionality experiment results at open line

Measurement speed(km/h)	Main location	From location	Response distance (m)	Actual distance (m)	Accuracy (m)
50	Before	After	390	383 ± 10	7 ± 10
50	After	Before	390	382 ± 10	6 ± 10
40	Before	After	371	377 ± 10	6 ± 10
40	After	Before	371	378 ± 10	7 ± 10

15.7 Summary

In order to strengthen the contactless train collision warning function, this paper find out the contactless collision warning system. This system choose chirp spread based on spectrum signal technology and symmetric bilateral two-way ranging method. Relying on this basis, the train meets the functional requirements of collision warning generally.

References

1. Li C (2014) Research about the development status of China subway and the future prospects. Sci J 14(7):121 (in Chinese)
2. Nan Y (2015) China's urban system in China: the rational layout need to be established for the accelerated development. China produced by the news (BO4) (in Chinese)
3. Yong C, Yang S (2006) Research and development of collision avoidance warning system for automobile. Comput Simul 23(12):239–243 (in Chinese)
4. Xia Y (2010) Mobile payment scheme based on 2.4G technology research. Changsha University of Science and Technology (in Chinese)
5. IEEE Computer Society (2007) IEEE Std 802.15.4a—2007. New York, USA, pp 121–138
6. Zheng C (2015) Research and implementation of the warning system for collision avoidance for train. Beijing Jiaotong University (in Chinese)
7. Zhang Y (2014) Design and implementation of ground testing equipment based on STM32. North Central University (in Chinese)
8. Liu L, Cheng L (2005) Study of selection embedded system hardware platform. Telecommun Eng Stand Periodicals 2:51–54 (in Chinese)

Chapter 16
Fault Decoupling Method of Railway Door System Based on the Extended Petri Net

Qi An, Yong Qin, Yunxiao Fu, Xiaoqing Cheng, Limin Jia, Shaohuang Pang and Jianlong Ding

Abstract Door system plays an important role in urban railway door system. Fault decoupling analysis and research of railway door system are of great significance for door system performance. This paper puts forward the fault decoupling method of compound causal chain based on the extended Petri net so as to locate and identify the compound fault in fuzzy discrete event systems (FDES) quickly and accurately. With the basis of ODDT analysis, the decoupling method is used to decouple the complex relations hierarchically in the model and clarify the fuzzy–casual relationships. Consequently, the difficulty of solving problems reduced; system state recognition and behavior analysis can be done at a lower cost.

Keywords Door system · Compound fault · Petri net · Causal chain · Decoupling method

16.1 Introduction

With the development in industrial technology, urban railway train system has gradually become large, its structure becoming more complicated and coupling degree of system components deepening constantly. Door system plays an important role in urban railway train system. The fault of door system severely threatens train operation safety. As a result, decoupling the fault chin effect of door system, together with finding the fault location, will improve the whole performance of door system. Compound fault [1] means two or more faults occurring simultaneously when the system has a partial or overall failure. Compound fault usually leads to fault casual chains mix. The following representations become more complicated, making the

Q. An · Y. Qin (✉) · Y. Fu · X. Cheng · L. Jia · S. Pang · J. Ding
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, No. 3 Shangyuancun, Haidian, Beijing, China
e-mail: qinyong2146@126.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_16

diagnosis more difficult. Fuzzy discrete event systems (FDES) [2] offer an effective modeling method for compound fault diagnosis.

There are some common methods of system fault analysis: fault tree analysis, failure mode effects and fatality analysis, Petri net, etc. The fault analysis with Petri net solves the fault correlation diversity problem which is always shown in traditional fault analysis. Carl Adam Petri, a Germany scientist, proposed Petri net for the first time. Foreign scholars started Petri net study earlier and they have done more research on it. Viswanadham N. and some scholars came up with a monitoring system based on hierarchical Petri net [3], which was used to implement fault diagnosis. Lo K.L. and some experts applied the extended Petri net into online fault diagnosis of transformer substation establishing fault relation model of complex system [4].

This paper puts forward fault decoupling method of compound causal chain based on the extended Petri net, which decouples system model hierarchically and establishes a less complex solving model. Therefore, the calculated load can be reduced in the complicated system with multiple relationship coupling. Compared with common decoupling methods of compound casual chain, the method in this paper has following characteristics: with the analysis of observable degree in dimensionality of time (ODDT), decoupling complicated relations in the model hierarchically, decreasing the solving difficulties, and doing system state recognition and behavior analysis at a lower cost. The decoupling model analysis shows that the model after decoupling has no difference in solving problems with the original model, which means decoupling process is lossless [5].

16.2 Casual Chain Analysis of Compound Fault

16.2.1 Extended Time Petri Net

If time intervals $tm_i = [a, b]$ and $tm_j = [c, d]$, a and c are the lower bounds of tm_i and tm_j, respectively, while b and d are the upper bounds of tm_i and tm_j, respectively. Σ_F represents a FDES. $e \in E$ is a fuzzy event, state identification $M_i[e > M_j$. State transition can be done in time interval tm_{ij}, with time constrain, which is denoted by $M_i[e \triangleright tm_{ij} > M_j$.

Compound fault casual chain model based on temporal casual relations is established by extended time Petri net (ETPN) [6–10]. ETPN is an eight tuple $\sum_{ET} = (S', T'; F', E', I', \delta', \tau', M_0')$, in which,

1. $(S', T'; F')$ is a prototype Petri net, with $\forall s \in S{:}\{s\} \cap \overrightarrow{s^{\bullet}} = \varnothing$, $\forall t \in T'{:}|^{\bullet}t| \leq 1$;
2. E' is the finite set of events;
3. $I' \subset F'$ is the finite set of inhibitor arcs;
4. $\delta'{:}E' \rightarrow 2^{|T'|}$ is event map in transition subset;
5. $\tau'{:}T' \rightarrow R_0 \times (R_0 \cup \{\infty\})$, $\tau'(t_i)$ is time delay interval related to transition t_i;
6. M_0' is the original state identification of Σ_{ET} and $M'(s_i) \in R_0 \times R_0$ is the token of s_i.

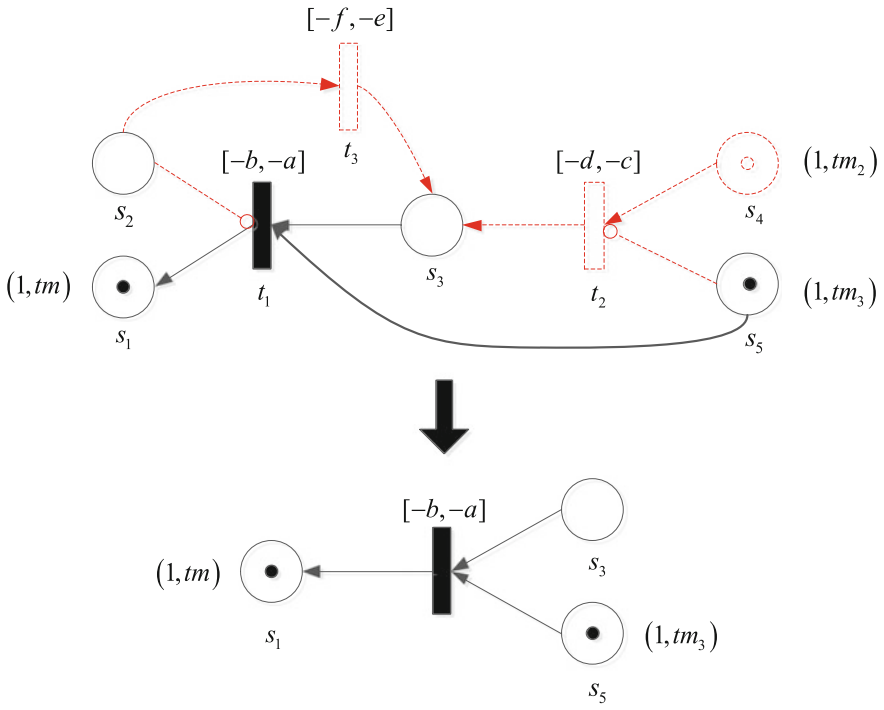


Fig. 16.1 TC-PPN reduction

16.2.2 Possibility Petri Net with Time Constraints

If Σ_{TP} is possibility Petri net with time constraints (TC-PPN), fuzzy state identification of Σ_{TP} is $M = [M(s_1), M(s_2), \ldots, M(s_n)]^T$, in which $n = |S|$.

Through ETPN model inversion, TC-PPN can be obtained for correct decoupling, which can eliminate the impacts of virtual event and timing error. Based on inhibitor arcs and original state, model reduction can be done in TC-PPN. We will get TC-PPN after reduction.

Figure 16.1 is the reduction process of TC-PPN, in which the dotted line elements can be deleted in the original net.

In Fig. 16.1, inhibitor arcs are represented by lines with circles in the tail.

16.2.3 Decoupling Net Construction

If Σ_{TP} is a FDES, time constraint of Σ_{TP} is defined as a three tuple $TG(\Sigma_{TP}) = (V, A_{rc}, H)$, in which,

1. $V = \{V_i | i = 1, 2, \ldots, n\}$, $V_i = \{(a, b) | a, b \in R : a \leq b\}$ together with $s_i \in S$ are time constraint, which is called vertex set $\mathrm{TG}(\Sigma_{\mathrm{TP}})$;
2. $A_{\mathrm{rc}} \subseteq F$ is the arc set of $\mathrm{TG}(\Sigma_{\mathrm{TP}})$;
3. $H : A_{\mathrm{rc}} \rightarrow 2^{|T|}$ is the mark of arc.

The time constraint graph of TC-PPN is established, which is used to calculate ODDT with time constraint:

If Σ_{TP} is a TC-PPN and $\mathrm{TG}(\Sigma_{\mathrm{TP}}) = (V, A_{\mathrm{rc}}, H)$ is the time constraint graph, ODDT of event $e \in E$ is:

$$d_e = (d_e^o, d_e^{uo}) \tag{16.2.1}$$

$$d_e^o = \frac{\sum_{t \in \delta(e)} \mu}{|\delta(e)| - |\varepsilon|} \tag{16.2.2}$$

$$d_e^o + d_e^{uo} = 1 \tag{16.2.3}$$

In which, $|\varepsilon|$ represents the number of $\mu = \varepsilon$ in $t \in \delta(e)$. The calculation method of μ is:

1. If $s_i \in {}^\bullet t, s_j \in t^\bullet : |V_i| > 1 \vee |V_j| > 1$, then,

$$\mu = \varphi\big(V_i, V_j, \tau(t)\big) \tag{16.2.4}$$

$$\varphi = \frac{\sum_{v_i \in V_i, v_j \in V_j, v_i + \tau(t) \neq v_j} \kappa\big(v_i, v_j, \tau(t)\big)}{|V_i| \bullet |V_j| - |v_i + \tau(t) \equiv v_j|} \tag{16.2.5}$$

   In which, $|v_i + \tau(t) \equiv v_j|$ is the time interval number of satisfying $v_i + \tau(t) = v_j$ and κ is calculated by Eq. (16.2.1). φ is non-self-voting function;
2. If $s_i \in {}^\bullet t, s_j \in t^\bullet : V_i \notin V \wedge V_j \notin V$, then $\mu = 0$;
3. If $s_i \in {}^\bullet t, s_j \in t^\bullet : |V_i| = 1 \wedge |V_j| = 1$, then $\mu = \varepsilon$, which means unknown state out of d_e^o calculation.

Set $\sum_{\mathrm{TP}} = (S, T; F, E, I, \delta, \tau, M_0)$ as a FDES with multiple casual coupling. Then, $D^T = \{d_e^T | e \in E\}$ is an ODDT set of all events in Σ_{TP}. With ODDT, multiple casual decoupling algorithm of FDES is established and used to do decoupling. The result is type I decoupling net in the first step.

Search leakage library in type I decoupling net. If the leakage library is not the source library of original ETPN, type II decoupling net needs to be constructed.

Σ_{I} is given as type I decoupling net of Σ_{TP} and the reachable state mark set is R_{I}. Σ_{I} only has one state mark $M_l = (M_{\sigma - l}, M_{\upsilon - l})$, which satisfies $\forall M_\upsilon = (M_\sigma, M_\upsilon)$, $M_\upsilon \in R_{\mathrm{I}} : M_{\sigma - l}(s) = \max\{M_\sigma(s)\} \wedge M_\upsilon(s) \subseteq M_{\upsilon - l}(s)$ M_l is the final state of Σ_{I}.

Σ_{I} is given as type I decoupling net of Σ_{TP}. R and R_{I} are the reachable state mark set of Σ_{TP} and Σ_{I}, respectively. Σ_{TP} projection in Σ_{I} is $P(R)$. And then $P(R) = R_{\mathrm{I}}$.

Σ_{I} is given as type I decoupling net of Σ_{TP}, $\sum_{\mathrm{TP}} = (S, T; F, E, I, \delta, \tau, M_0)$, $\sum_{\mathrm{I}} = (S_{\mathrm{I}}, T_{\mathrm{I}}; F_{\mathrm{I}}, E_{\mathrm{I}}, I_{\mathrm{I}}, \delta, \tau, M_0^{\mathrm{I}})$. M_l and M_l^{I} are given as the final state of Σ_{TP} and Σ_{I}, $S_{\mathrm{I}-in} = \{s | s \in S_{\mathrm{I}} \wedge s^{\bullet} = \varnothing\}$, $S_{in} = \{s | s \in S \wedge s^{\bullet} = \varnothing\}$. If $S_{\mathrm{int}} = S_{in} \cap S_{\mathrm{I}-in} \wedge S_{\mathrm{int}} \neq \varnothing$, as to $\forall s \in S_{\mathrm{int}}$, $M_l(s) = M_l^{\mathrm{I}}(s)$.

Set $\sum_{\mathrm{TP}} = (S, T; F, E, I, \delta, \tau, M_0)$, $\sum_{\mathrm{I}} = (S_{\mathrm{I}}, T_{\mathrm{I}}; F_{\mathrm{I}}, E_{\mathrm{I}}, I_{\mathrm{I}}, \delta, \tau, M_0^{\mathrm{I}})$, in which Σ_{II} is given as type II decoupling net of Σ_{TP}. R_{II} is the reachable state mark set of Σ_{II}. Σ_{II} has only one state mark $M_l = (M_{\sigma-l}, M_{\upsilon-l})$, satisfying $\forall M_{\upsilon} = (M_{\sigma}, M_{\upsilon})$, $M_{\upsilon} \in R_{\mathrm{I}}$:$M_{\sigma-l}(s) = \max\{M_{\sigma}(s)\} \wedge M_{\upsilon}(s) \subseteq M_{\upsilon-l}(s)$. M_l is the final state of Σ_{II}.

Σ_{II} is the reduction result of Σ_{I}, which means all the elements in Σ_{II} are subsets of Σ_{I} elements. Thereby,

$$\sum_{\mathrm{TP}} = (S, T; F, E, I, \delta, \tau, M_0) \tag{16.2.6}$$

$$\Sigma_{\mathrm{II}} = (S_{\mathrm{II}}, T_{\mathrm{II}}; F_{\mathrm{II}}, I_{\mathrm{II}}, E_{\mathrm{II}}, \delta, \tau, M_0^{\mathrm{II}}) \tag{16.2.7}$$

M_l and M_l^{II} are given as the final state of Σ_{TP} and Σ_{II}, respectively, $S_{\mathrm{II}-in} = \{s | s \in S_{\mathrm{II}} \wedge s^{\bullet} = \varnothing\}$, $S_{in} = \{s | s \in S \wedge s^{\bullet} = \varnothing\}$. If $S_{\mathrm{int}} = S_{in} \cap S_{\mathrm{II}-in} \wedge S_{\mathrm{int}} \neq \varnothing$, as to $\forall s \in S_{\mathrm{int}}$, $M_l(s) = M_l^{\mathrm{II}}(s)$.

This inference illustrates the final state in accordance with the original model which can be obtained by recalling the reduction model based on time interval and conditional probability. In the information system with complex causalities, such as backward inference, fault analysis, and diagnosis, analysis and reduction with this method will lead to a result consistent with the original model at a low computing cost.

16.3 Simulation Experiment

16.3.1 Door System Analysis

The door-opening process model of a door system model can be parted into three layers: electrical control layer, EDCU control layer and mechanical action layer. Electrical control layer includes relays, electrical loops, power supply, and electrical machines, and gives out signals of opening the door and driving the motor. EDCU control layer consists of EDCU units and their peripheral circuits give out signals of door movement velocity which is interactively controlled by EDCU and the electrical machine. The mechanical action layer includes the door and the connected mechanical parts (masts, screw rods, guide rails, and so on), with the function of accomplishing the movements of mechanical parts of the door. The three subnets correlate and interact with each other controlling the movement process of door system jointly.

16.3.2 The Decoupling Method of Opening-Door Fault

In order to test whether this method is effective or not, a decoupling deduce is made. This simulation experiment follows the opening-door process of rail transit vehicles. According to the recorded frequency of door fault, the input of compound fault during the opening-door process is set as the faults of screw nut and screw rod unlocking, masts, electrical machines, brakes, and popping of backing pins.

In the opening-door process carried out by the door system model in Fig. 16.2, the monitoring system will capture the following information (the unit is ms): $O(S_{38}) = 2951$, $O(S_{35}) = 2892$, $O(S_{28}) = 2243$, $O(S_{25}) = 1145$, $O(S_{19}) = 1077$, among which $O(x)$ refers to the time when x is captured for the first time. It aims to get the possible faulted running equipment and their triggered event chain.

Probability data of some nodes used in the experiment are given in the following Table 16.1.

Based on the above parameters and analysis, along with the causal chain of running equipment, an ETPN model is shown in the following Fig. 16.2.

$\sum_{ET} = (S', T'; F', E', I', \delta', \tau', M_0')$ is given as ETPN in Fig. 16.2, among them,

1. S' is a collection of libraries. Implications of each library can be found in Table 16.2. $T' = \{t_1, \ldots, t_{39}\}$, F' is shown as in the Fig.
2. $E' = \{e_1, \ldots, e_{39}\}$;
3. $I' = \{(S_{22}, t_{15})\}$;
4. $\delta'(e_1) = \{t_1\}$, $\delta'(e_2) = \{t_2\}, \delta'(e_3) = \{t_3\}, \ldots, \delta'(e_{39}) = \{t_{39}\}$;
5. $\tau(t_{15}) = \tau(t_{17}) = \tau(t_{18}) = \tau(t_{19}) = \tau(t_{22}) = \tau(t_{26}) = \tau(t_{28}) = \tau(t_{30}) = [10, 40]$, $\tau(t_4) = \tau(t_9) = \tau(t_{11}) = \tau(t_{12}) = \tau(t_{14}) = \tau(t_{32}) = \tau(t_{34}) = \tau(t_{37}) = [310, 340]$,

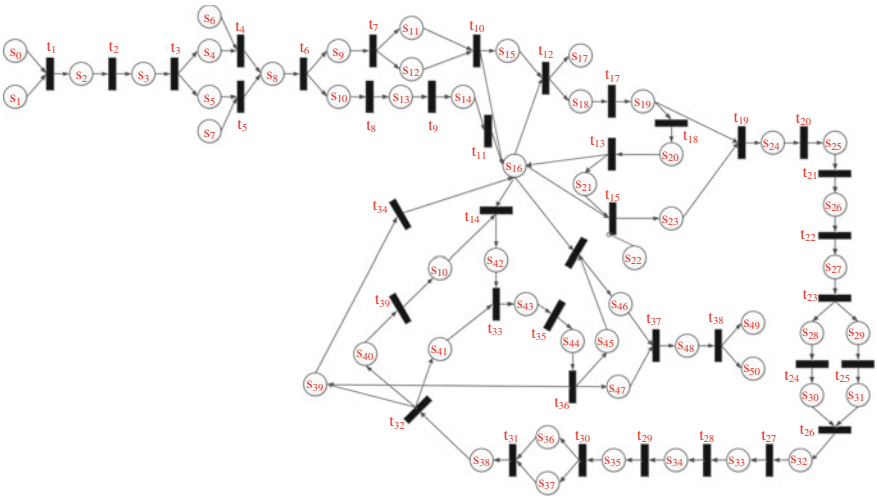


Fig. 16.2 ETPN model of the door system in the opening-door process

Table 16.1 Probability data of some nodes

Node	Probability data	Node	Probability data	Node	Probability data
1	0.0005	11	0.0015	21	0.0034
2	0.0010	12	Y0.0033	22	0.0028
3	0.0001	13	0.0010	23	0.0002
4	0.0002	14	(Y)0.0143	24	0.7793
5	Y0.0011	15	0.0001	25	0.2207
6	Y0.0042	16	(Y)0.0001	26	0.0010
7	0.0064	17	0.0001	27	0.0001
8	(Y)0.0220	18	0.0001	28	0.0001
9	0.0167	19	Y0.0046	29	(Y)0.0723
10	Y0.0029	20	0.0004	30	0.1187

$\tau(t_5) = \tau(t_6) = \tau(t_{21}) = \tau(t_{23}) = \tau(t_{24}) = \tau(t_{25}) = \tau(t_{39}) = [510, 540]$,
$\tau(t_1) = \tau(t_3) = \tau(t_7) = \tau(t_{16}) = \tau(t_{33}) = \tau(t_{35}) = \tau(t_{39}) = [110, 140]$,
$\tau(t_2) = \tau(t_8) = \tau(t_{10}) = \tau(t_{13}) = \tau(t_{20}) = \tau(t_{27}) = \tau(t_{29}) = \tau(t_{31}) = \tau(t_{36}) = [20, 40]$;

6. $M_0(S_{38}) = [2951, 2951]$, $M_0(S_{35}) = [2892, 2892]$, $M_0(S_{28}) = [2243, 2243]$, $M_0(S_{28}) = [2243, 2243]$, $M_0(S_{19}) = [1077, 1077]$, $M_0'(\text{else}) = \varepsilon$.

Inverse Σ_{ET} and $\Sigma_{TP}' = (S, T; F, E, I, \delta, \tau, M_0)$ is drawn. Moreover, $\Sigma_{TP} = (S, T; F, E, I, \delta, \tau, M_0)$ is got when Σ_{TP}' is reduced, as shown in Fig. 16.3.

Making the extension degree of time interval $\lambda = 0.5$, the ODDT of each event is calculated by the time constraint derived from TG: $d^o_{e_{31}} = \varepsilon$, $d^o_{e_{30}} = 1$, $d^o_{e_{29}} = 0$, $d^o_{e_{28}} = 0$, $d^o_{e_{27}} = 0$, $d^o_{e_{26}} = 0$, $d^o_{e_{24}} = 0.935$, $d^o_{e_{22}} = 0$, $d^o_{e_{21}} = 0.841$, $d^o_{e_{20}} = 0$, $d^o_{e_{19}} = 0.725$, $d^o_{e_{15}} = 0$, $d^o_{e_{13}} = 0$, $d^o_{e_{18}} = 0.5$, $d^o_{e_{17}} = 0$, $d^o_{e_{25}} = 0$, $d^o_{e_{23}} = 0$. Create type I decoupling net of Σ_{TP} based on ODDT of events is given in Fig. 16.4.

Construction of type II decoupling net of Σ_{TP}, Σ_{II}, based on type I decoupling net Σ_I is shown in the following Fig. 16.5.

From 16.2.3, if the final condition of Σ_{II} is obtained, the conclusion that it is in accordance with Σ_{TP} under the condition of ODDT may be drawn. Thus, in this example, possible faulted running equipment and their triggered event chains can be obtained based on the analysis of Σ_{II}.

Create a whole working model functioning in door opening created in the system emulator. In this model, if the inhibitor arcs are added to nodes where faults can be tested, possible faults are expressed. When several faults occur at the same time, the most likely casual chain can be decoupled based on faults tested. The fault chain deduced by simulation environment is shown in Fig. 16.6.

Table 16.2 Implications of each library in the ETPN model

Marking	Implication	Marking	Implication	Marking	Implication
S_0	Zero-speed signal	S_{18}	Driving signals of electrical machines	S_{36}	Upper guide rail
S_1	Stop-in-place signal	S_{19}	Electrical machines work	S_{37}	Lower guide rail
S_2	ATC	S_{20}	Locking devices turn on	S_{38}	Door opens to 85 % of the maximum
S_3	Door enabling signal	S_{21}	Minor control library 2	S_{39}	Position sensor of the door
S_4	HMI	S_{22}	S_2 moves	S_{40}	Backing pins on top of the door work
S_5	ATO	S_{23}	Door movements allowed	S_{41}	Minor control library 3
S_6	ATO switch not on automatic catch	S_{24}	Unlocking of electrical machines and screw-nut pairs	S_{42}	Minor control library 7
S_7	ATO switch on automatic catch	S_{25}	Screw nut locking devices quit the LS locking segment	S_{43}	Current controlled by electrical machines decline
S_8	Opening-door signal	S_{26}	S_1 switch off	S_{44}	Door opens to the maximum
S_9	VCU	S_{27}	S_4 switch off	S_{45}	Minor control 4
S_{10}	Relays in door system get electricity	S_{28}	Backing pin popping	S_{46}	Minor control 5
S_{11}	Zero-speed signal	S_{29}	Racks slide on the short guide pillar, with the long guide pillar moves laterally	S_{47}	Minor control 6
S_{12}	Opening-door signal	S_{30}	Lower suspension arm swings	S_{48}	Mechanical machines shut down
S_{13}	Normally open contacts of EDCU power circuits close	S_{31}	Idler wheel slides in the cross slide way	S_{49}	Yellow lights on
S_{14}	EDCU gets electricity	S_{32}	Idler wheel enters the lower slide way	S_{50}	Door stop
S_{15}	Minor control library 1	S_{33}	Screw rod	S_{51}	Motor current increases instantly

(continued)

Table 16.2 (continued)

Marking	Implication	Marking	Implication	Marking	Implication
S_{16}	EDCU	S_{34}	Screw nut		
S_{17}	Yellow lights wink	S_{35}	Masts drive the door moves longitudinally		

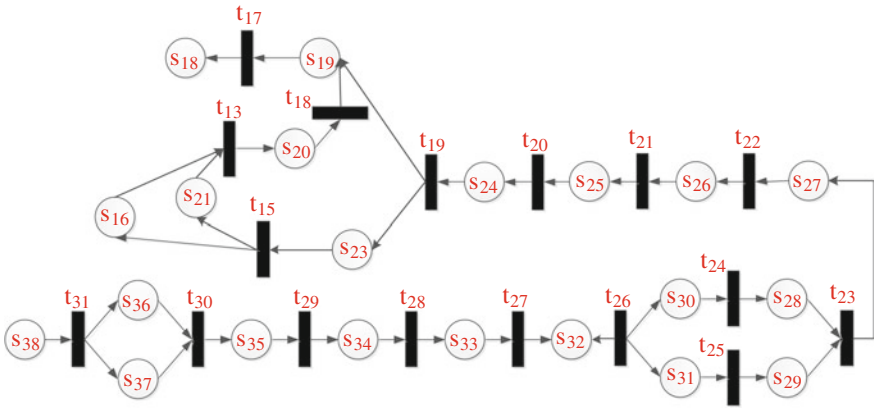


Fig. 16.3 Schematic diagram of TC-PPN of the door system during the opening-door process

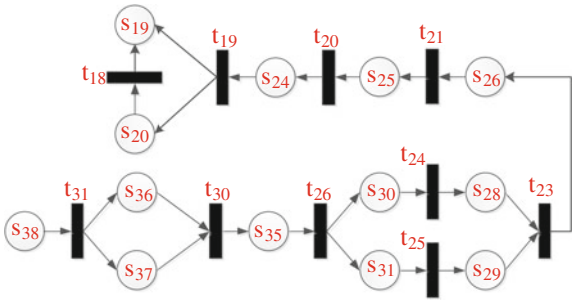


Fig. 16.4 Type I decoupling net of the door system during the opening-door process

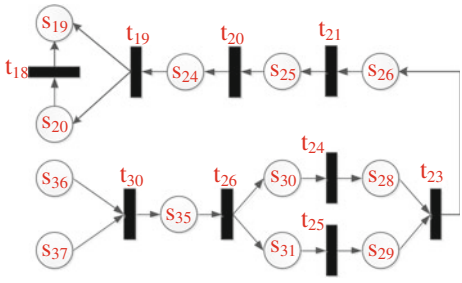


Fig. 16.5 Type II decoupling net of the door system during the opening-door process

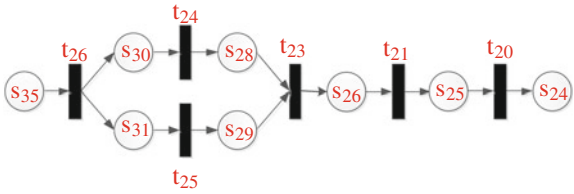


Fig. 16.6 Fault causal chain analyzed by decoupling

16.4 Conclusion

This paper focuses on fault decoupling analysis of compound faults in railway door system. Through decoupling the system model, layer by layer and creating solution models with less complicity, the calculating burden is decreased under the circumstance of multi-related coupling in the complex system. Applying ODDT analysis to TC-PPN model, two stage models of decoupling compound casual chains are created. Without damaging FDES compound casual chain constrained by one kind of time, the complex relationship in FDES can be decoupled at different layers. In this way, the obscure casual relationship may get clarity and the difficulty of problem solving is decreased. Further, identifying the system condition and puzzled relationship can be conducted at less cost. The property analysis of two stage models type I and type II decoupling net indicates that the decoupled model is in accordance with the previous model in terms of solving problems, which means no loss in the decoupling process.

Acknowledgments This work is supported by the Excellent Doctor Fund under Grant 2014YJS128, the Research Fund for the Doctoral Program of Higher Education of China under grand 20120009110035, the State Key Laboratory Program under Grant RCS2014ZT24. Thanks for the help.

References

1. Liu JB, Veneris A (2005) Incremental fault diagnosis. IEEE Trans Comput Aided Des Integr Circuits Syst 24(2):240–251
2. Lin F, Ying H (2001) Fuzzy discrete event systems and their observability. In: Proceedings of the joint 9th IFSA World Congress and 20th NAFIPS international conference. IEEE, pp 1271–1276
3. Viswanadham N, Johnson TL (1988) Fault detection and diagnosis of automated manufacturing systems.In: Proceedings of the 27th IEEE conference on decision and control, vol 3, pp 2301–2306
4. Lo KL, Ng HS, Grant DM (1999) Extended Petri-net models for fault diagnosis for substation automation. In: IEEE Proceedings on generation, transmission and distribution, vol 146(3), pp 229–234

5. Ho Y, Cassandras CG (1980) Computing co-state variables for discrete event systems. In: Proceedings of the 19th IEEE conference on decision and control including the symposium on adaptive processes. IEEE, pp 697–700
6. Jayasiri A, Mann GKI, Gosine RG (2009) Mobile robot behavior coordination using supervisory control of fuzzy discrete event systems. In: Proceedings of the 2009 IEEE/RSJ international conference on intelligent robots and systems. IEEE Computer Society, pp 690–695
7. Kilic E (2008) Diagnosability of fuzzy discrete event systems. Inf Sci 178(3):858–870
8. Huq R, Mann GKI, Gosine RG (2006) Behavior-modulation technique in mobile robotics using fuzzy discrete event system. IEEE Trans Rob 22(5):903–916
9. Nie M, Tan WW (2012) Theory of generalized fuzzy discrete event system. In: Proceedings of the 2012 IEEE international conference on fuzzy systems. IEEE, pp 1–7
10. Cao Y, Ying M (2006) Observability and decentralized control of fuzzy discrete-event systems. IEEE Trans Fuzzy Syst 14(2):202–216

Chapter 17
Estimation of Vertical Track Irregularity Based on NARX Neural Network

Jie Jiang, Fuge Wang, Yong Zhang, Yong Qin and Xudong Gao

Abstract A method based on NARX neural network was proposed to estimate the vertical track irregularities. First, the basic theory of track vertical irregularity and inertial methods used to monitor track irregularity were introduced respectively. Second, NARX neural network structure and Bayesian regularization network training algorithm were described in detail. Third, a vehicle-track dynamic model was developed. Then, the actual track irregularity data from high-speed line was used to obtain the simulation response data. Finally, the normalized data was used as input of the NARX model and the data of track irregularity was used as output. The root-mean-square error (RMSE) and correlation coefficients were applied to evaluate the network performance. The experimental results show the efficiency and accuracy of the presented method to assess the vertical track irregularities.

Keywords Track irregularity · NARX neural network · Bayesian regularization · Vehicle-track interactions

17.1 Introduction

With the modernization of the rail transit industry, it becomes more and more important to ensure the safety and stabilization of the train. Track irregularity is the main factor for train's random vibration, track structure fatigue, and train's operating status.

J. Jiang · F. Wang · X. Gao
School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Y. Zhang (✉)
School of Automation,
Nanjing University of Science and Technology, Nanjing 210094, China
e-mail: 34445721@qq.com

Y. Qin
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_17

At present, inertial methods are widely used to monitor track irregularity in service railway vehicles [1]. However, it is difficult to monitor track irregularity of different wavelength bands through a single inertial variable, which has its own defect. Neural network provides a possibility to link inertial variables with track irregularity. Neural network has a specialized function in the field of fault diagnosis, simulation modeling, and complex system controlling, and so on [2]. Li et al. [3] proposed an emerging and powerful tool based on the neural network approach in recognizing complex patterns and nonlinear relationships between many inputs and an output. Chaolong et al. [4] used an artificial neural network model to predict the short-term changing trend of track irregularity at unit section.

17.2 The Basic Theory

17.2.1 Vertical Track Irregularity

Power spectrum like periodogram spectrum estimation is widely used to estimate vertical track irregularity. If $\{x_n, n = 0, 1, 2\ldots, M\}$ is the discrete sampling sequence of track irregularity, and its Fourier transform is shown in Eq. (17.1) [5].

$$X(e^{jw}) = \sum_{n=0}^{M-1} x(n)e^{-jwn} \tag{17.1}$$

Then, we can get its periodogram spectrum estimation as follows:

$$\hat{S}_x(e^{jw}) = \frac{1}{N}\left|X(e^{jw})\right|^2 \tag{17.2}$$

We can get the wavelength and amplitude of track irregularity from Eq. (17.2).

17.2.2 Principle of Inertial Methods

A railway bogie is composed of bogie frame, wheelsets, suspensions, and dampers. Figure 17.1 shows equipment of inertial methods mounted on one axle box and on the bogie frame above [5]. Follows are principle of three inertial methods based on axle box vibration, bogie frame vibration, and bogie pitch rate.

17.2.2.1 Axle Box Vibration

As axle box is connected with wheelset rigidly, axle box acceleration is equal to the wheel's vertical acceleration. As shown in Eq. (17.3), axle box acceleration is

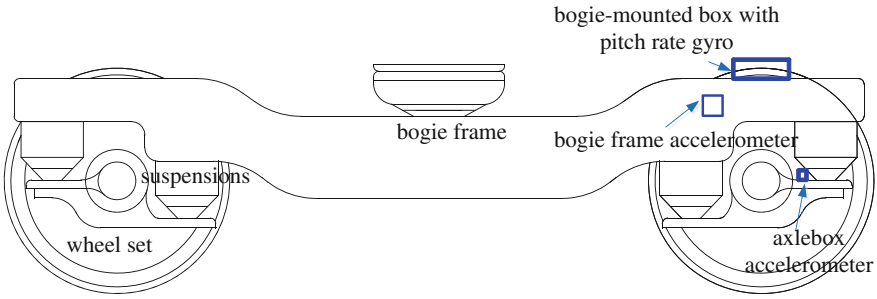


Fig. 17.1 Equipment of inertial methods

typically turned into track vertical displacement information by double integration of the signal with respect to time.

$$\hat{z}_a = \iint a \, \mathrm{d}t\mathrm{d}t = \iint \frac{a}{v^2} \, \mathrm{d}s\mathrm{d}s \tag{17.3}$$

where a is axle box acceleration, v is speed, dt is time sampling step length, ds is space interval, and $\hat{z}_a$ is vertical track irregularity.

17.2.2.2 Bogie Frame Vibration

If the primary and secondary suspension systems are ignored, bogie vertical motion can be obtained by the average of the leading and trailing wheelsets vertical motions and it is close to the track vertical geometry at wavelengths well above the bogie wheelbase.

The accelerometer mounted on the bogie is isolated from the high accelerations at the axle box. Therefore, bogie frame vibration can provide longer wavelength information with lower levels of linearity and noise. But the restricted condition still limits the final accuracy.

17.2.2.3 Bogie Pitch Rate

As shown in Fig. 17.2, an accelerometer mounted on a bogie senses acceleration roughly perpendicular to the trajectory of the sensor through space.

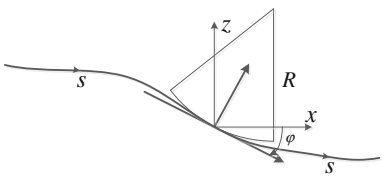


Fig. 17.2 Trajectory coordinate system

The pitch of the bogie is essentially the same as angle from horizontal of the track under the middle of the bogie, and the estimate of vertical displacement is given by Eq. (17.4).

$$\hat{z}_{\varphi} = -\int \varphi \, \mathrm{d}s = -\iint \frac{\dot{\varphi}}{v} \mathrm{d}s\mathrm{d}s = -\int v \int \dot{\varphi} \, \mathrm{d}t\mathrm{d}t \tag{17.4}$$

From Eq. (17.4), we can see that the estimate of vertical displacement can be obtained by double integrating pitch rate in time-domain sampling.

17.3 NARX Neural Network

The NARX model is a derivative of the powerful and unified nonlinear auto-regressive moving average with exogenous inputs (NARMAX) model [6].

17.3.1 The Network Structure

As shown in Fig. 17.3, NARX neural network can be divided into four layers: input layer, delay layer, hidden layer, and output layer.

The effects of input layer, hidden layer, and output layer are the same as that of BP neural network. Input layer transfers input signal to hidden layer, then hidden layer sends the signal after the main process to output layer. Output layer is

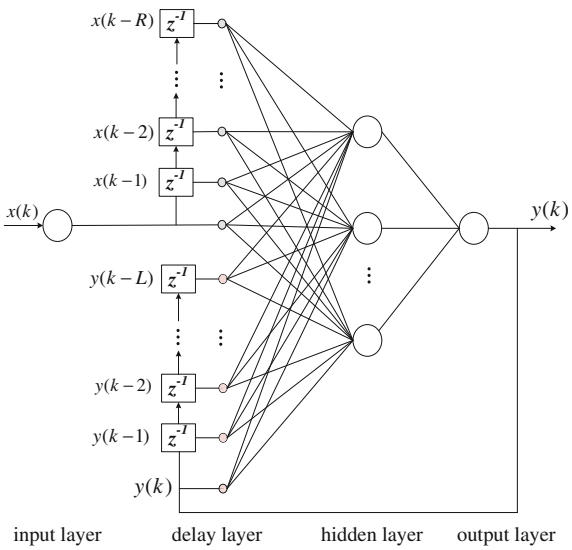


Fig. 17.3 The structure of NARX neural network

responsible for signal output after processing the signal in weighted aggregation from hidden layer. Delay layer regarded as delay operator takes input and output as network input with several steps delay.

17.3.2 Bayesian Regularization Algorithm

Compared with several other algorithms, Bayesian regularization algorithm has a small network structure and high learning efficiency, and needs simple calculation works, which improves its generalization ability by modifying training function of the neural network [7]. So, Bayesian regularization algorithm is adopted as NARX neural network learning algorithm.

The mean-square-error function is commonly used to train NARX neural network.

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}(t_i - a_i)^2 \tag{17.5}$$

where n is the total number of training value, t_i is the expected output value of the training value, and a_i is the network actual output value of the training value. Network performance function is stated as

$$F(\omega) = \alpha \cdot E_w + \beta \cdot E_D \tag{17.6}$$

where E_w denotes the square of network ownership value, E_D denotes the error square, α and β are parameters.

17.3.3 The Performance Evaluation of Neural Network

Root-mean-square error (RMSE) is adopted to evaluate the performance of neural network, and the expression is stated as

$$\mathrm{RMSE}(y, y_m) = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N}(y(i) - y_m(i))^2} \tag{17.7}$$

where y denotes target value of test sample, y_m denotes the output of neural network model, and N denotes the data sample number.

$$R(y, y_m) = \frac{\sum_{i=1}^{N}(y(i) - \overline{y})(y_m(i) - \overline{y_m})}{\sqrt{\sum_{i=1}^{N}(y(i) - \overline{y})^2 \cdot \sum_{i=1}^{N}(y_m(i) - \overline{y_m})^2}} \tag{17.8}$$

In order to further define the difference between the forecast output of neural network and the actual track irregularity, the correlation coefficient R is used to measure the correlation between the forecast data and actual data.

Where $\bar{y}$ is the average value of the test samples, and $\overline{y_m}$ is the average value of the neural network model output. The accuracy of the model is related to R. The more close to 1 R is, the higher accuracy the model has.

17.4 Experimental Validation

A vehicle-track dynamical model is used in this study for simulating axle box vibration, bogie frame vibration, and bogie pitch rate. Then these inertial variables are used as NARX neural network input units to assess the vertical track irregularity.

17.4.1 Vehicle-Track Dynamical Model

Figure 17.4 illustrates the dynamic model used in this study. The vehicle model is made of wheelsets, bogie frame, primary and secondary suspensions, and car body. Wheel and rail are coupled together with a linearized Hertzian spring, and the whole system is solved by a simple fast explicit integration method developed by Zhai et al. [8].

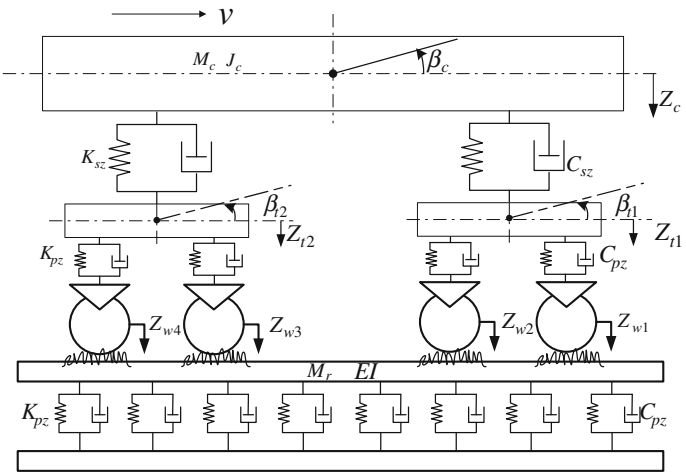


Fig. 17.4 The dynamic model for vertical vehicle-track interaction

17.4.2 Results Analysis

Input and output data of the system take the following normalization processing:

$$x_i^{\text{scal}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{17.9}$$

where x_i is the original data, $x_{\max}$ and $x_{\min}$ are, respectively, the maximum and the minimum of x.

A series of actual measured track irregularities from high-speed line are as vehicle-track dynamical model inputs, and the irregularities are discrete data sampled at 0.5 m intervals, with a length of 2 km. 3000 sets of data are collected, and the former 2500 sets are used for neural network training, the latter 500 sets are used for neural network testing. Figure 17.5 shows the input and output data of the vehicle-track dynamical model after normalization processing. The NARX model selects 20 as hidden nodes, 45 as input and output delay order number based on expertise and trial-and-error methods. Figure 17.6 shows the comparison between the actual track irregularity and the output of NARX neural network.

Figure 17.7 shows the relative analysis results between the actual track irregularity and the output of NARX neural network. As shown in Fig. 17.7, the training

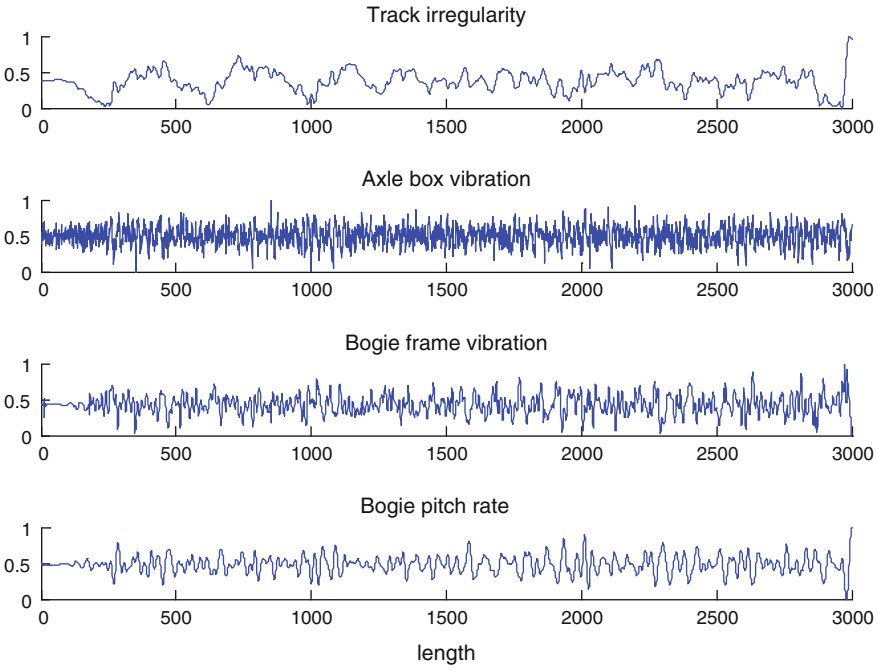


Fig. 17.5 The input and output data of vehicle-track dynamical model

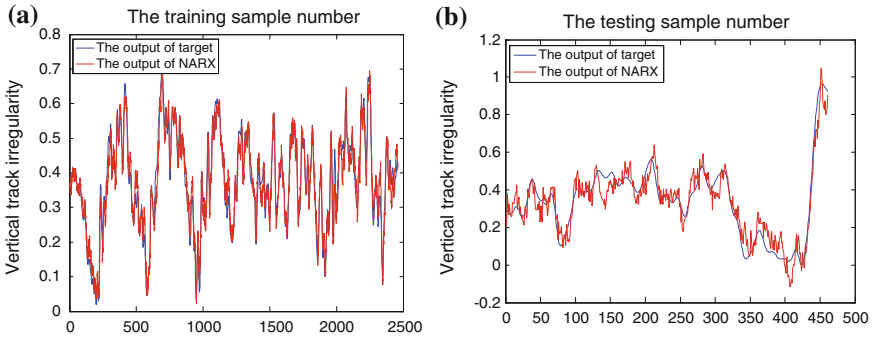


Fig. 17.6 The output of target and NARX

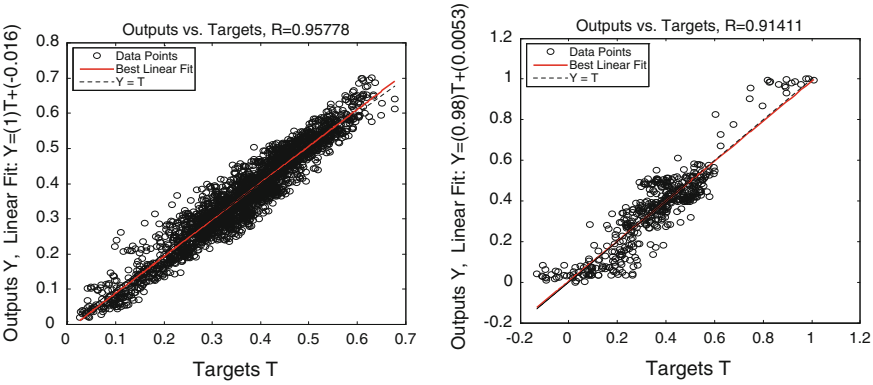


Fig. 17.7 The relativity between target and NARX output

correlation coefficient R is 0.9578 and the corresponding RMSE is 0.0386, the test correlation coefficient R is 0.9141 and the corresponding RMSE is 0.0821. The results show the efficiency of the neural model. It is practicable to use NARX neural network to assess vertical track irregularity.

17.5 Conclusions

An assessment method of vertical track irregularity using NARX neural network has been presented. First, axle box vibration, bogie frame vibration, and bogie pitch rate are obtained by vehicle-track dynamical model simulation. Then, with the simulation response data as multi-input and track irregularity as output, NARX model is used to assess the vertical track irregularity. The experimental results verify the efficiency of the neural model.

Acknowledgments The Fundamental Research Funds for the Central Universities (AE89454). The author gratefully acknowledge the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially.

References

1. Zhang J, Zhao Y, Zhang Y et al (2014) Identification of the power spectral density of vertical track irregularities based on inverse pseudo-excitation method and symplectic mathematical method. Inverse Prob Sci Eng 22(2):334–350
2. Zaiwei L, Songliang L (2012) Analysis of track irregularity signal by improved Hilbert-Huang transform. J Shenzhen Univ (Sci Eng) 29(3):270–275 (in Chinese)
3. Li C, Bovik AC, Wu X (2011) Blind image quality assessment using a general regression neural network. IEEE Trans Neural Networks 22(5):793–799
4. Chaolong J, Weixiang X, Futian W, et al (2012) Track irregularity time series analysis and trend forecasting[J]. Discrete Dyn Nat Soc 2012 Article ID 387857, 15 pages
5. Muramoto K, Nakamura T, Sakurai T (2012) A study of the effect of track irregularity prevention methods for the transition zone between different track structures. Q Rep RTRI 53 (4):211–215
6. Yassin IM, Taib MN, Rahim NA et al (2010) Particle swarm optimization for NARX structure selection—application on DC motor model. In: 2010 IEEE symposium on industrial electronics and applications (ISIEA), IEEE, pp 456–462
7. Hirschen K, Schafer M (2006) Bayesian regularization neural networks for optimizing fluid flow processes. Comput Methods Appl Mech Eng 195:481–500
8. Zhai W, Wang K, Cai C (2009) Fundamentals of vehicle–track coupled dynamics. Chin Veh Syst Dyn 47(11):1349–1376 (in Chinese)

Chapter 18
Train Rolling Bearing Degradation Condition Assessment Based on Local Mean Decomposition and Support Vector Data Description

Dandan Wang, Yong Qin, Xiaoqing Cheng, Zhilong Zhang, Hengkui Li and Xiaojun Deng

Abstract For effective utilization of a large amount of vibration data which are collected during the normal operation of train rolling bearing, this paper puts forward a new method for rolling bearing degradation condition assessment which combines the local mean decomposition (LMD) and support vector data description (SVDD). LMD is used to decompose the vibration signal, after the decomposition, we extract feature vector from three points of view: time–frequency, energy and entropy, statistical characteristic value. Principal component analysis can help to reduce dimension. Therefore, we just need to collect the data when rolling bearing normally operates to establish the evaluation model, and then realize the rolling bearing degradation status quantitative evaluation.

Keywords Support vector data description · Local mean decomposition · Principal component analysis · Degradation condition assessment · Rolling bearing

18.1 Introduction

Rolling bearing is the key component of urban rail train, whose performance directly affects the safety operation of train. In recent years, most research focus on status identification based on failure data. However, under normal working conditions, failure data is difficult to obtain. So how to use the bearing data during

D. Wang · Y. Qin (✉) · X. Cheng
State Key Laboratory of Rail Traffic Control and Safety, School of Traffic and Transportation, Beijing Jiaotong University, Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing 100044, China
e-mail: qinyong2146@126.com

Z. Zhang · H. Li · X. Deng
CSR Qingdao Sifang Co., Ltd., No.88 Jinhongdong Road, Chengyang, Qingdao 266111, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_18

normal operation to assess the state of rolling bearing degradation performance has become a popular method for research these years.

In the 1970s, Jay Lee put forward the concept of equipment performance degradation. It aims to forecast the fault of product and manage its health. This study attracted much attention since it appeared both at home and abroad. In 2001, the University of Wisconsin and Michigan joined with nearly 40 companies to build an intelligent maintenance system research center; some performance degradation assessment methods are put forward in succession, such as the cerebellum model of neural network [1], the self-organizing neural network [2], logistic regression [3], hidden markov models [4], etc.

In order to extract more comprehensive signal features, in this paper, we extract feature vector from three aspects separately after local mean decomposition (LMD): time–frequency, energy and entropy, statistical characteristic value. Principal component analysis can be used to reduce the dimension and save most of the effective information at the same time. Then we build a hypersphere to measure the degradation degree by support vector data description (SVDD).

18.2 Basic Principle of the Research

18.2.1 Local Mean Decomposition

In recent years, Jonathan S. Smith put forward a new adaptive nonstationary signal processing method called LMD, which was used to analyze the computer signal and achieved a good result. After that, some scholars applied this method to the field of mechanical vibration and obtained good effect too [5]. In this paper, we use the LMD based on cubic spline function to process the vibration data. Here are the steps for LMD [6].

Step 1: $x(t)$ is the original signal, find all local extreme values and connect them with cubic spline curve, get the up envelope $E_{\max}(t)$ and low envelope $E_{\min}(t)$;

Step 2: $m_{11}(t)$ is the local mean function and the envelope estimation function $c_{11}(t)$;

$$m_{11}(t) = \frac{E_{\max}(t) + E_{\min}(t)}{2} \tag{18.2.1}$$

$$c_{11}(t) = \frac{|E_{\max}(t) - E_{\min}(t)|}{2} \tag{18.2.2}$$

Step 3: $m_{11}(t)$ is subtracted from the original data $x(t)$;

$$h_{11}(t) = x(t) - m_{11}(t) \tag{18.2.3}$$

Step 4: $h_{11}(t)$ is shown as amplitude demodulated by dividing it by $c_{11}(t)$;

$$s_{11}(t) = \frac{h_{11}(t)}{c_{11}(t)} \tag{18.2.4}$$

Step 5: A smoothed local mean $m_{12}(t)$ is calculated for $s_{11}(t)$, subtracted from demodulated using $s_{11}(t)$, and the resulting function is amplitude demodulated using $c_{12}(t)$. This iteration process continues n times until a purely frequency modulated signal $s_{1n}(t)$ is obtained. So

$$\begin{cases} h_{11}(t) = x(t) - m_{11}(t) \\ h_{12}(t) = s_{11}(t) - m_{12}(t) \\ \vdots \\ h_{1n}(t) = s_{1(n-1)}(t) - m_{1n}(t) \end{cases} \tag{18.2.5}$$

where

$$\begin{cases} s_{11}(t) = \frac{h_{11}(t)}{c_{11}(t)} \\ s_{12}(t) = \frac{h_{12}(t)}{c_{12}(t)} \\ \vdots \\ s_{1n}(t) = \frac{h_{1n}(t)}{c_{1n}(t)} \end{cases} \tag{18.2.6}$$

where the objective is that $\lim_{n\to\infty} c_{1n}(t) = 1$.

Step 6: $c_1(t)$ is the envelope;

$$c_1(t) = c_{11}(t) \cdot c_{12}(t) \cdots c_{1n}(t) = \prod_{i=1}^{n} c_{1i}(t) \tag{18.2.7}$$

Step 7: $PF_1(\mathrm{t})$ is the component of the decomposition;

$$PF_1(t) = c_1(t) \cdot s_{1n}(t) \tag{18.2.8}$$

Step 8: $r_1(t)$ now becomes the new data and the whole process is repeated k times until $r_k(t)$ is a constant or contains no more oscillations;

$$\begin{cases} r_1(t) = x(t) - PF_1(t) \\ r_2(t) = r_1(t) - PF_2(t) \\ \vdots \\ r_k(t) = r_{k-1}(t) - PF_k(t) \end{cases} \tag{18.2.9}$$

The scheme is complete in the sense that the original signal can be reconstructed according to

$$x(t) = \sum_{v=1}^{k} PF_v(t) + r_k(t) \tag{18.2.10}$$

18.2.2 Support Vector Data Description

SVDD, proposed by Tax and Duin [7] in the year 2004, is a model which aims at finding spherically shaped boundary around a dataset. Given a set of training data $x(i) \in R^n, i = 1, 2, \ldots, l$, Tax and Duin solved the following optimization problem:

$$\min_{R,a,\xi} \left(R^2 + C \sum_{i=1}^{l} \xi^2 \right) \tag{18.2.11}$$

$$\text{subject to} \quad \begin{array}{l} \|\Phi(x_i) - a\|^2 \le R^2 + \xi_i, \quad i = 1, \ldots, l, \\ \xi_i \ge 0, \quad i = 1, \ldots, l \end{array}$$

where Φ is a function mapping the data to a higher dimensional space, and $C > 0$ is a user-specified parameter. After (18.2.11) is solved, a hyperspherical model is characterized by the center a and the radius R. A testing instance x is detected as an outlier if

$$\|\Phi(x) - a\|^2 \ge R^2 \tag{18.2.12}$$

As can been seen from the formula, what we need to do is to find a smallest super ball to obtain all of the normal data. During this process, there must be some point far away form the center of the sphere, controlling the value of ξ could help adjust the size of the sphere. Figure 18.8 shows a sphere of SVDD in two-dimensional space, the data come from a banana dataset, the red points are the support vector.

18.3 Train Rolling Bearing Degradation Condition Assessment

18.3.1 Data Sources and the Technology Roadmap

The data which this paper used was generated by the NSF I/UCR Center for Intelligent Maintenance Systems with support from Rexnord Corp. in Milwaukee, WI. The rotation speed was kept constant at 2000 RPM by an AC motor coupled to the shaft via rub belts. A radial load of 6000 lbs is applied onto the shaft and bearing by a spring mechanism. Each dataset consists of individual files that are 1-s

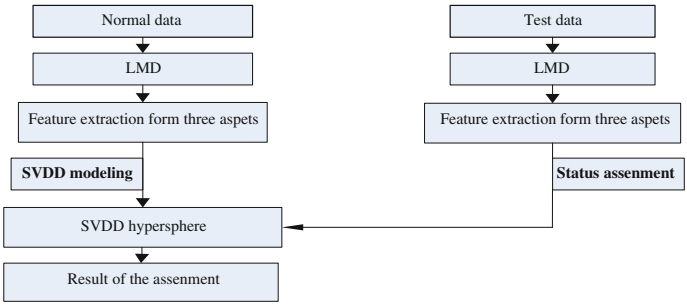


Fig. 18.1 The technology roadmap of this research

vibration signal snapshots recorded at specific intervals. Each file consists of 20,480 points with the sampling rate set at 20 kHz. Figure 18.1 is the technology roadmap.

18.3.2 Example of Train Rolling Bearing Degradation Condition Assessment

Based on the technology roadmap mentioned, decomposing signal is the first step to processing the data. 1-s vibration signal can be divided into 34 pieces (600 points/piece, one revolution), every piece of data is decomposed by LMD. From the result of the decomposition, amount of the PF is not constant, some pieces are five, but the others are six. To solve this problem, we pick the minimum number of the PF, and cancel the other PF.

Figure 18.2 provides a process of decomposition which comes from one piece of data. As we can see from it, LMD could obtain less PF, it is convenient for extracting feature vectors. And for strengthening the difference, we add the 34 pieces feature together.

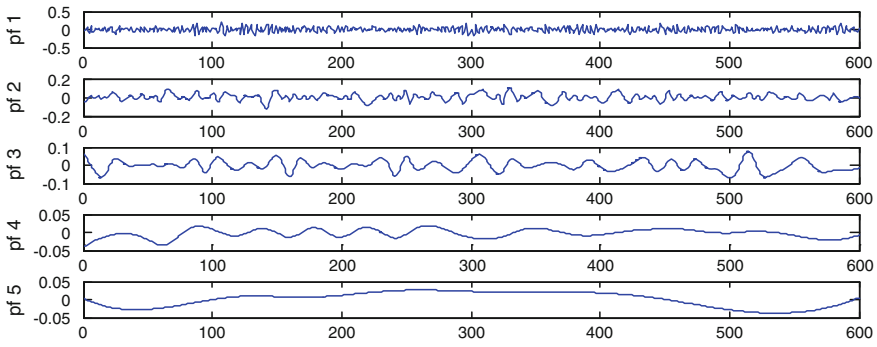


Fig. 18.2 A vibration signal and its PFs

1. Time–frequency domain: root mean square (RMS), peak, kurtosis, Xr, mean, crest factor, kurtosis factor, shape factor;
2. Energy and entropy: energy, Shanon entropy, energy moment;
3. Statistical characteristic value: Weibull shape parameter, Weibull scale parameters.

Figures 18.3, 18.4, and 18.6 are the features from three domains which obtain the whole life signal of the first PF.

Before fitting statistical distribution type, the original PF signal should be processed. Here the original vibration data is transferred by Hilbert to easily to detect

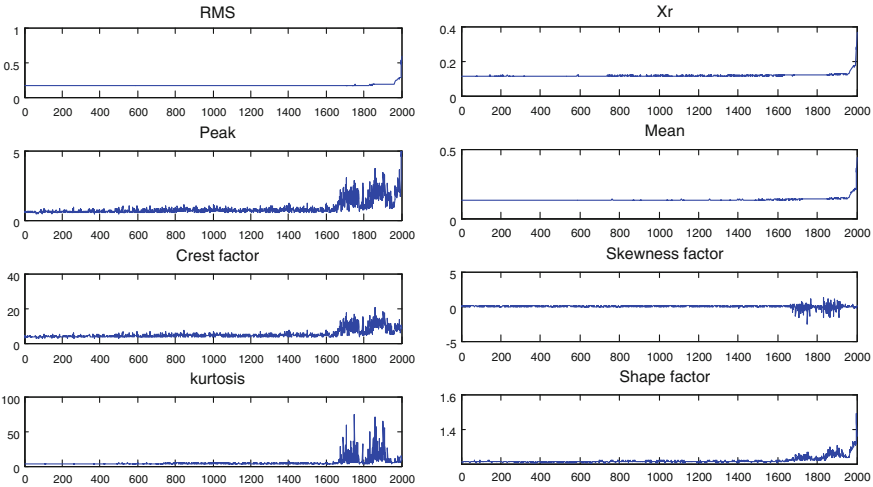


Fig. 18.3 Time–frequency feature

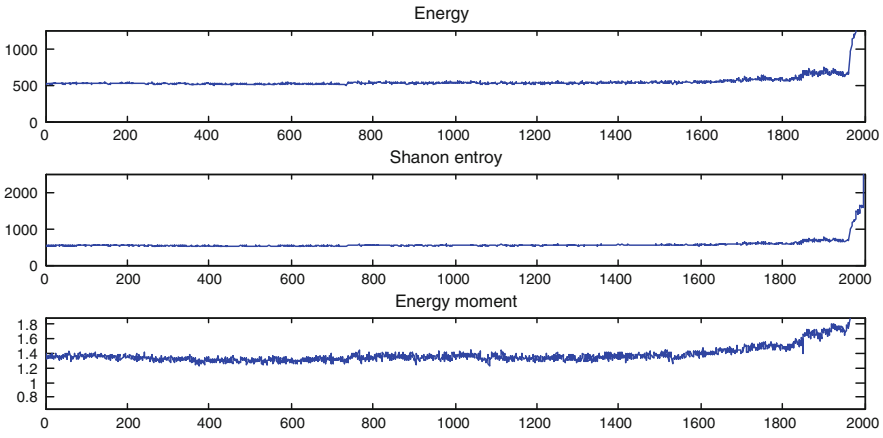


Fig. 18.4 Energy and entropy feature

an early failure, and then get the spectral envelope [8]. Figure 18.5 is checking for fitting, if the relationship is linear, the processed data is supposed to fit the statistical distribution, as we can see in the picture, it fits the distribution well (Fig. 18.6).

PCA is a method to reduce the dimension at the same time without losing important information. The principle to choose the main component is the cumulative variance contribution rate, according to it, we pick the component that occupies most of the information to be a feature vector and put it into the SVDD (Fig. 18.7).

In this paper, the first 200 pieces (here every 34 groups become one piece, which represent the status of one second) of signal data (test data) are put into the SVDD to build a hypersphere, Gauss is chosen to be the kernel function. And the last 515 pieces of the whole life data are the test data. Here the problem is, the points and hypersphere of two-dimensional feature can be seen, but if the amount of feature dimension is above two, we cannot see any feature points and boundary, so how to adjust the parameter is the key point. Figure 18.8 is an example to show different parameters C and kernel and makes different borders, and the differences between

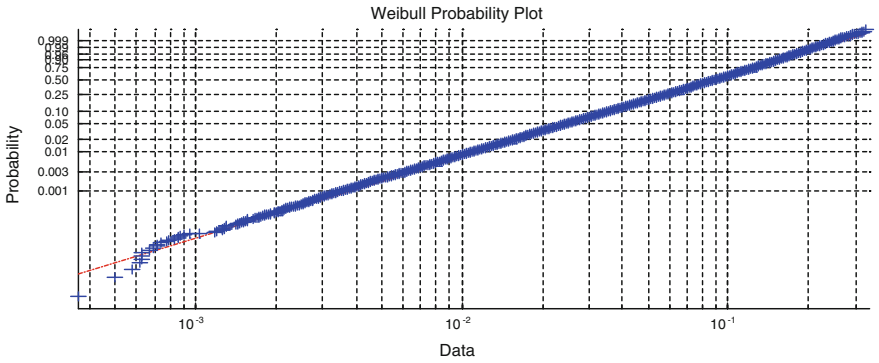


Fig. 18.5 Weibull distribution fitting check

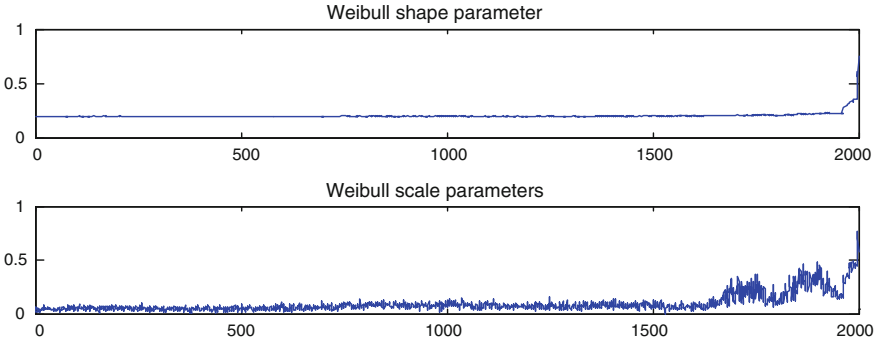


Fig. 18.6 Statistical characteristic feature

them are obviously (the remarked points are the support vectors). From the figure, we can see that the parameter of kernel affects the shape most, to choose an appropriate parameter, this paper uses the test data to check if these points in the hypersphere we build, and change the parameter of kernel to check the error rate and then find 12.4 is the appropriate parameter kernel, as can be seen from the Fig. 18.9.

Degradation condition assessment can be measured by the distance between the spot of the test data in the space and the center of the super ball. Figure 18.10 is the SVDD distance for the whole rolling bearing life. As we can see from Fig. 18.10,

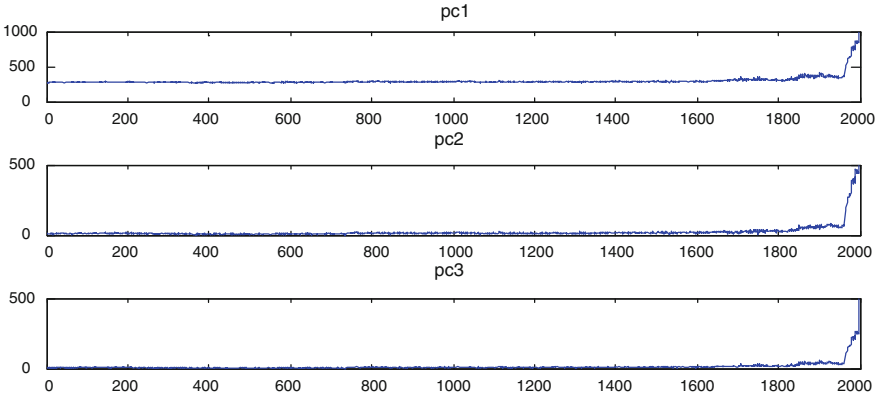


Fig. 18.7 The whole life signal of the principal components

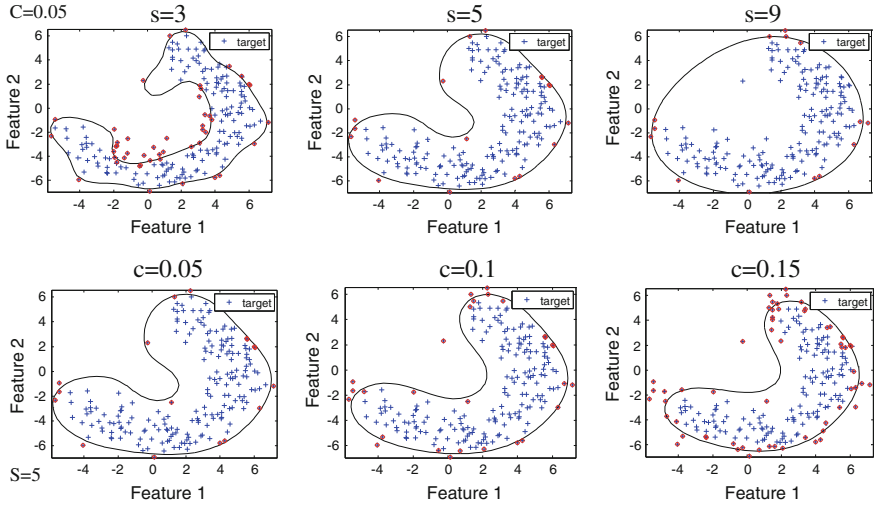


Fig. 18.8 A sphere of SVDD in two-dimensional space

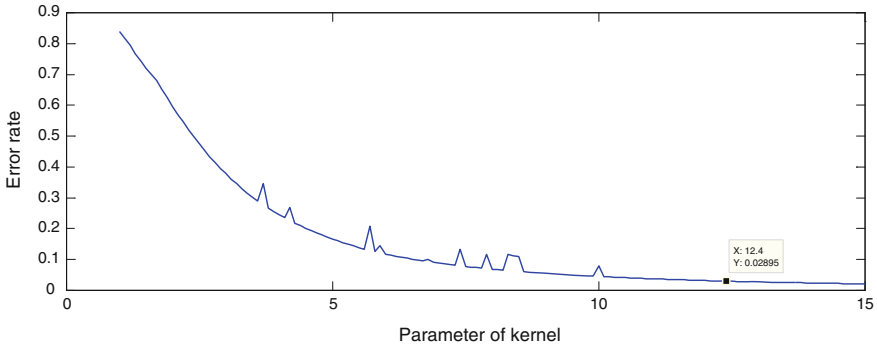


Fig. 18.9 Relationship between parameter of kernel and error rate

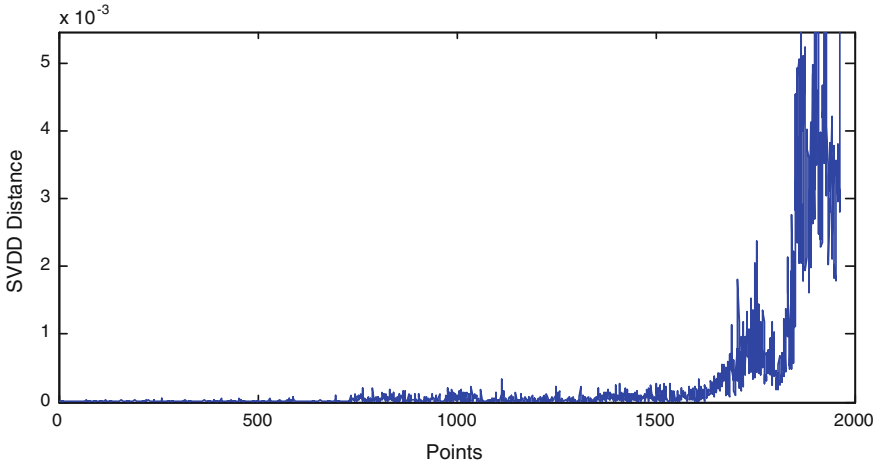


Fig. 18.10 Distance of SVDD

the bearing begins to degenerate at the point 700, and from the point 1600, the bearing degenerates deeply.

18.4 Conclusion

In this paper, the method of assessing train rolling bearing degradation status is investigated. The result show that the way we extract feature vector could better save the most of the key information, and it is also convenient for computing. What is more, SVDD can help to realize the rolling bearing degradation status

quantitative evaluation using the normal operation data before the bearing completely runs out, then it is important to ensure the safety of passengers' life and property.

Acknowledgments This work is supported by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20120009110035 and theory of mass transit train system reliability and safety assessment (No. I14K00451).

References

1. Lee J (1996) Measurement of machine performance degradation using a neural network model. Comput Ind 30:193–209
2. Huang R, Xi L, Li X et al (2007) Residual life predictions for ball bearing based on self-organizing map and back propagation neural network methods. Mech Syst Signal Process 21:193–207
3. Yan JH, Li J (2005) Degradation assessment and fault modes classification using logistic regression. J Manuf Sci Eng Trans ASME 127:912–914
4. Ocak H, Loparo KA, Discenzo FM (2007) Online tracking od bearing wear using wavelet packet decomposition and probabilistic modeling a method for bearing prognostics. J Sound Vib 302:951–961
5. Wang WQ (2004) Prognosis of machine health condition using neuro-fuzzy systems. Mech Syst Signal Process 18:813–831
6. Zhang Y (2014) Service identification and prediction method research based on safety region estimation for the key experiments in rail vehicles. Beijing Jiaotong University
7. Tax DMJ, Duin RPW (1999) Data domain description using support vectors. Eur Symp Artif Neural Netw Bruges 21–23:251–256
8. Chen C, Tang B, Lv Z (2014) Prediction for performance degradation status of rolling bearing based on Weibull distribution and SVDD. J Vib Shock 33:20–26

Chapter 19
Application of Reverse FTF in Metro Door Failure Analysis

Xinrong Liu, Jun Xia, Zongyi Xing, Limin Jia and Yong Qin

Abstract Door system is a key subsystem of metro train, and its reliability is related to the maintenance strategy of Metro Corporation and the safety of passengers. On the basis of failure mode effects criticality analysis (FMECA) and fault tree analysis (FTA), t reverse FTA-FMECA (FTF) was proposed for the reliability analysis of metro door. Firstly, the qualitative analysis of FTA was applied to door failure to obtain bottom events. Then, based on the historical failure data and criticality matrix, the quantitative analysis of FMECA was conducted on the bottom events and, as a result, the hazard degree of each bottom event was determined. At last, the failure modes which have great damage on door system were found out according to the hazard degree. The result shows that the proposed method can provide reference and technical support for designing door system and improving its reliability.

Keywords Reliability · Metro door · FMECA · FTA · Reverse FTF · Criticality

19.1 Introduction

Metro door system is a complex electromechanical product. As an aisle that passengers on and off subway trains, it requires high operability, environmental adaptability, and reliability. In the positive line operations, door systems have been

X. Liu · Z. Xing (✉)
School of Automation, Nanjing University of Science and Technology,
Nanjing 210094, China
e-mail: xingzongyi@163.com

J. Xia
School of Mechanical Engineering, Nanjing University of Science and Technology,
Nanjing 210094, China

L. Jia · Y. Qin
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University,
Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_19

paid growing attention by subway operating company and door design units due to its frequent failure-prone. Thus, it is particularly important to identify critical failure mode and key components through reliability analysis.

Ding et al. [1] has made fault tree analysis (FTA) in car use and maintenance, providing a powerful tool for the repair and diagnostic in car systems. Du [2] applied the FTA to the railway signal devices fault diagnosis of railway station. The results show that it can effectively improve the efficiency of the expert system. Qin et al. [3] used FMECA to analyse the double seal doors system, and found that the key components and failure modes lead to system failure.

While the FTA and FMECA have been widely used in reliability analysis of products, but the construction of fault tree and the process of quantitative calculation is complex, and large-scale fault tree is difficult to understand. FMECA is prone to omissions in practical application, so it is not suitable for cause and effect analysis with complex relationship of multifactor and multilevel. In this paper, FTA and FMECA were combined [4, 5], the reverse FTF was proposed for reliability analysis in subway doors. According to failure information found in operation and maintenance, select door failure to conduct reverse FTF, and then find out the fault mode and key components of this failure.

19.2 Brief Introduction of Door System

Metro door systems typically include suspension and guiding mechanism door, door drives, door leaf, emergency unlocking device, emergency inlet device, body sealing profiles mounted on vehicle, and other mechanical parts. Electronic door control unit (EDCU) [6], electrical connections, and various types of limit switches, lights and other electrical, or pneumatic components are responsible for the detection.

The movement of door system is controlled by an EDCU and driven by a motor. Motor is connected with lead screw through the coupling or pulley or other connecting devices. Screw nuts are connected to portable door frame with hinges; leaves connected to portable door frame provide swinging and translational motion of doors. The control during swinging and translational motion of door leaf, is completed by the system which consists of guide rollers and guide rails. When opening the door, the signal is automatically generated through the number devices or pushing the control button by the driver, then through a signal line reaching EDCU, door leaf begins to move from the fully closed state, the motor drives the screw nut by connecting means, and then the motions of portable door frame, long guide pillar, pylons, lower trolley-guiding, and other guiding components are caused, so the door leaf is ultimately put out under the guidance of the guiding device [7]. After the door leaf is put out to the full state, the guiding device will control the door leaf to offset and the door leaf will go on moving under a state parallel to the train sidewall. In the offset process, the portable door frame makes the door leaf slide smoothly along the long guide rail until the door leaf reaches the state of fully open.

19.3 Reverse FTF Analysis of Metro Doors

In order to ensure the safety of passengers traveling in the subway, when operating reliability design of the door system, it must be prevented that the system automatically opens after shut down or automatically shuts down after open. Now select the auto subway door to operate reverse FTF fault analysis. Firstly, in order to get basic bottom event, we set auto door failure as the top event and do FTA qualitative analysis about it [8]. Secondly, combined with historical fault data, we operate FMECA analysis of basic bottom event to determine the hazards of each bottom events, and finally will find out the harmful fault mode of metro doors.

19.3.1 Qualitative Analysis of FTA

Fault tree qualitative analysis aims to, find out the causes or the causes combination leading to door failure and identify all failure modes. It can help ascertain the potential failure to improve the design, guide the troubleshooting, and improve the operation and maintenance programs.

The effects of input layer, hidden layer, and output layer are the same as that of BP neural network. Input layer transfers input signal to hidden layer, then hidden layer sends the signal after the main process to output layer. Output layer is responsible for signal output after processing the signal in weighted aggregation from hidden layer. Delay layer regarded as delay operator takes input and output as network input with several steps delay.

Fault tree should include all fault events of door systems to maintain its strict logic, ignores the human factor, circuit design issues, and communication equipment in this article fault tree, the fault tree established by setting the fault of auto door system as top events is shown in Figs. 19.1, 19.2 and 19.3.

According to Figs. 19.1, 19.2 and 19.3, you can learn fault tree model of top event that “fault of auto door system” has 13 bottom event, and obtained 13 first-order minimum cut sets using the downlink law. Letters and meanings of bottom events are shown in Table 19.1.

19.3.2 FMECA Analysis

FMECA analysis includes FMEA and CA two parts. FMEA analysis mainly determines the cause of failure, fault impact, and fault level. Fault level refers to

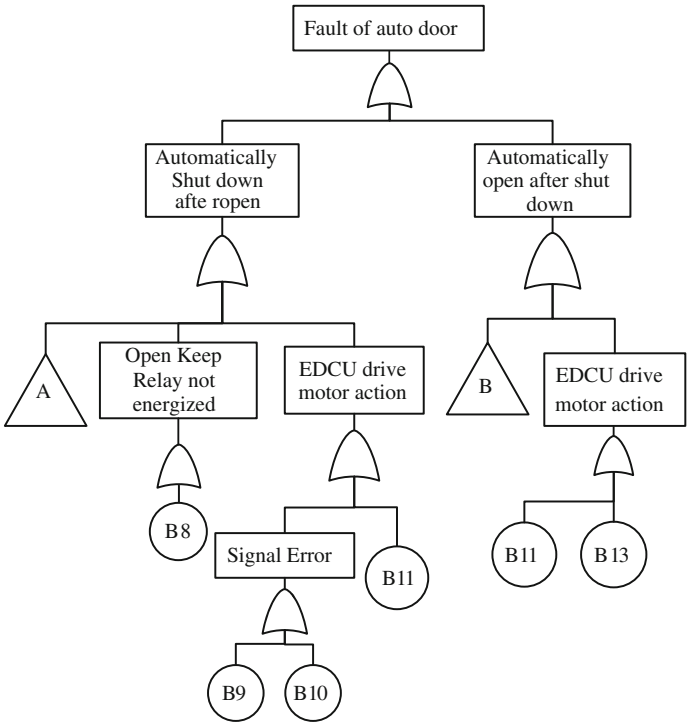
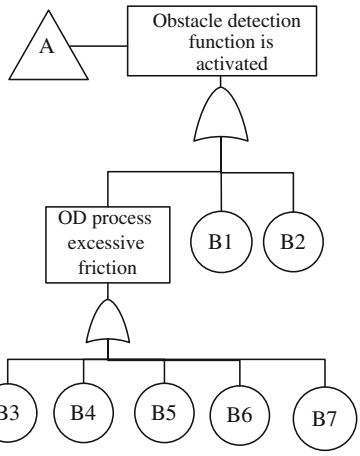


Fig. 19.1 Fault tree of metro door system

Fig. 19.2 Fault tree of metro door system



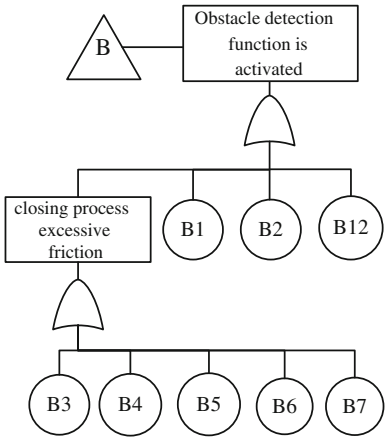


Fig. 19.3 Fault tree of metro door system

Table 19.1 Letters and meanings of bottom events

Letters	Meanings of bottom events	Letters	Meanings of bottom events
B1	Current sensor fault	B8	Switch relay fault
B2	Position sensor fault	B9	Speed sensor fault
B3	Screw poor lubrication	B10	Close button fault
B4	Guide post poor lubrication	B11	EDCU fault
B5	Upper rail position offset	B12	Closing gap S1 is too small
B6	Positioning pin mill locating slot	B13	Open button failure
B7	Pressure wheel tight		

Table 19.2 Fault classification table

Fault classification	Definitions of fault classification
I	Passenger gets hurt due to the door fault when train is operating
II	Door failure cannot be ruled out, causing the train not to continue to operate, which requires clear off outage
III	The door fault cannot be solved in a short time, train doors must be removed, then the train can continue operating
IV	Train doors failed and requires the driver or train inspection stations to troubleshoot, but caused positive line stop late, short delays
V	No effect on train operating, the train can run with the door malfunctioning, can overhaul after back to the library

consequence severity caused by system failure, the paper according to the actual situation and reference literature divides the door fault level into five classifications, as shown in Table 19.2.

Harmful CA analysis is mainly used to determine the criticality of each failure mode, the formula of i-th failure mode criticality follows:

$$C_i = \lambda\alpha\beta t \tag{19.1}$$

where λ denotes failure rate of the product; α denotes failure mode frequency ratio, indicating a failure of its total product failure modes percentage rate; β denotes probability of failure modes effects; t denotes working hours of product.

Component failure rate λ is calculated using the average failure rate [9], it is obtained through the subway door field fault data, which is calculated as:

$$\lambda = N/t \tag{19.2}$$

where N denotes total number of a component or fault mode within a specified time.

Fault Probability refers to conditional probability of ultimate impact when a system fault mode occurred to determine the failure levels.

In subway plug door system FMECA analysis, values of β are as follows [10]: $\beta = 1.0$, indicates that the system certainly has fault; $\beta = 0.5$, indicates that the system may be faulty; $\beta = 0.1$, indicates that the system is relatively trouble; $\beta = 0$, indicates that there is no effect on the system.

According to the Metro Corporation door system fault data history, the bottom incident will be FMECA analyzed; the results are shown in Table 19.3. According to the fault levels and failure mode criticality from Table 19.3, hazardous matrix can be obtained as shown in Fig. 19.4.

Figure 19.4 shows the relationship between the fault level and criticality of the 13 bottom incidents, and failure mode distribution points are away from the origin along the diagonal direction, the greater the distance, the greater the hazard of failure mode, indicating the need to strengthen maintenance.

From Fig. 19.4, it can be seen that the damage to the automatic switch of the door from high to low order is:

11 (EDCU internal fault), 4 (guide post poor lubrication), 7 (pressure wheel too tight), 3 (screw poor lubrication), 8 (switch relay failure), 9 (speed sensor fault), 10 and 13 (close button and open button failure), 12 (S1 gap is too small), 5 (top rail position offset), 1 and 2 (current and position sensors failure). It can clearly be seen that EDCU internal fault, guide post poor lubrication, pressure wheel tight, and screw poor lubrication are significantly of greater harm than other failure modes, which should be regarded as the key failure modes of the automatic switch failure of the door, and should be the focus of attention in the subway door daily maintenance. Further, it should be combined with the actual repair situation, to give the switch relay failure mode corresponding improvement measures.

Table 19.3 FMECA analysis of door system

Serial number	Failure mode	Causes	The ultimate impact	Grade	λ	α	β	t (10,000 h)	C_i
1	Current sensor fault	Quality defects; to the life cycle	Position detection function is impaired, easily leading to door auto close or open	III	0.04	1	0.1	26.6084	0.11
2	Position sensor fault	Quality defects; to the life cycle	Position detection function is impaired, easily leading to door auto close or open	III	0.04	1	0.1	26.6084	0.11
3	Screw poor lubrication	Not timely lubrication; lubricant quality problems, must be replaced	Resulting in automatic door switch failure; open slow; auto door abnormal noise	III	0.83	0.41	0.5	26.6084	4.53
4	Guide post poor lubrication	Not timely lubrication; lubricant quality problems, must be replaced	Resulting in automatic door switch failure; auto door abnormal noise	III	0.56	0.93	0.5	26.6084	6.93
5	Upper rail position offset	Routine maintenance adjustment is not in place	Door cannot switch; the door shut lax	III	0.04	1	0.5	26.6084	0.53
6	Positioning pin mill locating slot	Positioning pin displacement, and the positioning groove gap is too small	Unable to switch or door closed lax; door switch abnormal sound	IV	0.19	0.6	0.5	26.6084	1.52
7	Pressure wheel is tight	Routine maintenance adjustment is not in place	Door cannot switch; leading to the auto door fault	III	0.94	0.4	0.5	26.6084	5.01
8	Switch relay failure	Quality defects; to the life cycle	Lead to auto door failure; the train doors could not be closed;	II	0.08	1	0.1	26.6084	0.21
9	Speed sensor fault	Quality defects; to the life cycle	Lead to automatic closing fault	II	0.04	1	0.1	26.6084	0.11
10	Close button failure	Normally closed contact over contact	Lead to automatic closing fault	II	0.04	0.5	0.1	26.6084	0.053

(continued)

Table 19.3 (continued)

Serial number	Failure mode	Causes	The ultimate impact	Grade	λ	α	β	t (10,000 h)	C_i
11	EDCU failure	Door controller internal safety relay or communication interface component damaged, no output signal	Door does not properly switch; door switches normal but DDU is red; light does not shine	III	5.19	0.86	0.1	26.6084	11.88
12	Close trip switch S1 triggers excessive	Unadjusted in place when overhaul	Light does not shine; start pinch, the door cannot be normally switch; the train cannot start	III	2.29	0.07	0.5	26.6084	2.13
13	Open button failure	Normally open contact is bad	Leading to Auto door failure	II	0.04	0.5	0.1	26.6084	0.053

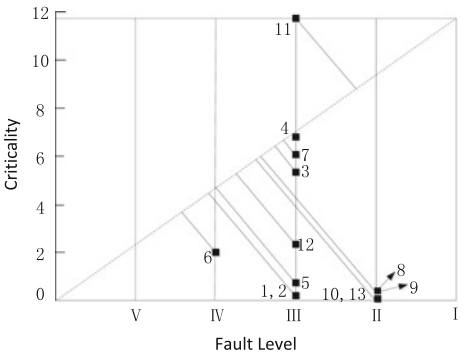


Fig. 19.4 Criticality matrix diagram of failure modes

19.4 Conclusions

On the basis of qualitative analysis of FTA in the subway train auto door failure, through further FMECA analysis of 13 bottom incidents, it can be obtained that EDCU internal fault, poor lubrication of guide posts, roller screw tight, and poor lubrication four-failure modes are four key links leading to subway train auto door failure, and should be the focus of the maintenance department. This conclusion is consistent with the experience of field maintenance engineers, verifying the effectiveness of the method in this paper, and it will provide the theoretical basis for the design and improvement of the door system and maintenance decision making.

Acknowledgments This study was supported by the Key Technologies Research and Development Program of China (2011BAG01B05), and the Fundamental Research Funds for the Central Universities (AE89454). The author gratefully acknowledge the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially.

References

1. Ding JB, Ding SQ, Song YD (2009) Application of fault tree analysis in car use and maintenance. Mech Eng Autom 05:103–104 + 107 (in Chinese)
2. Du J (2009) Implementation method of fault diagnosis expert system for railway signal equipment based on fault tree technology. Beijing Jiaotong University (in Chinese)
3. Qin SJ, Yao DM, Xu TJ et al (2008) Analysis double seal door of the ITER remote handling transfer cask system by FMEA. Mach Design Manuf 11:48–50 (in Chinese)
4. Zhan JG, Huang WM (2000) The integrated reliability analysis method using FMEA and FTA. Mach Design Manuf 1:1–3 (in Chinese)
5. Gao JY, Gong JW (2003) Safety analysis for tracked mobile robot system based on inverse FTF. Meas Control Tech 22(6):49–51 (in Chinese)
6. Guan SJ, Yang K, Pan ZY (2011) Chief introduction for the door of metro made in China. Procedia Eng 16:423–427
7. Yang K (2011) The passenger door system and its control principle on Shenzhen Metro Line 2. In: Proceedings of mechanic automation and control engineering, pp 855–857

8. Zhu XJ, Wang JB, Yin ZM (2006) On the reliability of passenger compartment door in Shanghai metro trains. Urban Mass Transit 3:31–34 (in Chinese)
9. Scorsetti M, Signori C, Lattuada P (2010) Applying failure mode effects and criticality analysis in radio therapy: lessons learned and perspectives of enhancement. Radiother Oncol 94:367–374
10. Shen GX, Li HY, Zhang YZ et al (2011) Failure analysis of turret system of NC lathe. Mach Tool Hydraulics 39(19):126–129 (in Chinese)

Chapter 20
Fatigue Life Prediction of a Top-Suspended Centrifuge Basket

Xujuan Yang, Rui Yan, Zonghua Wu, Yulong Li and Guangheng Xu

Abstract The fatigue life of a top-suspended centrifuge basket is predicted and analyzed. The static strength analysis results of the top-suspended centrifuge basket obtained by ANSYS are imported into FE-SAFE. According to its actual working process, the fatigue load spectrum of the top-suspended centrifuge basket is simplified. The Brown-Miller fatigue algorithm is employed to calculate fatigue life, and then the fatigue life distribution of the top-suspended centrifuge basket is obtained. Results show that the weakest places of fatigue life are on the both sides along the vertical direction of the basket holes.

Keywords ANSYS/FE-SAFE · Top-suspended centrifuge · Fatigue life

20.1 Introduction

The top-suspended centrifuge is mainly applied to the separation of dextrose massecuite and A-massecuite. In its working cycle, the rotational speed of the top-suspended centrifuge would be raised and decreased repeatedly, so the centrifugal force produced by massecuite on the inner wall of the basket is a periodic changed force and then the basket fatigue would be induced. With the expansion of production scale and improvement of production efficiency, the top-suspended centrifuge is developed toward larger scale and higher speed, which would make its working condition worse. So the fatigue life of the top-suspended centrifuge is facing a severe challenge and the effective prediction of fatigue life is an important part in the design process of the top-suspended centrifuge.

X. Yang (✉) · R. Yan · Z. Wu · Y. Li · G. Xu
Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology, School of Mechanical Engineering, Guangxi University, Nanning, Guangxi, China
e-mail: yxj413@163.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_20

Current research on the top-suspended centrifuge mostly focuses on the basket strength analysis [1–5], such as the stress analysis and optimization of the top-suspended centrifuge basket [1, 2], the dynamic analysis of the top-suspended centrifuge basket based on ANSYS Workbench [3], the dynamic performance analysis, and optimization of the top-suspended centrifuge basket [4]. There are a few articles on centrifuge spindle fatigue and the bottom of the basket fatigue [5, 6], but the study on the fatigue life prediction of the top-suspended centrifuge basket is rarely reported in the literature.

In this paper, the finite element model of the top-suspended centrifuge basket is established by ANSYS software and then the static strength analysis results obtained by ANSYS are imported into FE-SAFE, the Brown-Miller fatigue algorithm is employed to calculate fatigue life.

20.2 Fatigue Life Analysis Process

The fatigue life analysis process by ANSYS/FE-SAFE is shown in Fig. 20.1. First, the finite element model is established in the ANSYS software, and the static analysis of the top-suspended centrifuge basket is carried out, then the static analysis results are obtained. In the second step, static analysis result file is imported into FE-SAFE software; meanwhile the load spectrum and the material parameters are set. After these preparation works, fatigue analysis is carried out. Finally, the fatigue analysis result file of FE-SAFE is imported into ANSYS, and then the fatigue life cloud of the basket is obtained.

In the process of analysis, ANSYS provides data interface and a part of data analysis function. FE-SAFE software bears the main task of fatigue life analysis.

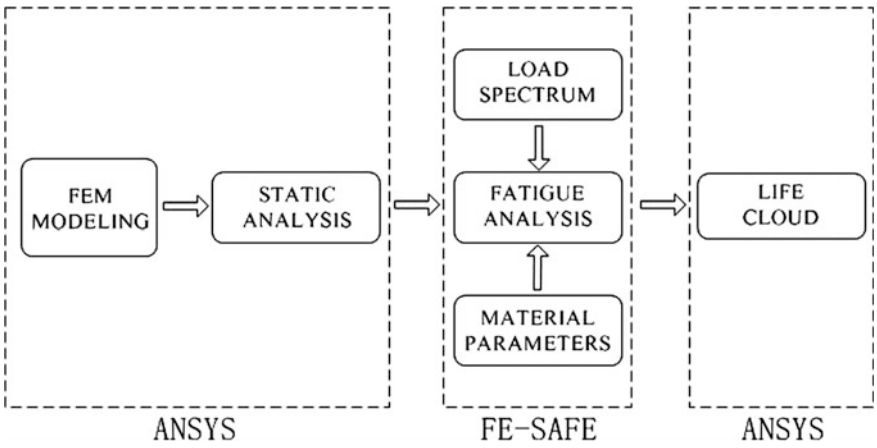


Fig. 20.1 Fatigue life analysis process

20.3 FEM Modeling and Static Analysis

According to the structure parameters by ANSYS, the top-suspended centrifuge basket model can be established. Due to the symmetry of structure and load, the finite element model of the 1/72 of top-suspended centrifuge basket is established to improve the computing speed in this paper. The material of the top-suspended centrifuge basket is SAF2205 duplex stainless steel, and its elastic modulus and Poisson's ratio are shown in Table 20.1. According to the structural characteristics of the basket, element type is defined as SOLID186.

In the working process of high-speed rotation, the basket sidewall bears great pressure from centrifugal force of massecuite, mesh liner, and strainer. The bottom surface bears pressure from the gravity of the massecuite.

The pressure of the massecuite on the sidewall can be expressed as

$$P_1 = \rho b_2 \omega^2 \frac{r_a^2 - (r_a - db_1)^2}{2} \tag{20.1}$$

where ρ is the massecuite density, b_2 is the changing coefficient of the massecuite density, ω is the rotational speed of the centrifuge, r_a is the inner radius of the basket, d is the thickness of the massecuite, b_1 is the thickness coefficient of the massecuite.

The pressure of the mesh liner and strainer on the sidewall can be expressed as

$$P_2 = \frac{m\omega^2}{2\pi h} \tag{20.2}$$

where h is the basket height, m is the mass of the mesh liner and strainer. Then the total pressure on the inner surface of the basket can be represented as follows:

$$P = P_1 + P_2 \tag{20.3}$$

In addition, the pressure on the bottom surface of the basket can be expressed as

$$P_G = h\rho b_2 g \tag{20.4}$$

In the ANSYS software, the model can be loaded and solved, and then the static analysis results can be obtained. The maximum static stress distribution of the basket can be expressed by the von Mises stress distribution in Fig. 20.2, which is the boundary condition of the fatigue life analysis. During the actual working cycle

Table 20.1 Mechanical properties of duplex stainless steel SAF2205

Steel	Heat treatment	σ_b (MPa)	σ_s (MPa)	E	μ
SAF2205	1020–1100 water quenching	620	450	200 GPa	0.3

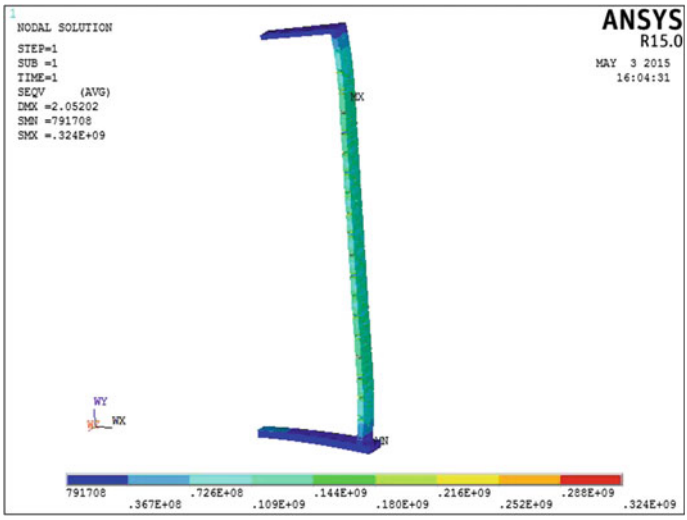


Fig. 20.2 Static analysis results of Von Mises stress distribution

of the top-suspended centrifuge, actual stress of each area in the basket is changed periodically, which cannot exceed the maximum static stress.

20.4 Fatigue Life Analysis

20.4.1 Material Parameters

In the material library of FE-SAFE, SAF2205 duplex stainless steel is selected. In addition, Segger material approximation algorithm in the FE-SAFE database provides a way to estimate the performance parameters of the material by its tensile strength (UTS) and elastic modulus [7]. According to the mechanical properties of SAF2205 duplex stainless steel, the tensile strength is 620 MPa, and the elastic modulus is 200 GPa. So the strain–life curve is obtained as shown in Fig. 20.3.

20.4.2 Fatigue Analysis Load Spectrum

The whole working cycle of the top-suspended centrifuge includes feeding, accelerating, separating, decelerating, and unloading. Taking the work process of a top-suspended centrifuge as an example, during the feeding process, the centrifuge keeps a low rotational speed of 160 r/min for 23 s. During the accelerating process, the rotational speed accelerates from 160 to 1060 r/min in 60 s. In the separating

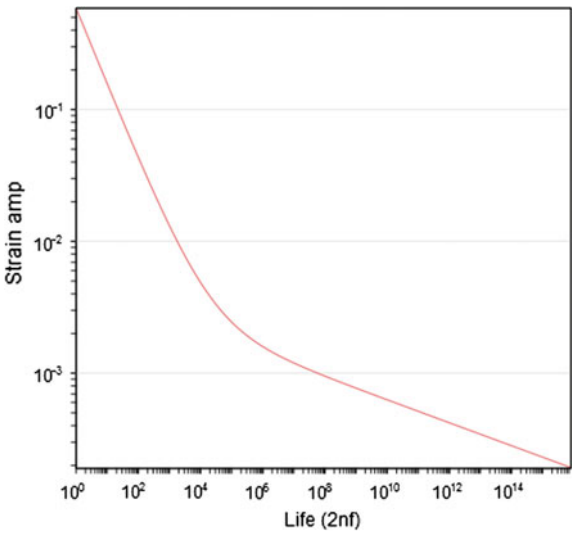


Fig. 20.3 Strain-life curve

process, the rotational speed will maintain 1060 r/min for 25 s. During the decelerating process, the rotational speed lasts for 48 s and the rotational speed will drop from 1060 to 50 r/min. Finally, the centrifuge will unload sugar at a rotational speed of 50 r/min. The whole working cycle will last for about 180–200 s.

The working cycle is simplified for convenience, the rotational speed of the top-suspended centrifuge is assumed to increase or decrease uniformly in the process of accelerating and decelerating, and then the load spectrum of the whole working cycle of the top-suspended centrifuge will be shown in Fig. 20.4.

20.4.3 Fatigue Algorithm

As SAF2205 is plastic steel, the fatigue analysis of the top-suspended centrifuge basket belongs to high cycle fatigue analysis. Brown-Miller fatigue criterion with morrow mean stress correction is used [8], and then the Brown-Miller fatigue criterion formula can be expressed as

$$\frac{\Delta\gamma_{\max}}{2}+\frac{\Delta\varepsilon_n}{2}=1.65\frac{\left(\sigma_{\mathrm{f}}'-\sigma_m\right)}{E}(2N_{\mathrm{f}})^b+1.75\varepsilon_{\mathrm{f}}'(2N_{\mathrm{f}})^c \tag{20.5}$$

where $\Delta\gamma_{\max}$ is the maximum shear strain, $\Delta\varepsilon_n$ is the positive strain, σ_{f}' is the fatigue enhancement coefficient, σ_m is the average normal stress, N_{f} is the cycle number of fatigue loading, b is the fatigue strengthening index, $\varepsilon_{\mathrm{f}}'$ is the fatigue ductility coefficient, c is the fatigue ductility exponent.

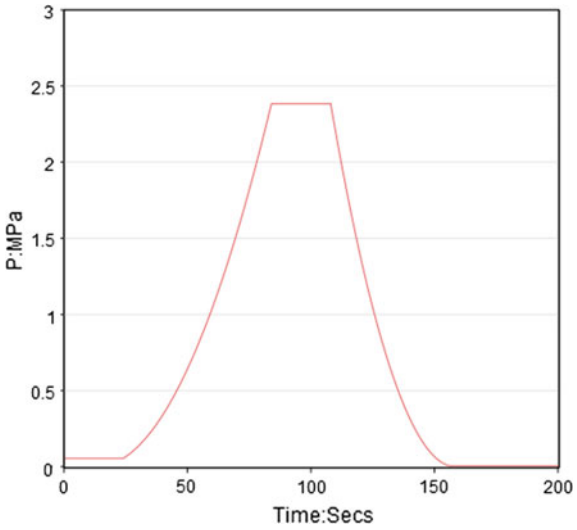


Fig. 20.4 Load spectrum of the whole working cycle

20.4.4 Simulation Results

The static analysis results obtained from ANSYS are imported into FE-SAFE, the material, load spectrum, and algorithm are defined and the surface roughness is set from 16 to 40 μm, then the fatigue life can be obtained and analyzed. The fatigue analysis results obtained from FE-SAFE are imported into ANSYS, the fatigue life cloud is shown as Fig. 20.5.

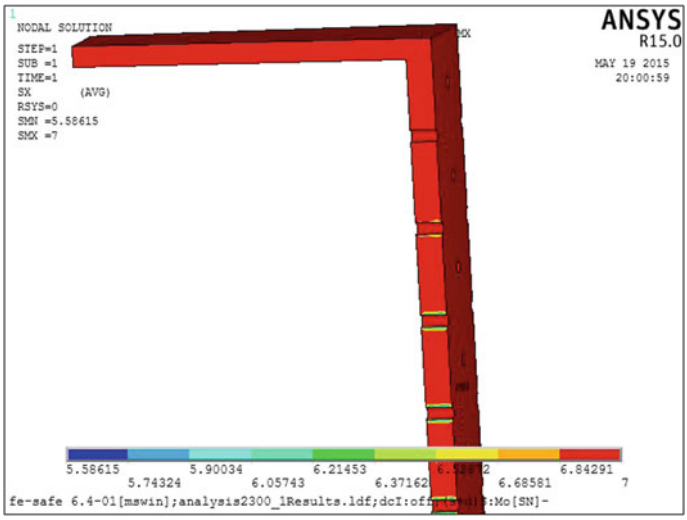


Fig. 20.5 Fatigue life cloud of the top-suspended centrifuge basket

As the fatigue life cloud is shown, the minimum life of the basket is distributed on the both sides along the vertical direction of the basket holes, and the minimum logarithmic life is 5.58615, which means that the minimum of working cycle is $10^{5.58615} = 3.856 \times 10^5$.

Assuming that the working cycle lasts for 180 s and the top-suspended centrifuge works 18 h a day, 200 days a year, so the fatigue life of the top-suspended centrifuge basket will be 5.3 years.

20.5 Conclusions

The fatigue test is a destructive test, and it is a complex task which demands lots of equipment and manpower. This paper has presented an effective method to predict the fatigue life of the top-suspended centrifuge basket. According to the simulation results, the weakest places of the fatigue life are on both the sides along the vertical direction of the basket holes, which fit the actual fatigue crack of the top-suspended centrifuge basket well.

Acknowledgments This research is supported by Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology (Grant No. 13-051-09S08) and Guangxi Higher Education Science and Technology Project (YB2014011).

References

1. Chen Y, Guangheng X, Yang X et al (2014) Stress analysis and optimization of the hanging centrifuge drum perforation. Adv Mater Res 971–973:731–735
2. Zhou L, Tan W (2007) Simulation calculation and analysis on the stress of centrifuge basket. Fluid Mach 35(8):19–21
3. Liu J, Zhao L, Zhou Q (2013) ANSYS workbench based dynamic analysis of vertical centrifuge basket. Mechanical Engineering and Automation, pp 56–57 (in Chinese)
4. Wei X, Jiang Y, Liu J (2013) Dynamic analysis and optimization for basket system of hang type centrifuge. J Mach Des 27(8):69–72 (in Chinese)
5. Huang F (1997) Fatigue fracture failure analysis of horizontal scraper centrifuge main shaft. Ordnance Mater Sci Eng 20(3):51–53 (in Chinese)
6. Xu Z, Chen F, Wang Q, Xu Y (1995) Allowable defect analysis of the bowl body of the centrifuge. Nat Sci J Xiangtan Univ 17(04):96–100 (in Chinese)
7. Jiang N (2006) Engineering and examples of ANSYS and ANSYS/FE-SAF. Hohai University Press, Nanjing, pp 75–76 (in Chinese)
8. Brown MW, Miller KJ (1973) A theory for fatigue under multiaxial stress-strain conditions. Proc Inst Mech Eng 187(65):745–755

Chapter 21
Development of Reliability Models for Railway Segment-Track Systems

Mingdian Chen, Rengkui Liu, Futian Wang and Ling Zhang

Abstract Evaluation of railway track reliability is important for the safety and security of railway operation and control of maintenance costs. This paper develops reliability models for railway segment-track systems (RS-TS), including reliability structure models and reliability assessment models. The reliability structure relationships, which include serial structure, quasi-parallel structure, and hybrid structure, among track segments and a track system are analyzed in the reliability structure models. The reliability index of track segments and a serial structure track system is calculated in reliability assessment models with the method of Markov process. It draws on track geometry data and rail wear data from Beijing Subway Line 13 to validate the models.

Keywords Railway · Track system · Segment · Reliability structure · Reliability assessment · Markov process

21.1 Introduction

The quality states of railway tracks are important to the security of railway operation. Reliability, which is the probability that a track system can perform a required function under given conditions for a given time interval, is significant for track quality state index [1]. Track quality states can be accurately mastered through the assessment of track reliability. Thus, maintenance for tracks can be effectively guided, and then safety and security of railway operation can be ensured. In addition, assessments of track reliability are the foundation of mastering track quality state degradation rules. Maintenance resource can be reasonably allocated,

M. Chen (✉) · F. Wang · L. Zhang
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Haidian District, Beijing, China
e-mail: 13120825@bjtu.edu.cn

R. Liu
School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_21

maintenance plan can be scientifically compiled, excessive repair and lack of repair can be avoided and maintenance costs can be precisely controlled by forecasting the track quality state degradation rules.

A number of researchers have attempted to analyze railway track reliability. Nguyen et al. [2] have presented a reliability assessment of railway track buckling during extreme heat wave events. Zhao et al. [3] have developed a model which can be used to evaluate the reliability of a sleeper system when the sleeper condition is known. Zheng [4] has analyzed the failure mode of rail from the point of structural reliability and calculated the reliability of 50 kg/m rail and 60 kg/m rail by means of binding the finite element method and Monte Carlo method. Li [5] has analyzed the structural failure mode of type III sleepers and calculated their failure probability and reliability index in high-speed railway. Zhang and Chen [6] have analyzed the reliability of continuously welded rail track stability by means of the Monte Carlo method. Kang et al. [7] has proposed an evaluation system based on reliability, availability, maintainability, and safety (RAMS) for the analysis of track irregularity status.

Those researches have mainly aimed at reliability of certain track facilities from the perspective of structural reliability. They have neither explicitly analyzed the comprehensive reliability of track segments or track systems, nor studied the reliability structure relationships among segments and a track system. This paper will propose reliability models for railway segment-track systems (RS-TS). Reliability structure relationships among segments and a track system will be discussed through the reliability structure models for RS-TS, and the reliability of RS-TS will be assessed through the reliability assessment models for RS-TS.

21.2 Reliability Structure Models for RS-TS

21.2.1 Analysis of Railway Track Degradation Characteristics

Railway track degradation has many characteristics such as memorability, periodicity, co-movement and heterogeneity. Degradation characteristics of a track facility in given spatial position between two successive maintenance events are similar. Occurred defects will recur generally after maintenance, which is called memorability. Memorability causes the periodicity of degradation rules, which manifests that degradation time interval are approximately same. Degradation period of track facilities in different spatial positions is different. Degradation processes between different types of track facilities among same mileage range are mutually affected, which is called the co-movement.

Railway track quality states are influenced by factors like geographical environment, transportation organization, design and manufacture, maintenance management, and so on. The key factor is the geographical environment factor

generally, which causes that the track facilities in a same type in different spatial position having different degradation rules. It is called heterogeneity, and the factors mentioned above are called heterogeneity factors [8]. These factors also affect the reliability of the track system.

Railway track unitization analysis (RTUA) method is to divide the railway track into several track segments with fixed length and to analyze the track quality state of each segment. A track segment is viewed as a whole, although it has several track facilities. Length of a segment varies from 100 to 500 m because of the difference of management standard of railway authorities in different countries, and it generally fetches 200 m [9]. The memorability, periodicity, co-movement and heterogeneity of railway track degradation determine that RUTA can well describe the degradation rules of track facilities. Therefore, RUTA has advantages in analyzing the reliability of RS-TS.

21.2.2 Modeling of Reliability Structure for RS-TS

System reliability structure theory is widely applied to mechanical engineering, electronic engineering, software engineering, and other fields. However, it is rarely used in railway track engineering. In this paper, we first apply the theory of system reliability structure to RS-TS in order to analyze the relationships among segments and a track system.

21.2.2.1 Serial Structure

This structure is suitable for a main line track system or single station track system. It is assumed that the track system consists of n serial segments, $S_1, S_2, \ldots, S_n$. The system will fail as long as any one of the segments, $S_i (i = 1, 2, \ldots, n)$, fails, as shown in Fig. 21.1.

21.2.2.2 Quasi-parallel Structure

This structure is suitable for a track system with several parallel tracks joined by an identical turnout. A turnout connects at most three parallel tracks in general, so it needs more turnouts to set more tracks. We simplify the switch area with several connected turnouts into a single turnout for convenience. It is supposed that the track system consists of the turnout T_1 and n parallel segments, $S_1, S_2, \ldots, S_n$.

Fig. 21.1 Serial structure

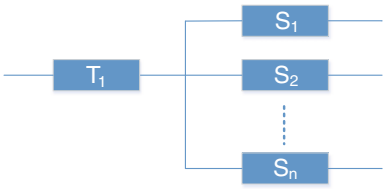


Fig. 21.2 Quasi-parallel structure

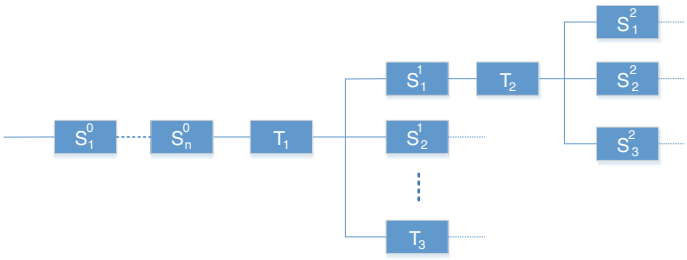


Fig. 21.3 Hybrid structure

The turnout T_1 connects the segments. Only when all segments fail will the system fail. However, if T_1 fails, the system will fail. We call the reliability structure of the system "quasi-parallel structure," as shown in Fig. 21.2.

21.2.2.3 Hybrid Structure

This structure is suitable for a complex station track system or a yard track system. It is assumed that the track system consists of several segments and several turnouts. Reliability structure relationships among segments include both serial structure and quasi-parallel structure, as shown in Fig. 21.3.

21.3 Reliability Assessment Models for RS-TS

21.3.1 Assumptions of Models

We use the method of Markov process to assess the reliability of RS-TS in the models. The following assumptions are made prior to the assessment:

1. Lives of all segments obey exponential distribution, i.e., the distribution of life X_i of segment S_i can be expressed as

$$P_i\{X_i \le t\} = 1 - e^{-\lambda_i t}, \quad t \ge 0, \quad \lambda_i > 0, \quad i = 1, 2, \ldots, n \tag{21.1}$$

2. Maintenance time of all segments after failure obeys exponential distribution, i.e., the distribution of maintenance time Y_i of segment S_i

$$P_i\{Y_i \le t\} = 1 - e^{-\mu_i t}, \quad t \ge 0, \quad \mu_i > 0, \quad i = 1, 2, \ldots, n \tag{21.2}$$

3. X_i, X_j, Y_k, Y_m are mutually independent, $i \ne j \ne k \ne m$, $i, j, k, m = 1, 2, \ldots, n$.
4. The maintenance effective is perfect.

21.3.2 Definition of Operation States of RS-TS

RAM failure categories suitable for use in railway applications include significant, major and minor failure according to the Standard EN50126 [1]. Failure categories of RS-TS can also be defined in this way, as shown in Table 21.1.

We set the major failure to the threshold between normal and failure in RS-TS, as shown in Fig. 21.4.

Table 21.1 Failure categories of RS-TS

Failure category	Definition
Significant (immobilizing failure)	A failure that prevents train movement or causes a delay to service greater than a specified time and/or generates a cost greater than a specified level
Major (service failure)	A failure that must be rectified for the system to achieve its specified performance and does not cause a delay or cost greater than the minimum threshold specified for a significant failure
Minor	A failure that does not prevent a system achieving its specified performance and does not meet criteria for significant or major failures

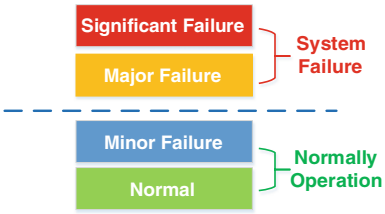


Fig. 21.4 Definition of operation states of RS-TS

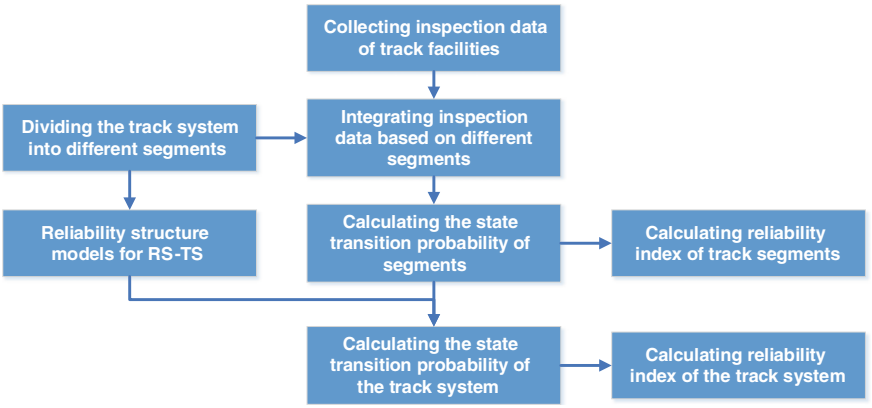


Fig. 21.5 Reliability assessing procedure for RS-TS

Then we need to determine the relationships among states of track facilities and a segment. It needs to set thresholds for different types of track facilities. When the state of a certain facility is poorer than the threshold, the segment cannot operate normally, i.e., there is occurrence of segment failure.

21.3.3 Reliability Assessing Procedure for RS-TS

The reliability assessing procedure for RS-TS is shown in Fig. 21.5.

21.3.4 Reliability Index Calculation (RIC) Models for RS-TS

21.3.4.1 RIC odel for a Single Track Segment

Considering the single segment $S_i (i = 1, 2, \ldots, n)$, we define the state zero as the segment normally operating and the state one as the segment having failures. $E_i = \{0, 1\}$ is the state set of segment S_i. The state of segment S_i at the moment t, $X_i(t)$, can be expressed as

$$X_i(t) = \begin{cases} 0, & \text{if segment } S_i \text{ is normal at the moment } t \\ 1, & \text{if segment } S_i \text{ fails at the moment } t \end{cases}$$

$\{X_i(t), t \geq 0\}$ is an invariant Markov process on the state space E_i because of the assumptions proposed above [10]. Thus, transition probability matrix $\boldsymbol{P^i}(t)$ can be expressed as

$$\boldsymbol{P^i}(t) = \begin{pmatrix} 1-\lambda_i & \lambda_i \\ \mu_i & 1-\mu_i \end{pmatrix} \tag{21.3}$$

In practical calculation, we can estimate the transition probability matrix $\boldsymbol{P^i}(t)$ by statistics method, and then λ_i and μ_i can be obtained.

The steady-state availability A_i, reliability $R_i(t)$, mean time to first failure MTTFF_i of segment S_i are given by

$$\begin{cases} A_i = \frac{\mu_i}{\lambda_i + \mu_i} \\ R_i(t) = \mathrm{e}^{-\lambda_i t} \\ \text{MTTFF}_i = \frac{1}{\lambda_i} \end{cases} \tag{21.4}$$

21.3.4.2 RIC Model for a Serial Structure Track System

We define the state zero as all segments normally operating and the state i as the segment S_i $(i = 1, 2, \ldots, n)$ having failures. $E_i = \{0, 1, 2, \ldots, n\}$ is the state set of the serial track system. The state of the system at the moment t, $X(t)$, can be expressed as

$$X(t) = \begin{cases} 0, & f \text{ all segments are normal at the moment } t \\ i, & \text{if segment } S_i \text{ fails at the moment } t \end{cases}$$

Thus transition probability matrix $\boldsymbol{P}(t)$ can be expressed as

$$\boldsymbol{P}(t) = \begin{pmatrix} 1-\Lambda & \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \mu_1 & 1-\mu_1 & 0 & \cdots & 0 \\ \mu_2 & 0 & 1-\mu_2 & \cdots & \mu_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_n & 0 & 0 & \cdots & 1-\mu_n \end{pmatrix} \tag{21.5}$$

where

$$\Lambda = \sum_{i=1}^{n} \lambda_i \tag{21.6}$$

The steady-state availability A, reliability $R(t)$, mean time to first failure MTTFF of the system are given by

$$\begin{cases} A = \left(1 + \sum_{i=1}^{n} \frac{\lambda_i}{\mu_i}\right)^{-1} \\ R(t) = \mathrm{e}^{-\Lambda t} \\ \mathrm{MTTFF} = \frac{1}{\Lambda} \end{cases} \tag{21.7}$$

21.4 Case Verification

Beijing Subway Line 13 (hereinafter using abbreviation Line 13) plays an important role in Beijing Subway. The length of its main line is 40.5 km. There are 60 turnouts and 141 curves in Line 13. The type of rail in Line 13 is CHN60.

We select the track geometry data and rail wear data from the uplink track with mileage from K11 + 000 to K12 + 000 in Line 13. The time horizon of track geometry data is from January, 2013 to October, 2014, and the inspection interval is 1.5 months. The time horizon of rail wear data is from January, 2012 to September, 2014, and the inspection interval is 3 months. We divide the track system into five segments, with which each length is 200 m.

We calculate the track quality index (TQI) and single index, such as longitudinal level, alignment, gauge, cross level, and twist, from track geometry data. We select criterions of the thresholds of TQI and damage of rail wear in Beijing Subway as the thresholds of the state of segments, as shown in Table 21.2.

After the statistical analysis, we obtain the frequency of state changes of TQI, single index, and side rail wear, as shown in Table 21.3.

Table 21.2 Criterions of the thresholds of TQI and damage of rail wear (unit: mm)

	Longitudinal level	Alignment	Gauge	Cross level	Twist	TQI	Rail wear
Thresholds	2.41 × 2	2.26 × 2	1.65	1.91	1.61	14.51	12

Table 21.3 Frequency of state changes

Segment ID	Frequency of TQI state changes				Frequency of rail wear state changes			
	$0 \to 0$	$0 \to 1$	$1 \to 0$	$1 \to 1$	$0 \to 0$	$0 \to 1$	$1 \to 0$	$1 \to 1$
S_1	92	2	1	9	44	0	0	0
S_2	95	1	1	7	43	0	0	0
S_3	35	3	3	63	112	0	0	0
S_4	47	4	3	50	70	1	1	0
S_5	72	3	3	26	118	1	1	0

Table 21.4 State transition probability and parameters of segments

i	$P_{00}^{i}(t)$	$P_{01}^{i}(t)$	$P_{10}^{i}(t)$	$P_{11}^{i}(t)$	λ_i	μ_i
1	0.986	0.014	0.1	0.9	0.014	0.1
2	0.993	0.007	0.125	0.875	0.007	0.125
3	0.980	0.020	0.045	0.955	0.020	0.045
4	0.959	0.041	0.074	0.926	0.041	0.074
5	0.979	0.021	0.133	0.867	0.021	0.133

Table 21.5 Reliability parameters of segments

i	A_i	$R_i(t)$	MTTFF_i
1	0.877	$e^{-0.014t}$	71.429
2	0.947	$e^{-0.007t}$	142.857
3	0.692	$e^{-0.02t}$	50.000
4	0.643	$e^{-0.041t}$	24.390
5	0.864	$e^{-0.021t}$	47.619

Fig. 21.6 Reliability structure of selected track system

We can get the state transition probability $P_{jk}^{i}(t)$ and parameters $\lambda_i, \mu_i (i = 1, 2, 3, 4, 5)$ according to the statistical result, as shown in Table 21.4.

The steady-state availability A_i, reliability $R_i(t)$, mean time to first failure MTTFF_i of segment S_i are calculated, as is shown in Table 21.5.

The selected track system is a serial structure system, and its reliability structure is shown as Fig. 21.6.

Therefore, the steady-state availability A, reliability $R(t)$, mean time to first failure MTTFF of the system are calculated by

$$\begin{cases} A = \left(1 + \sum_{i=1}^{n} \frac{\lambda_i}{\mu_i}\right)^{-1} = 0.425 \\ R(t) = e^{-\Lambda t} = e^{-0.103t} \\ \text{MTTFF} = \frac{1}{\Lambda} = 9.709 \end{cases} \tag{21.8}$$

21.5 Conclusions and Prospects

This paper develops reliability models for RS-TS, including reliability structure models and reliability assessment models. The research contents include:

1. We analyze the track degradation characteristics and factors affecting the reliability of a track system.

2. In the reliability structure models for RS-TS, we study three reliability structure relationships, which include serial structure, quasi-parallel structure and hybrid structure, among track segments and the track system.
3. In the reliability assessment models for RS-TS, we define, the failure categories of RS-TS and calculate the reliability index of track segments and a serial structure track system with the method of Markov process.
4. We select the track inspection data of Beijing Subway Line 13 to validate the models developed in this paper.

However, there are still some questions worth deeply studying:

1. This paper only proposes the calculating method of reliability index of RS-TS in serial structure. The calculating method of reliability index of a quasi-parallel structure track system and a hybrid structure track system needs to be further studied.
2. The reliability structure relationships among track facilities and a track segment need to be further studied.

References

1. EN50126 (1999) Railway applications—the specification and demonstration of reliability, availability, maintainability and safety (RAMS)
2. Nguyen MN, Wang X, Wang CH (2012) A reliability assessment of railway track buckling during an extreme heatwave. In: Proceedings of the institution of mechanical engineers. Part F: J Rail Rapid Transit 226(5):513–517
3. Zhao J, Chan AHC, Burrow MPN (2006) Reliability analysis and maintenance decision for railway sleepers using track condition information. J Oper Res Soc 58(8):1047–1055
4. Zheng J (2002) Analysis of rail reliability. Southwest Jiaotong University, Leshan, p 61 (in Chinese)
5. Li B (2011) Research of random vibration of track structure and reliability of sleeper III. Southwest Jiaoting University, Leshan, p 146 (in Chinese)
6. Zhang X, Chen X (2007) Calculation of reliability of CWR track stability and parameter sensitivity analysis. J China Railway Soc 04:70–73 (in Chinese)
7. Kang X, Wang W, Liu J (2013) Research on comprehensive evaluation system for track irregularity of high-speed railway based on RAMS. China Railway Sci 02:13–17 (in Chinese)
8. Wang F, Liu R (2013) Theory and realization technique of high-speed railway grid management. China Railway Publishing House, Beijing (in Chinese)
9. Gao J, Zhai W, Xu Y (2009) Analysis of management length of track irregularity sections on existing main lines. J Railway Eng 05:105–108 (in Chinese)
10. Cao J (2006) Introduction of reliability mathematics. Higher education press, Beijing, p 484 (in Chinese)

Chapter 22
The Research of Vehicle Classification Method Based on the Frequency Domain Features

Yi Tang, Honghui Dong, Limin Jia, Junqing Tang and Xianpeng Xia

Abstract On the basis of the time-variable signal generated by a geomagnetic sensor, a vehicle classification method based on the fusion of time domain and frequency domain features is presented. First, mutation value removal, weighted smoothing and cubic spline interpolation method are selected as a pretreatment. Then, FFT is adapted to obtain the frequency domain features. And vehicle classification based on frequency spectrum template matching method is proposed. Finally, this method is used in practice to verify its feasibility. Example analysis proves that the vehicle classification method can make full use of the frequency domain features of the signal waveform to realize vehicle reclassification. This method has finer classification and higher classification accuracy, so it can provide new idea and method for vehicle classification based on geomagnetic sensor.

Keywords Vehicle classification · Template matching · Geomagnetic sensor · Intelligent transportation

22.1 Introduction

Road traffic information detection is a foundation to realize the intelligent control of traffic system. And it is the basic to obtain vehicle flow; vehicle speed and lane occupy rate, etc. It has a significant impact on traffic scheme and the adaptive control of signal lamp. Therefore, a simple and efficient vehicle classification method has the important significance to ease road traffic congestion and improve the operation effect of ITS management.

Y. Tang · L. Jia · J. Tang · X. Xia
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing 100044, China

Y. Tang · H. Dong (✉) · L. Jia · J. Tang · X. Xia
Beijing Engineering Research Center of Urban Traffic Information
Intelligent Sensing and Service Technologies, Beijing, China
e-mail: hhdong@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_22

There were many experts who has made many years of research in the vehicle identification at home and abroad, proposing the relevant models and identification algorithm. Stencel [1] proposed eight indicators which includes the mean, the variance and the kurtosis, etc. Gorski and Zarzycki [2] proposed a parametric method that the signal sample of high dimensional was mapped to the low dimensional feature vector space. Meta and Cinsdikici [3] proposed a vehicle classification method based on the principal component analysis and neural network. The domestic scholars selected some features, which include the peak, the mean, the variance, and the peak number [4]. In the aspect of vehicle identification algorithm, some common peak pattern algorithms have been proposed, which include support vector machine algorithm, naive bayes classification algorithm, clustering algorithm, K-nearest neighbor algorithm and BP neural network algorithm [5, 6]. At present, the study of vehicle identification is mainly researched in the time domain, and it is seldom related to the frequency domain analysis. In addition, most of the classification algorithm has the defects of large amount of computation, coarse classification and low recognition accuracy, and they cannot identify some similar vehicles.

In this paper, the geomagnetic sensor is introduced to finish data acquisition. And a vehicle classification method based on the frequency domain features is put forward. First, the data is preprocessed through some methods, which include removing mutation value, moving weighted smoothing and cubic spline interpolation. Second, the frequency domain characteristics of the vehicle can be obtained based on FFT. Then, vehicle classification based on frequency spectrum template matching method. Finally, this method is applied in practice to verify its reliability.

22.2 Frequency Domain Feature and Vehicle Classification

22.2.1 Vehicle Signal Extraction

Because of the influence of environmental conditions and frequency sampling error, the magnetic disturbance signal detected by geomagnetic sensor is incomplete and contains some noisy data, which will reduce the accuracy of classification algorithm. In this paper, the methods of removing mutation value and moving weighted smoothing are introduced to process the initial data, and extract a more appropriate data [7]. On the basis of removal of anomaly, data smoothing processing can eliminate the noisy data, make the waveform smooth. In this way the frequency domain feature can be extracted easily. The moving weighted smoothing formula is as follows.

$$Z_i' = \frac{\alpha_1 * Z_{i-2} + \alpha_2 * Z_{i-1} + \alpha_3 * Z_i + \alpha_4 * Z_{i+1} + \alpha_5 * Z_{i+2}}{(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5)} \tag{22.1}$$

where: Z_I means to the output of the geomagnetic sensor in the Z axis. α means to the smooth weight.

The sample frequency of geomagnetic sensor is FS. And the duration of the magnetic disturbance signal, which will happen when each vehicle passes the geomagnetic sensor, is t. And the length of the waveform is $L\,(L = \mathrm{FS} \times t)$.

Because vehicles have different speed when they pass geomagnetic sensor, the length of the waveform is not equal. Cubic spline interpolation and data sampling method is introduced to make the waveform signal have the same sequence length. In the interval $[t_1, t_L]$, amplitude of disturbance signal is defined as $\left[Z'_1, Z'_2, \ldots Z'_L\right]$. In order to calculate easily, the sampling point is defined as $k = 2^j\,(j = 1, 2, 3\ldots)$. The new data sequence $(s = [s_1, s_2, s_3 \ldots s_k])$ is obtained through Cubic spline interpolation and data sampling. After data preprocessing, frequency domain feature extraction can be completed more easily.

22.2.2 Frequency Domain Feature Extraction

The complex signal is composed of many harmonic signals with different frequency. And the intensity and phase of each harmonic signal will influence the whole characteristic of the final signal. Fourier transform is a signal analysis method, which can transform a signal from the time domain to the frequency domain. It is difficult to distinguish some signals in the time domain, but it can be easy in the frequency domain. Each point in Fourier transform refers to a frequency point. There is a new data sequence $[s_1, s_2, s_3 \ldots s_k]$, which is obtained by processing a vehicle signal waveform based on cubic spline interpolation method. This data sequence can be changed to an amplitude vector, $[C_1, C_2, C_3 \ldots C_k]$, by fourier transform. The amplitude in the spectrogram can be defined as a frequency domain features. In this paper, FFT is introduced to convert vehicle waveform signals from time domain to frequency domain. The spectrogram, which is concerned to 50 suvs and 40 microbuses, is showed in Figs. 22.1 and 22.2.

There is no obvious difference between suvs and microbuses. The average amplitude template can be obtained by calculating the average amplitude of the suvs and microbuses. There are obvious differences in two kinds of amplitude in interval [4, 14]. And there does exist the cluster in two kinds of spectrograms, which leads a good discrimination (Fig. 22.3).

Therefore, the spectrum relations, which are obtained by processing vehicle signal waveform through FFT, can be considered as the frequency domain features. Each kind of vehicle has own features. There is a good discrimination in corresponding frequency range of the spectrogram.

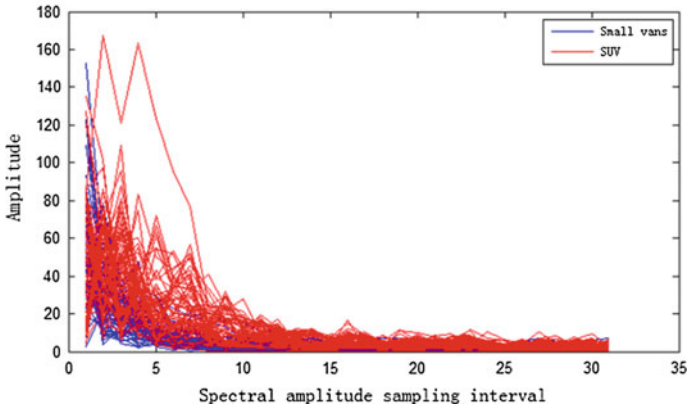


Fig. 22.1 Spectrum of suv and small van

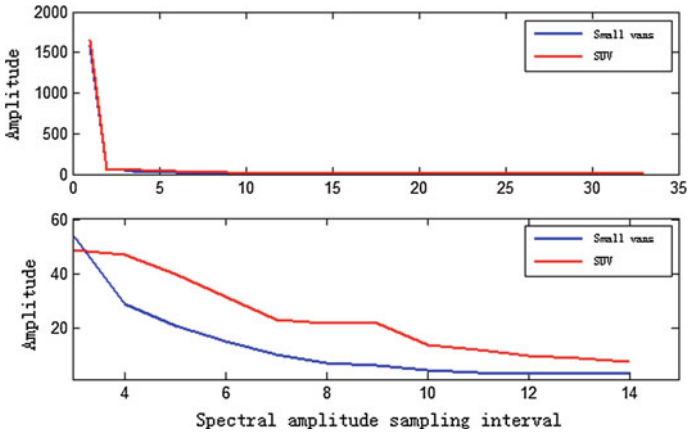


Fig. 22.2 The average spectrum of suv and small van

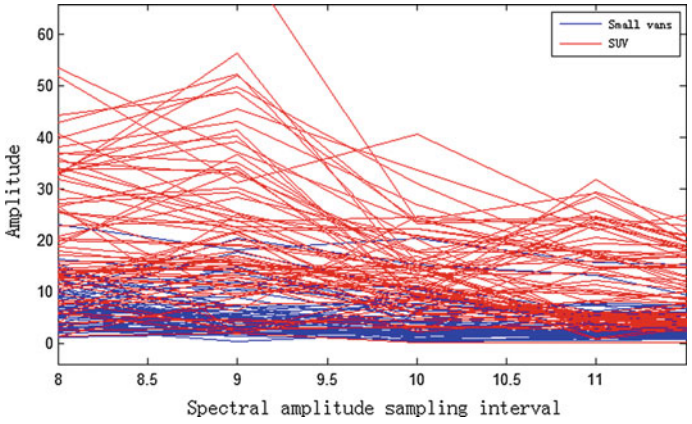


Fig. 22.3 The spectrum of suv and small van in the interval [8.5, 11]

22.3 Vehicle Classification Algorithm

The frequency domain feature of each signal can be obtained by FFT, which can determine each kind of the average amplitude templates. In this way, the vehicle signal which is waiting for recognition can be matched to the existing templates as long as do some simple processing. In the process of matching, both the amplitude vector from the existing templates and the vehicle signal which is waiting for recognition should have the same coordinate extents. Otherwise, the result of recognition will have some deviations. The vehicle classification based on frequency spectrum template matching method goes as follows.

Step 1: The spectral features of all the vehicles are obtained by FFT. The average amplitude of each training sample sets is consider as the amplitude template vector, $A_{\mathrm{I}} = (a_{i1}, a_{i2} \ldots a_{im})\,(i = 1, 2, 3 \ldots)$ whose length is m. a_{im} is the average amplitude which is concerned to the vehicle type I and the position m.

Step 2: The dynamic region search algorithm (DRSA) is used to find optimal sampling area of each different vehicle type. The optimal sampling area is defined as $H = [e,f]\,(f \geq e)$. The entire subinterval can be traversed through DRSA, which can guarantee complete the optimal search process in the acceptable time.

Step 3: The amplitude vector corresponding to the optimal sampling interval for different vehicle types is obtained through step 2. And it is defined as the standard feature vector, $B_i[e,f] = \left(a_{ie}, a_{ie+1} \ldots a_{if}\right)$. The amplitude vector of the vehicle signal which is waiting for recognition is $C_i[e,f] = (c_{ie}, c_{ie+1} \ldots c_{if})$. The optimal matching results can be obtained by using template matching method, which realizes the purpose of automatic identification. The template matching methods are as follows.

(1) Absolute value difference algorithm: The distance between amplitude vector of vehicle type waiting for recognition (C_j) and amplitude template (B_i) is $D(i,j) = \left\|B_i - C_j\right\|_2$. Euclidean distance is used to represent the correlation of the two waveforms. And the minimum value principle is used to determine the vehicle classification.

(2) Optimal energy point algorithm: The total energy of the optimal classification sampling interval for each vehicle is defined as C_j. The energy of the vehicle type waiting for recognition is defined as $E_j = \sum_{k=e}^{f} \left|c_{jk}\right|^2$. The vehicle type can be obtained based on similar principle.

(3) Improved absolute value difference algorithm: Even though two amplitude vectors are approximate, there are large differences in a few individual sampling points. The two methods are easy to be mistaken for the two originally matched waveforms. Therefore, in the basis of absolute value difference algorithm, an absolute threshold (T) refers to amplitude difference is added. If the two point difference is less than or equal to the

threshold T, the two sampling points are matched. The correlation of the two waveforms is represented by the matched logarithm of the sampling points. And the best matching template is determined by maximum principle [8].

$$\forall k \in [1,m], \quad \text{Numbers}(i,j) = \sum_{k} F\left[X_i(k), Y_j(k)\right] \tag{22.2}$$

$$F\left[X_i(k), Y_j(k)\right] = \begin{cases} 1, & \left|X_i(k), Y_j(k)\right| \le T \\ 0, & \left|X_i(k), Y_j(k)\right| > T \end{cases} \tag{22.3}$$

22.4 Algorithm Verification

The experimental data come from 174 cars, 76 suvs, 61 small vans, and 21 medium trucks. The sampling frequency of geomagnetic sensor is 100 Hz. The signal duration when each vehicle passes through the geomagnetic sensor is about 0.6–1.4 s, which contains more than 100 data points. Firstly, the smoothing weight is valued in 0.3, 0.7, 1, 0.7, and 0.3. The sampling point is selected in 8, 16, 32, 64, and 128. Cubic spline interpolation is used to obtain a sample signal waveform of a car as shown in Fig. 22.4.

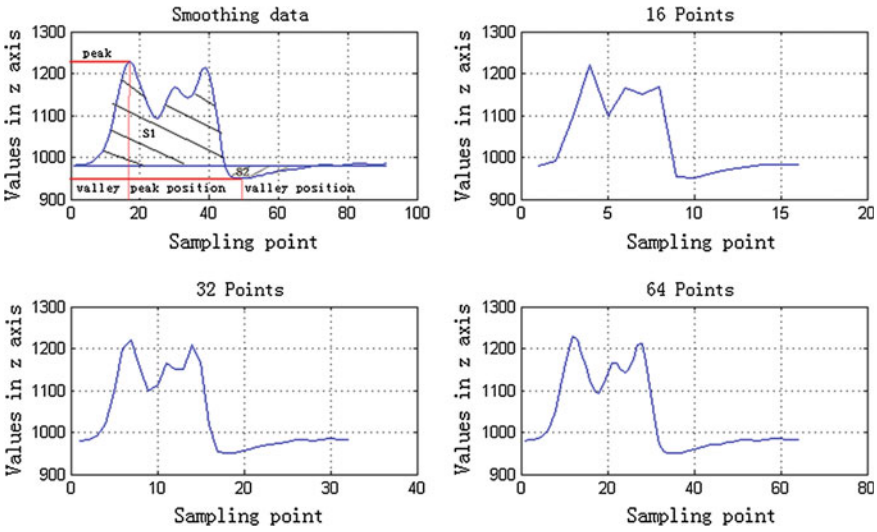


Fig. 22.4 Data sampling comparison chart

Table 22.1 The classification results of the template matching

Vehicle types	Running time (s)	Optimum interval	Accuracy (%)
Cars and suvs	0.712	(1,2)	95.20
Cars and small vans	0.67	(1,5)	89.79
Cars and medium trucks	0.553	(3,3)	92.82
Suvs and small vans	0.392	(7,30)	89.78
Suvs and medium trucks	0.281	(1,1)	89.69
Small vans and medium trucks	0.238	(8,8)	86.59

Table 22.2 The template matching result of suv and small van

Template matching	Input vehicle types		Total
	Suvs	Small vans	
Suvs	21	2	23
Small vans	5	19	24
Total	26	21	47
Discrimination	0.808	0.905	0.851

Based on the analysis of the waveform signal of a large number of sample vehicles, it is obvious that some data information will be lost in 16 or 32 sampling points. And the original waveform signal cannot be completely displayed. However, useful data information is completely displayed in 64 or 128 sampling points. Considering the length of the data set, 64 sampling points are selected.

332 cars are divided into two categories, training set and test set. The optimal classification sampling interval of any two vehicle types from training set can be obtained by DRSA. Then, the test set is classified by the template matching method based on the improved absolute value difference algorithm. The operating time and classification accuracy of the template matching are shown in Table 22.1.

It is shown that the optimal sampling interval for different vehicle types is inconsistent in Table 22.2. Suvs and small vans can be well identified by using the template matching method. And the optimal interval is [7, 30], which is consistent with the interval which is showed in Fig. 22.3. There is a set of experimental data referring to recognition of suv and small van shown in Table 22.2.

It is obvious that suv and small van are well identified through the template matching, which is just the result that we expected.

22.5 Conclusion

In this paper, a vehicle classification method based on the frequency domain features is put forward. In the aspect of frequency domain feature, FFT is introduced to extract frequency domain features. On this basis, optimal sampling interval of any two vehicle types is obtained by DRSA. And vehicle classification is realized by

using spectrum template. Finally, this method is used in actual experiment to verify the method is reliable and practical in the research of vehicle classification.

The research still has great room for improvement. Because, all work we do is identification of any two vehicle types. Further work will be carried out in study of identification of multiple vehicle types.

Acknowledgments This work was supported by Beijing New-star Plan of Science and Technology (Z1211106002512027) and the National Science and Technology Pillar Program of China (2014BAG01B02).

References

1. Stencel M (2004) Signal parameterization vs orthogonalization on example of vehicle's magnetic signature recognition. In: Proceedings of the 21st IEEE instrumentation and measurement technology conference IMTC04, vol 2, pp 1416–1418
2. Gorski M, Zarzycki J (2012) Feature extraction in vehicle classification. In: 2012 international conference on signals and electronic systems (ICSES), pp 1–6
3. Meta S, Cinsdikici MG (2010) Vehicle-classification algorithm based on component analysis for single-loop inductive detector. IEEE Trans Veh Technol 59(6):2795–2805
4. Haijian L, Honghui D, Yuanchao S, Limin J, WeiFeng G (2015) Vehicle classification with a single magnetic sensor for urban road. J Jilin Univ Eng Technol Ed 45(1):97–103 (in Chinese)
5. Bo Y, Wenbin N (2013) Vehicle detection and classification algorithm based on anisotropic magnetoresistive sensor. Chin J Sci Instrum 34(3):537–544 (in Chinese)
6. De Luca A, Lanari L, Oriolo G (1991) A sensitivity approach to optimal spline robot trajectories. Automatica 27(3):535–539
7. Hong C, Hai L (2013) T analysis of vehicle driving data based on cubic spline interpolation. Automob Technol 8:54–57
8. Weng G, He T, Chen M (2004) The vehicle's classification recognition system based on DTW algorithm. In: Fifth world congress on intelligent control and automation 2004, WCICA 2004. IEEE, New York, vol 5, pp 4169–417

Chapter 23
Experimental Research on Fiber Bragg Gratings High-Speed Wavelength Demodulation Method

Lijun Meng and Fang Zhu

Abstract In order to detect weak signals of high frequency with fiber Bragg gratings (FBG), the FBG demodulation system based on matched filter is established. First, the output expression and maximum wavelength sensitivity as well as the main way to improve its demodulation property were analyzed in detail. On these bases, the FBG-matched filter demodulation device was set up to measure the vibration signals of an ultrasonic cleaning tank. A SLED broadband light source was used to provide light energy; an amplified InGaAs photo detector was used to measure the optical power; a digital oscilloscope was employed to display and record the output signals. The experiment results showed that ultrasonic signal with main frequency of 43.945 kHz and strain range of about 55με has been achieved.

Keywords FBG · Matched filtering demodulation · Resolution · Ultrasonic wave

23.1 Introduction

Fiber Bragg gratings (FBG) is a new kind of optical passive component which is developing most rapidly in recent years. It has been used in many fields such as civil engineering, aerospace, shipping, electric power industry, etc [1]. FBG is used to measure the external parameters such as stress, temperature, displacement, vibration, and acceleration according to its reflection wavelength change. In practical application, it is one of the key technologies to obtain the reflection wavelength change quickly and accurately.

L. Meng (✉)
School of Electromechanical and Architectural Engineering,
Jianghan University, No. 8 Delta Lake Road, Wuhan, Hubei, China
e-mail: menglijun0408@163.com

F. Zhu
Department of Aircraft Maintenance and Engineering,
Guangzhou Civil Aviation College, Guangzhou, China
e-mail: zhufang@caac.net

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_23

The main FBG wavelength demodulation methods are interference method, filtering method, and chirped grating detection method [2–7]. However, the universal shortcomings for existing wavelength demodulation methods are relatively low demodulation frequency and not high demodulation resolution. In some FBG detection fields that require higher demodulation speed and resolution such as the structural damage detection with ultrasonic wave, high-frequency blade flutter measurement, submarine hydroacoustic detection, petroleum and natural gas exploring, and seismic geophone testing, ordinary FBG demodulation system cannot meet the requirements.

As the traditional FBG demodulation technology is relatively mature, it can increase the demodulation frequency and resolution by improving the appropriate equipment parameters. In this paper, the principle and the output characteristics of the transmission-matched filter demodulation technology were analyzed first. Based on this, the demodulation performance of this system was improved by increasing the light source intensity and the detector response. Finally, a matched filter wavelength demodulation system was established and applied to detect the vibration signal of ultrasonic cleaning tank. In the experiment, it measured the ultrasonic signal with the main frequency of 43.945 kHz successfully.

23.2 Demodulation Principle of Matched Filter Method

FBG-matched filter demodulation systems based on broadband light source are divided into two types: reflection type and transmission type. Owing to the relatively large S/N (Signal-to-noise ration) of the reflection one and the high luminous energy utility rate of the transmission type, we mainly analyzed the transmission type-matched filter demodulation method in this paper.

Figure 23.1 shows the structure of matched filter demodulation system for transmission type. The sensing FBGs were joined to a 3 dB coupler, and then the 3 dB coupler was connected to two parts: one to a broadband light source; another one first to a matched FBG, and then was linked to a photodetector which converted the optical power signal into electrical signals; finally, the electrical signals were

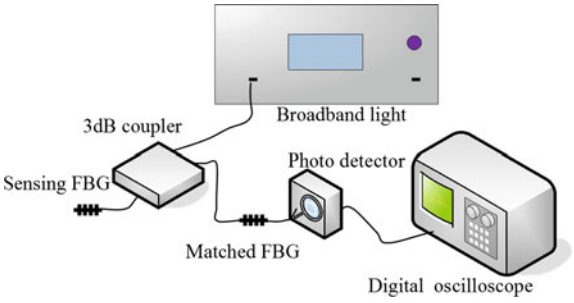


Fig. 23.1 Structure of matched transmission system based on broadband light source

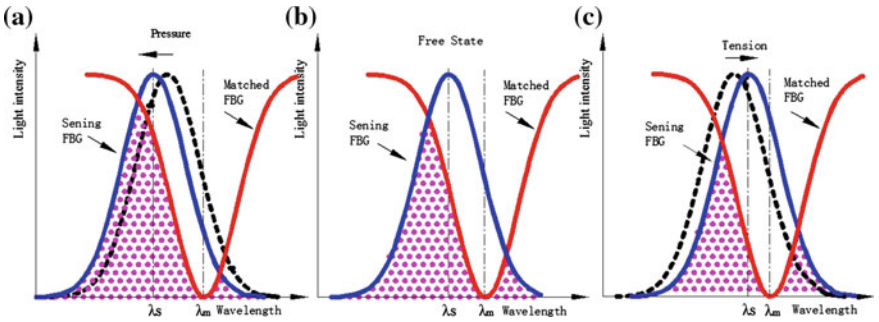


Fig. 23.2 Demodulation principle of matched transmission system based on broadband light source

displayed on the digital oscilloscope or collected by a data acquisition card for computer processing.

The demodulation principle is shown in Fig. 23.2 if that the initial wavelength of sensing FBG is slightly shorter than the wavelength of matched FBG. When the sensing FBG sustains tension, its center wavelength will increase compared to the initial state, so as to reduce the sectional area (shaded area) formed by the spectral lines of sensing FBG and matched FBG, as well as horizontal ordinate; in other words, the light power reaching photo detector is reduced. On the other hand, if the sensing FBG bears pressure, its wavelength will decrease contrasted with the initial state, so the corresponding sectional area (shaded area) expands, which means the luminous power arriving at photo detector is increased. The center wavelength change of the FBG can be achieved by detecting the output signal change of the photo detector. The demodulation principle is similar to this, if the initial wavelength of sensing FBG is slightly longer than the wavelength of matched fiber grating.

23.3 Demodulation Performance Analysis

According to the reference documentation [8] which gave the analysis about the output power spectrum curve of the transmission type-matched filter demodulation system, the output voltage of this system is obtained:

$$U_T = KIA_1A_2R_sR_m\frac{\sqrt{\pi}}{2\sqrt{\ln 2}}\Delta\lambda_s \cdot \left\{1 - \frac{\Delta\lambda_m}{(\Delta\lambda_s^2 + \Delta\lambda_m^2)^{\frac{1}{2}}} \times \exp[-4\ln 2 \cdot \frac{(\lambda_s - \lambda_m)^2}{\Delta\lambda_s^2 + \Delta_m^2}]\right\} \tag{23.1}$$

where K is the luminous energy utility rate of the coupler. A 3 dB optical coupler was used in the system, so K = 1/4, if there was no light loss in the transmission

process; I is the incident light intensity of the broadband light source at its center wavelength; A_1 is the transimpedance gain for the photo detector; A_2 is the voltage current conversion factor for the photo detector; R_s, λ_s, $\Delta\lambda_s$ are the sensing FBG reflectivity at its center wavelength, center wavelength, and full width at half maximum (FWHM), respectively; R_m, λ_m, $\Delta\lambda_m$ are the matched FBG reflectivity at its center wavelength, center wavelength, and FWHM separately.

Wavelength measurement resolution can be considered as the wavelength change per unit voltage output, so according to Formula (23.1) and the reference [9], the maximum wavelength measurement resolution of this system is

$$\begin{aligned}\delta\lambda &= \frac{\delta U_T}{(\mathrm{d}U_T/\mathrm{d}\lambda_s)T} \\ &= \sqrt{\frac{e}{2\pi}}\frac{\delta U_1}{TKIA_1A_2R_sR_m}\frac{|\Delta\lambda_s^2+|\Delta\lambda_m^2}{|\Delta\lambda_s|\Delta\lambda_m}\end{aligned} \tag{23.2}$$

where T is the normalized amplitude of the signal.

From the demodulation system structure shown in Fig. 23.1, the frequency response characteristic of the photoelectric detector needs to be improved for a higher demodulation speed. The main materials for the photo detector are InSb, HgCdTe, PtSi, and InGaAs/InP presently. InGaAs/InP has the advantages of direct band gap, room temperature operation, and high purity, and the photodiodes made of this material have a very low dark current and noise. As a result, it is commonly used in the photoelectric detection device [10].

The electrical signal directly converted from the optical signal reaching photoelectric is very weak, and it is easy to be submerged in the noise, so it is necessary to design the corresponding preamplifier circuit. The photoelectric detector preamplifier circuit improves the photoelectric conversion sensitivity while decreases the frequency response of the photoelectric conversion.

In addition, from Formula (23.2), the wavelength measurement resolution of the demodulation system is influenced by the output voltage resolution, the light utilization rate of the system, transimpedance gain A_1, voltage current conversion factor A_2, the optical intensity, and the FBG peak reflectivity. So the main methods to improve the system resolution are as follows: choose a broadband light source of high power, select FBGs with high reflectivity, and improve the transimpedance gain and voltage current conversion factor of the optical detector. Improving the output power of the light source and the voltage current conversion coefficient of the photo detector are the effective measures in practice.

According to the above analysis, in order to improve the demodulation speed and resolution of the matched filter demodulation system, the broadband light with high power was chosen to increase light energy, and the high bandwidth InGaAs photodiode was selected as a photo detector in this paper.

23.4 Experimental Studies

A matched filter demodulation system was constructed and used to measure the vibration signal of ultrasonic cleaning tank in this paper, and the experimental configuration is shown in Fig. 23.3. A piezoelectric ceramic plate was pasted on the bottom center of the tank's downward surface, and the piezoelectric ceramic was subjected to a sinusoidal alternating current with amplitude of 220 V and frequency of 43 kHz. A sensing FBG was glued at the bottom upper surface of the ultrasonic cleaning tank, 4 cm away from its bottom center to measure the vibration signal of the ultrasonic cleaning tank. The parameters of the sensing FBG were as follows: initial wavelength λ_s = 1308.791 nm, FWHM $\Delta\lambda_s$ = 0.485 nm, reflectivity R_s = 0.9, and FBG length was 3 mm. Matched FBG was fixed on a flat paper board, and its main parameters were λ_m = 1308.890 nm, FWHM $\Delta\lambda_m$ = 0.485 nm, reflectivity R_m = 0.9, and FBG length was 3 mm. In the test process, the paper board was kept in static state and was not affected by external force.

The broadband light source used in the experiment was an output power-adjusted SLED light made by OPEAK, with a center wavelength of 1322.8 nm and FWHM of 48.5 nm. During the test, the output power of SLED light source was set to 9 mW, and its output light intensity was I = 0.1743 mW/nm based on the literature [8]. The InGaAs photodiode with an adjustable gain was selected as a photo detector (Thorlabs-PDA10CS). In the experiment, the photo detector gain was set to 30 dB, with the bandwidth of 775 kHz, noise equivalent power of 1.25e−12 W$\sqrt{\text{Hz}}$, which met the frequency requirements of ultrasonic measurement. Transimpedance gain A_1 and voltage current conversion factor A_2 for the photo detector were A_1 = 0.913 A/W and A_2 = 2.38e4 V/A, respectively. The digital oscilloscope (DS1102E RIGOL) was used to display the signal, its bandwidth was 100 MHz, the real-time sampling rate was 1GSa/s, and the output voltage resolution δU_1 was 1 mV.

It can be known from the demodulation system structure that the demodulation speed of the system can reach 775 kHz. The optimal wavelength resolution of the system was 2.4 pm in theory according to Formula (23.2). The theoretical output

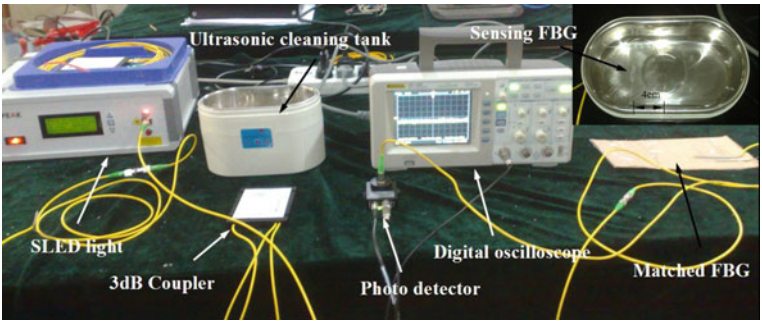


Fig. 23.3 Experimental setup of matched filter demodulation system

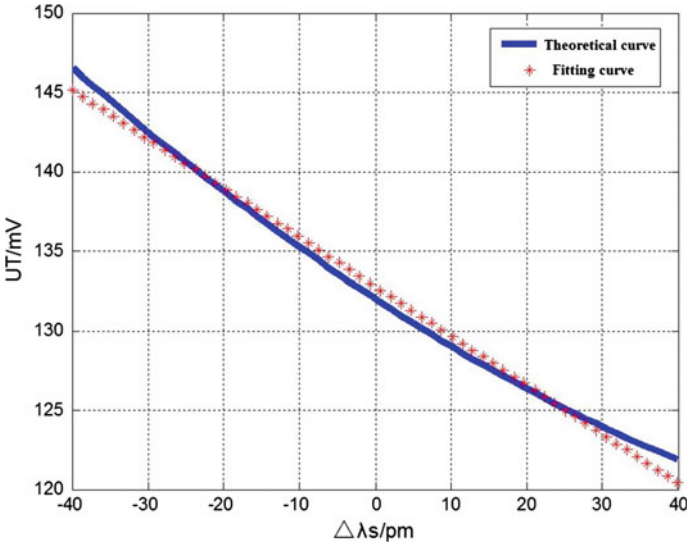


Fig. 23.4 Output characteristic of matched filter demodulation system

characteristic curve is shown in Fig. 23.4. It can be seen that the output voltage signal had good linear relationship with the center wavelength when sensing FBG wavelength change was relatively small. The sensitivity of the system was 0.309 mV/pm, and the system linearity was 0.94 when the measurement range was 40 pm.

The vibration signal of the ultrasonic cleaning tank measured by the demodulation system is shown in Fig. 23.5, where the figure (a) was the time-domain signals of the digital oscilloscope, and the figure (b) was the frequency signal obtained by FFT transform. It can be seen from Fig. 23.5 that the ultrasonic signal was approximately sinusoidal signal, with the main frequency of 43.945 kHz and the peak value of 17 mV. According to preceding analysis about the system sensitivity, the center wavelength change range of the sensing FBG was

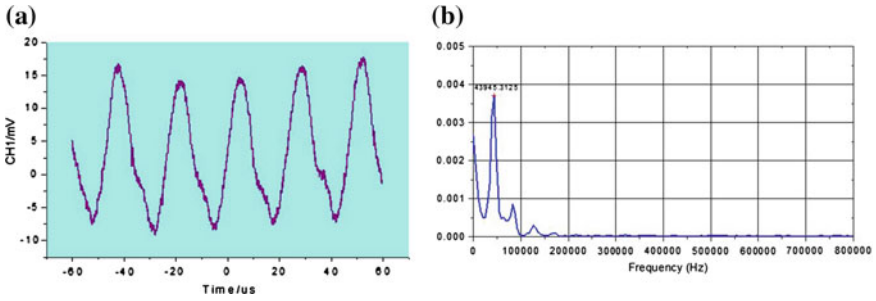


Fig. 23.5 Ultrasonic wave-cleaning tank signals detected by the demodulation system. **a** Time domain signal. **b** FFT signal

17/0.309 = 55.0162 pm. Due to the high ultrasonic vibration frequency, the temperature influence on the FBG wavelength change measurement can be ignored, so the strain measurement sensitivity of the FBG was about 1 pm/με. It can be inferred that strain range during ultrasonic cleaning tank vibration was about 55 με.

23.5 Summary

In this paper, the working principle, output characteristic, and measurement resolution of FBG-matched filter demodulation system were analyzed, and the main approaches to improve the demodulation speed and sensitivity were discussed. The testing equipments of the matched filter demodulation system were built, and they were used to measure the vibration signal of 43.945 kHz about an ultrasonic cleaning tank. The results showed that the system had high demodulation speed and good wavelength resolution, which can meet the need of high-frequency weak signal measurement.

Acknowledgment This paper is funded and supported by the New Doctoral Research Projects of Jianghan University No. 1006-06550001.

References

1. Lin J, Wang W, Wang X (2004) Study of application and evolution of fiber grating sensor technique. J Dalian Univ Technol 44(6):931–936 (in Chinese)
2. Kersey AD, Berkoff TA, Morey WW (1992) High resolution fiber-grating based strain sensor with interferometric wavelength-shift detection. Electron Lett 28(03):236–238
3. Liu B, Sun G et al (2005) Application for high birefringence fiber sagnac loop mirror filter of fiber bragg grating sensing system biomolecular interactions. Acta Scientiarum Naturallum (Universitatis Nakaiensis) 38(05):9–12 (in Chinese)
4. Ferreira LA (1997) Pseudoheterodyne demodulation technique for fiber Bragg grating sensors using two matched gratings. IEEE Photonics Technol Lett 9(4):487–489
5. Davis MA (1995) Matched-filter interrogation technique for fiber Bragg gratings array. Electron Lett 31(10):822–823
6. Seim J, Schulz W, Udd E et al (1998) Low cost, high speed fiber optic grating demodulation system for monitoring composite structures. Proc SPIE 3326:390–395
7. Wang H, Zhang J, Qiao X, et al (2007) Prospect of research for signal demodulation in fiber Bragg grating sensing system. Semicond Technol 32(03):188–192 (in Chinese)
8. Lobo AB, Ferreira LA, Santos JL et al (1997) Analysis of the reflective-matched fiber Bragg grating sensing interrogation scheme. Appl Opt 36(7):934–939
9. Liu Y, Liu Z, Guo Z, Dong X (2001) Theoretical and experimental researches on tunable filtering detection technique of fiber grating senso. Acta Optica Sinica 21(1):88–92 (in Chinese)
10. Li T, Wang Y, Li YF et al (2010) Investigation on dark current and low frequency noise of mesa type InGaAs infrared detector. J Optoelectron Laser 21(4):500–503 (in Chinese)

Chapter 24
Test Pattern Generation for Multiple Victim Lines of Crosstalk Effect in Digital Circuits by Binary Decision Diagram

Zhongliang Pan and Ling Chen

Abstract The continual increase of the integrated circuit complexity results in the more signal lines close to each other; this may produce the coupling effects among the signal lines. One of main coupling effects is crosstalk effect. In this paper, a new method is presented for the test pattern generation for the multiple victim lines of crosstalk effects; the method makes use of the binary decision diagrams (BDD). First, the BDDs of the circuit under test are built to express the circuit structure and logic values related to crosstalk effects. Second, the test patterns of crosstalk faults are produced by performing the operations on these BDDs and searching for the input assignments of the decision diagrams. The experimental results show that the method proposed in this paper can produce the test patterns of crosstalk effects in shorter time.

Keywords Digital circuits · Crosstalk effects · Multiple victim lines · Test pattern generation · Binary decision diagram

24.1 Introduction

The complexity of integrated circuits (IC) and the circuit speeds are continually increasing with the progress in the IC process technology that makes the interconnect feature size down to nanometer scale. The small physical size of IC makes the signal lines coupling and the signal integrity increasing in the circuit design [1]. The crosstalk is one of coupling noises due to the mutual capacitance and inductance between the interconnect signal lines; the crosstalk may result in the IC functional errors [2]. For the given two signal lines, when the signal transit on a line generates coupling effect on other line, the line is called aggressor line and the other line is called victim line.

Z. Pan (✉) · L. Chen
School of Physics and Telecommunications Engineering, South China Normal University, Guangzhou 510006, China
e-mail: panz@scnu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_24

Performing the analysis of the crosstalk faults is needed to acquire the characteristics of crosstalk. In this aspect, Yao et al. [3] presented an input-aware statistical timing analysis approach which was based on signal probability; the approach was used to tackle the crosstalk-induced delay. Kumar et al. [4] investigated the approach to model the crosstalk effects in the CMOS-gate-driven coupled RLC interconnects, and the approach utilized the *n*th power law model and the time-domain finite difference. Kavicharan et al. [5] discussed the estimation models of the delay and crosstalk effects for high speed interconnects in VLSI circuits, a computation approach of finite ramp responses for the current mode resistance, inductance, and capacitance interconnects was proposed. Kaushik et al. [6] proposed a circuit design scheme of the bus encoder for the reduction of the crosstalk effects, power dissipation, and area in the VLSI interconnects, where the bus invert strategy was used to decrease the signal coupling. Fatemi and Tehrani [7] discussed the crosstalk timing analysis and made use of the timing window overlap for the coupled signal lines.

The main technique for tackling the crosstalk faults is performing the test; in this aspect, Pour and Mousavian [8] presented a parallel built-in self-test (BIST) method that can detect the crosstalk faults on the inter-switch links in network on chip. Hasan et al. [9] investigated the test pattern generation approaches which simultaneously consider the coupling capacitance, timing, and functional incompatibilities between the victim lines and the aggressor lines. Bengtsson et al. [10] proposed test approaches for the glitch faults and the crosstalk-induced delay in network on chip which can implement the asynchronous communication protocol. The BIST hardware implementation was given for the delay test method represented at register transfer level. Shumin et al. [11] investigated the effects of the glitch and crosstalk-induced delay in system-on-chip interconnect bus, and presented the methods to detect them, where a pulse detector having an adjustable detection threshold was designed and used. Sanyal et al. [12] discussed the test pattern generation for the multiple aggressor crosstalks under the gate leakage being loaded in the presence of gate delays, and proposed a test pattern generation approach which utilizes the 0–1 integer linear programming strategy.

In general, an aggressor line may produce coupling on several adjacent victim lines; therefore, it is needed to consider the situation of multiple victim lines. In this paper, a new method is presented for the test pattern generation of the crosstalk faults with multiple victim lines using binary decision diagram (BDD).

24.2 Binary Decision Diagrams

The binary decision diagram (BDD) is an efficient data structure for the representation of logic Boolean functions [13]. Suppose the Boolean functions have the variables $x_1, x_2, \ldots, x_n$. The BDD of a Boolean function can be constructed by the following procedure. First, construct a decision tree for the given Boolean function, such that, along any path from the root node to leaf node, no variable appears more

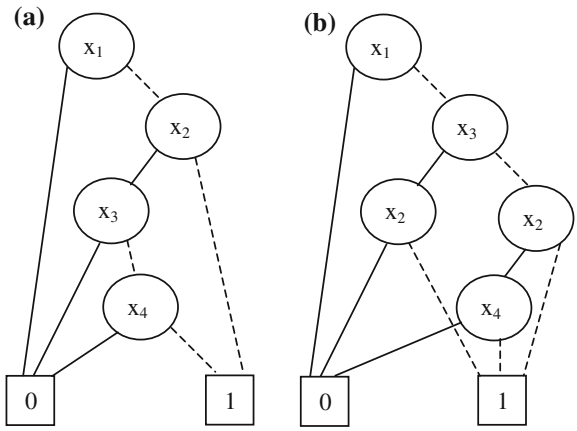


Fig. 24.1 The BDD of function g

than once, and along every path from the root node to leaf node, the variables must appear in the same order. Second, perform the following two reduction strategies repeatedly: (a) Merge the nodes that have same label (variable) and same children nodes; and (b) Delete redundant nodes. If both children pointers of a given node point to the same children nodes, then the node is redundant, and it should be deleted. For example, the BDD is given in Fig. 24.1a for the logic Boolean function $g = \bar{x}_1\bar{x}_2 + \bar{x}_1\bar{x}_3\bar{x}_4$.

The BDD is a directed acyclic graph; there are two types of nodes: leaf nodes and nonterminal nodes [14]. A leaf node in the BDD is labeled by the 0 or 1 which is corresponding to the value of the Boolean function. A nonterminal node is labeled by a variable, and it has two edges which direct toward two children nodes; one of the edges is named as 0-edge which is corresponding to the case where the variable is assigned the value 0, and another edge is named as 1-edge which is corresponding to the case where the variable is assigned the value 1. In Fig. 24.1, the 0-edge is shown by a dashed line, and the 1-edge is shown by a solid line.

The number of nodes in the BDD being corresponding to a given logic Boolean function depends on the variable ordering. For example, for the Boolean function $g = \bar{x}_1\bar{x}_2 + \bar{x}_1\bar{x}_3\bar{x}_4$, the BDD of g is shown in Fig. 24.1a if the variable ordering is $x_1 < x_2 < x_3 < x_4$; the number of nodes in the variable ordering is 6. The BDD of g is shown in Fig. 24.1b if the variable ordering is $x_1 < x_3 < x_2 < x_4$; the number of nodes in the variable ordering is 7.

24.3 Test Pattern Generation of Crosstalk Faults

The crosstalk is the effect of the coupling capacitance between an aggressor line and a victim line. The strength of the crosstalk depends on the transitions and states of the aggressor line.

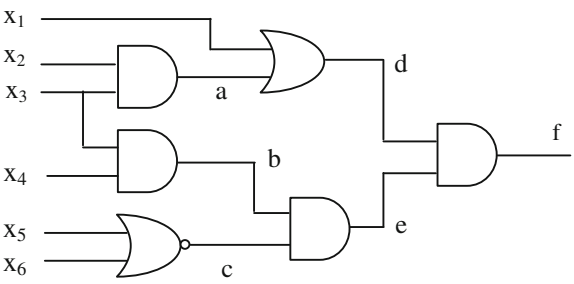


Fig. 24.2 An example of circuit

24.3.1 Test Vector of Crosstalk Fault with Multiple Victim Lines

The crosstalk faults in VLSI circuits can be detected using the test vectors. A test vector of a crosstalk fault is a pair of circuit input vectors and consists of two circuit input vectors. Utilize V to denote a test vector, a crosstalk fault $V = (V_1, V_2)$, where V_1 and V_2 are the circuit input vectors. The V_1 is called as the first vector; the V_2 is called as the second vector. The function of the test vectors is as follows: when we apply the test vectors to the normal circuit and to the faulty circuit, respectively, this will result in that the normal circuit behavior and the faulty circuit behavior produce different values, for example, the values of primary output lines in the normal circuit and in the faulty circuit are different. The example of test vectors is given for the circuit shown in Fig. 24.2.

In the circuit shown in Fig. 24.2, suppose there be a crosstalk fault among the signal lines a, b, and c, where the line b is aggressor line, and the lines a and c are victim lines, and suppose the crosstalk fault is produced by a transition (0–1) in the line b to generate a glitch (0–1) in the lines a and c. The generation procedure of test vector is to search for the input vectors of circuit such that the vectors can detect the crosstalk fault. A test vector $V = (V_1, V_2)$ being obtained is as follows: the $V_1 = (x_1, x_2, x_3, x_4, x_5, x_6) = (0, 0, 1, 0, 1, 1)$, the $V_2 = (x_1, x_2, x_3, x_4, x_5, x_6) = (0, 0, 1, 1, 0, 1)$. Apply the V_1 and V_2 to the circuit sequentially; if the value of circuit output is $f = 0$ for V_1, and $f = 0$ for V_2, then there is no crosstalk. If the value of circuit output is $f = 0$ for V_1, and $f = 1$ for V_2, then there is crosstalk. Thus, the test vector V can detect the crosstalk fault where b is the aggressor line, and a and c are the victim lines.

24.3.2 Test Pattern Generation by Binary Decision Diagram

For a given crosstalk fault in a circuit, maybe there exist test vectors or there are no test vectors. If there are test vectors, then we can use the test algorithm to produce the test vectors. In the following, a method for the test pattern generation is given

for the crosstalk faults with multiple victim lines, and the method is based on BDD. The implementation procedure of the method consists of the following steps.

Algorithm 1

Step 1: Let the circuit D_1 be a subcircuit of the normal circuit, and it consists of the signal lines which can take effects on the values of aggressor line.
Step 2: Build the BDD of the circuit D_1. The BDD is named as DE_1.
Step 3: For the DE_1, search for the input assignments that lead to the leaf nodes with the special values 0 of aggressor line. Let the set of input assignments is DS_1.
Step 4: Let the circuit D_2 be a subcircuit of normal circuit, and it consists of the signal lines which can take effects on the values of victim line.
Step 5: Build the BDD of the circuit D_2. The BDD is named as DE_2.
Step 6: For the DE_2, search for the input assignments that lead to the leaf nodes with the special values 0 of victim line. Let the set of the input assignments be DS_2.
Step 7: Build the BDD of the circuit that the value of aggressor line is set to 1 and the value of all victim line is set to 0 for the normal circuit. The BDD is named as DE_3.
Step 8: Build the BDD of the circuit that the value of aggressor line is set to 1 and the value of victim line is set to 1 for the normal circuit. The BDD is named as DE_4.
Step 9: Carry out the XOR operation of DE_3 and DE_4. Name the BDD obtained as DE_5.
Step 10: For the DE_5, search for the input assignments that lead to the leaf node with attribute value 1. Let the set of the input assignments be φ.
Step 11: Let ψ be the intersection set of the DS_1 and DS_2. If both the sets ψ and φ are not empty sets, then there are test vectors $V = (V_1, V_2)$ for the crosstalk faults, where the V_1 can be an arbitrary vector that are chosen from the ψ, and the V_2 can be an arbitrary vector that are chosen from the φ.

It is needed to carry out the Step 4 to Step 6 in Algorithm 1 repeatedly for each of multiple victim lines. Besides, Algorithm 1 is aimed at the crosstalk faults with the transition (0–1) in the aggressor line. The similar steps can tackle the crosstalk faults with the transition (1–0).

24.4 Experimental Results

The test pattern generation method in this paper for the crosstalk faults has been implemented in C++ language, and run them on a personal computer with 3.06 GHz and 1GB main memory. We have carried out a lot of experiments for ISCAS89 benchmark circuits. Table 24.1 gives the characteristics of the sequential benchmark circuits.

Table 24.1 The ISCAS89 benchmark circuits

Circuit	PI	PO	Flip-flops	Gates	Crosstalk faults	Test time
s298	3	6	14	119	50	0.2
s344	9	11	15	160	50	0.2
s382	3	6	21	158	50	0.2
s641	35	24	19	379	50	0.4
s713	35	23	19	393	50	0.4
s820	18	19	5	289	50	0.3
s953	16	23	29	418	50	0.6
s1196	14	14	18	529	50	0.6
s1423	17	5	74	657	50	0.8

In Table 24.1, the column "Circuit" indicates the name of a benchmark circuit. The two columns "PI" and "PO" denote the number of primary inputs and primary outputs in the benchmark circuits, respectively. The column "Flip-flops" gives the total number of flip-flops in a circuit. The column "Gates" indicates the total number of gates in a circuit.

We have performed the test pattern generation using BDDs for the crosstalk faults with multiple victim lines. For the crosstalk faults in the ISCAS89 benchmark circuits, we randomly select 50 crosstalk faults from these circuits, and the test vectors are produced using Algorithm 1. In the experiments for the crosstalk faults with one victim line, first, the following five BDDs are built for each crosstalk fault, i.e., DE_1, DE_2, DE_3, DE_4, and DE_5. Second, the two sets ψ and φ are obtained by manipulating the five BDDs. In these experiments, all the 50 crosstalk faults in the circuits s298, s344, s382, s641, s713, and s820 are with two victim lines and in the circuits s953, s1196, and s1423 are with three victim lines.

The total time needed to perform the test pattern generation for all 50 crosstalk faults in a benchmark circuit is given in Table 24.1. The column "Test time" indicates the total time needed, and the time unit is minute. The time needed for the circuits s298, s344, and s382 is less than 0.2 min, for the s641 and s713 is less than 0.4 min, for the s820 is less than 0.3 min, for the s953 and s1196 is less than 0.6 min, and for the s1423 is less than 0.8 min. The experimental results also show that the variable ordering of BDD can produce the effects on the number of nodes in the BDDs, and the test time of crosstalk faults can be cut down by selecting an appropriate variable ordering.

24.5 Conclusions

The generation of test vectors for crosstalk faults is related to performing the operations for both circuit structure and a lot of signal lines. The function and structure of a circuit can be expressed efficiently by the BDDs. In this paper, a new

method is presented for the test pattern generation of crosstalk faults with multiple victim lines. The method is based on the representations and operations of several BDDs of a given circuit. The performance of the method in this paper can be enhanced further by designing better variable ordering in order to reduce the number of nodes in the BDDs.

Acknowledgements This work was supported by Guangdong Provincial Natural Science Foundation of China (No. 2014A030313441), Guangdong Province Science and Technology Project (No. 2013B090600063, No. 2014B090901005), Guangzhou Science and Technology Project (No. 201510010169), and National Natural Science Foundation of China (No. 61072028).

References

1. Jun X, Shuo W (2015) Investigating a guard trace ring to suppress the crosstalk due to a clock trace on a power electronics DSP control board. IEEE Trans Electromag Compat 57 (3):546–554
2. Majumder MK, Kaushik BK, Manhas SK (2014) Analysis of delay and dynamic crosstalk in bundled carbon nanotube interconnects. IEEE Trans Electromag Compat 56(6):1666–1673
3. Yao C, Kahng A, Bao L, Wenjun W (2015) Crosstalk-aware signal probability-based dynamic statistical timing analysis. In: Sixteenth international symposium on quality electronic design, pp 424–429
4. Kumar V, Kaushik BK, Patnaik A (2014) An accurate FDTD model for crosstalk analysis of CMOS-gate-driven coupled RLC interconnects. IEEE Trans Electromag Compat 56 (5):1185–1193
5. Kavicharan M, Murthy NS, Rao NB (2014) Efficient delay and crosstalk estimation models for current-mode high speed interconnects under ramp input. J Circuits Syst Comput 23 (6):1450082
6. Kaushik B, Agarwal D, Babu NG (2013) Bus encoder design for reduced crosstalk, power and area in coupled VLSI interconnects. Microelectron J 44(9):827–833
7. Fatemi H, Tehrani P (2013) Crosstalk timing windows overlap in statistical static timing analysis. In: International symposium on quality electronic design, pp 245–251
8. Pour R, Mousavian N (2013) A fully parallel BIST-based method to test the crosstalk defects on the inter-switch links in NoC. Microelectron J 44(3):248–257
9. Hasan S, Palit A K, Anheier W (2010) Test pattern generation and compaction for crosstalk induced glitches and delay faults. In: International conference on VLSI design, pp 345–350
10. Bengtsson T, Kumar S, Ubar RJ (2008) Test methods for crosstalk-induced delay and glitch faults in network-on-chip interconnects implementing asynchronous communication protocols. IET Comput Digital Tech 2(6):445–460
11. Shumin L, Chunglen L, Chauchin S, Jwu EC (2009) A unified detection scheme for crosstalk effects in interconnection bus. IEEE Trans VLSI 17(2):306–311
12. Sanyal A, Ganeshpure K, Kundu S (2012) Test pattern generation for multiple aggressor crosstalk effects considering gate leakage loading in presence of gate delays. IEEE Trans VLSI 20(3):424–436
13. Kvassay M, Kostolny J (2014) Minimal cut sets and path sets in binary decision diagrams and logical differential calculus. In: International conference on digital technologies, pp 179–186
14. Stergiou S, Jain J (2013) Optimizing BDDs for time-series dataset manipulation. In: Design, automation and test in Europe conference and exhibition (DATE), pp 1018–1021

Chapter 25
Test Vectors Generation for Crosstalk Coupling Delay Faults by Boolean Satisfiability

Zhongliang Pan and Ling Chen

Abstract The crosstalk effect among the signal lines is produced because of the increase of density of integrated circuits; it is primarily the capacitive coupling and inductance coupling of signal lines, where a signal transition on a signal line can take effects on the adjacent signal lines. The delay faults induced by crosstalk effects is generated when the aggressor line and victim line have simultaneous or near simultaneous transitions. In this paper, a new test generation method for the delay faults induced by crosstalk effects is presented; the method can produce the test vectors by using the circuit structure and Boolean satisfiability. First of all, the conjunctive normal form (CNF) formula of a circuit is constructed. Second, the crosstalk faults that consist of the aggressor and victim lines are activated, the path that can propagate the delay to the primary outputs are selected, and the Boolean satisfiability operations are used to get the value assignments of signal lines in the circuits. The experimental results for a lot of benchmark circuits show that the test vectors of the delay faults induced by crosstalk effects can be generated effectively by using the method proposed in this paper.

Keywords Digital circuits · Delay faults · Crosstalk effects · Test generation · Boolean satisfiability

25.1 Introduction

In recent years, the circuit density and switching speed are increasing, which has led to the signal integrity related problems in the VLSI circuits; for example, the crosstalk effect may occurs in the adjacent signal lines, which is the parasitic

Z. Pan (✉) · L. Chen
School of Physics and Telecommunications Engineering,
South China Normal University, Guangzhou 510006, China
e-mail: panz@scnu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_25

coupling among the adjacent lines [1, 2]. For the given two signal lines, if the signal transit on a line generates coupling effect on another line, then the line is called aggressor line, the another line is called victim line. The crosstalk effects can be classified into two following types: the crosstalk-induced pulse and the crosstalk-induced delay. For the crosstalk-induced pulse, the aggressor line is switching, and the victim line is in a static state. For the crosstalk-induced delay, both the aggressor and victim lines have near simultaneous signal transitions [3]. The crosstalk will decrease the reliability of circuits.

For the models and analysis of crosstalk faults, Sayil and Yuan [4] discussed the transient pulses caused by single event (SE) particles, investigated the signal speedup effects caused by SE crosstalks. Dhiman and Chandel [5] presented an analytical method that can provide closed form expressions for the dynamic crosstalk in the signal lines interconnects under sub-threshold circuit operation condition. Shang et al. [6] discussed the crosstalk avoidance codes to alleviate the crosstalk effects, proposed two types of memoryless crosstalk avoidance codes: forbidden overlap codes and forbidden transition codes. Qin et al. [7] investigated the statistical timing analysis considering crosstalk effects, used a piecewise linear delay change curve model to compute the statistical interconnect delay induced by crosstalk effects. Shirmohammadi and Miremadi [8] presented a crosstalk avoidance code based on numerical, which can omit the triplet opposite direction transitions produced by the crosstalk faults. Yiyuan et al. [9] discussed the crosstalk noise in the optical network on chip based on mesh structure, made use of the formal method to perform the systematical modeling and analysis for the crosstalk noise.

For the test generation of the delay faults induced by crosstalk effects, Peng et al. [10] discussed the test for the small-delay defects under the crosstalk and process variations, presented an evaluation and selection procedure for the test pattern generation. Jayanthy et al. [11] discussed the testing of crosstalk delay faults in the asynchronous circuits, the simulation-based test pattern generation and multi-objective genetic algorithms were used to produce the test vectors of the delay faults caused by crosstalks. Sinha et al. [12] discussed the multi-valued algebra for the crosstalk delay faults being induced by capacitance, made use of the timing operations on the multi-valued algebra as well as the backtrace procedure and search procedure to generate the test vectors of delay faults induced by crosstalk. Garbolino et al. [13] investigated the test of interconnect delays between cores in SoC, presented the approaches to generate the test vectors never inducing crosstalks and to decrease of power consumption during the test. Gope and Walker [14] discussed the maximization of crosstalk-induced slowdown during the path delay test, proposed a timing-driven test approach which can detect the delay faults such as the slowdown induced by crosstalk.

In this paper, a new test generation method for the delay faults induced by crosstalk effects is presented; the method makes use of the circuit structure and Boolean satisfiability.

25.2 Crosstalk Effects

The crosstalk effect is the one of the interference effects being caused by parasitic capacitance and inductance coupling. The crosstalk effect is injected into a victim line due to the capacitive and inductive coupling with neighboring lines. The injection of crosstalk in these signal lines may lead to different kinds of unexpected behaviors such as the signal delay and the reduction of achievable clock speed. The crosstalk delay induced by crosstalk effects is generated when the aggressor line and victim line have simultaneous or near simultaneous transitions. If the aggressor and victim lines transit in the same direction such as (from 0 to 1) or (from 1 to 0), their transition times are cut down, thus the signal delay is reduced, this phenomenon is called as crosstalk speedup. If the aggressor and victim lines transit in the opposite direction, for example, the aggressor line transits (from 0 to 1) and the victim line transits (from 1 to 0), thus the signal delay is increased, this phenomenon is called as crosstalk slowdown.

The uncertainty of the propagation delay of a signal may bring about the catastrophic violations of timing constraints in the circuits; it can cause incorrect operations of circuit. Therefore, it is necessary to carry out the test for the delay faults induced by crosstalk effects. One main task of test is to produce the test vectors of the faults. The example of test vectors is given for the circuit in Fig. 25.1.

In the circuit being shown in Fig. 25.1, we suppose there be a crosstalk effect between the signal lines b and c, where the line b is the aggressor line, the line c is victim line, and suppose the delay fault is generated by crosstalk slowdown, i.e., the aggressor line b and victim line c transit in the opposite direction, the line b transits from 1 to 0, and the line c transits from 0 to 1. There will be an increase in delay of line c, this may make the delay of line c to be out of range, thus results in the generation of path delay fault. Therefore, the crosstalk slowdown makes the following six paths that contain the line c to produce delay fault: x_2–a–c–e, x_3–a–c–e, x_4–c–e, x_2–a–c–f, x_3–a–c–f, x_4–c–f.

A pair of circuit input vectors V_1 and V_2 is needed for the detection of the delay fault induced by crosstalk effects. Let $V = (V_1, V_2)$. Apply the first vector V_1 at time t_0, after all the signals in the circuit have stabilized, apply the second vector V_2 at another time t_1. Sample the path output at time t_2, we can ascertain whether there is a delay fault in the path or not. The test vectors of the delay faults induced by

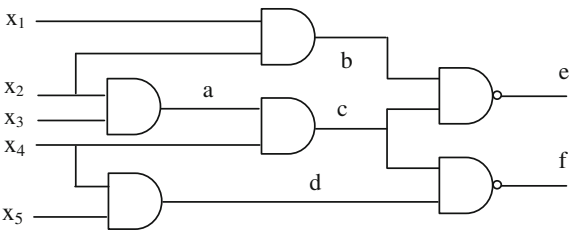


Fig. 25.1 The structure of a circuit

crosstalk effects can be obtained by utilizing the circuit structure. For example, the test vector is $V_1 = (x_1, x_2, x_3, x_4, x_5) = (1, 1, 0, 1, 1)$ and the $V_2 = (x_1, x_2, x_3, x_4) = (0, 1, 1, 1, 1)$ for the delay fault on the path x_3–a–c–f. The values of signal lines x_3, a, c, and f are 0, 0, 0, and 1, respectively, under the vector V_1 is applied at the primary inputs of circuit. The values of lines x_3, a, c, and f are 1, 1, 1, and 0, respectively, under the vector V_2 is applied at the primary inputs.

25.3 Test Generation by Boolean Satisfiability

For the detection of the delay faults induced by crosstalk effects, it is necessary to produce the test vectors. The test vectors are applied at the primary inputs of circuit under test, the output values of circuit are sampled at given time ranges. In the following, a test generation method for test vectors is given; the method is based on the circuit structure and Boolean satisfiability [15].

Let the primary inputs of the circuit under test be $x_1, x_2, \ldots, x_n$. Let the primary outputs of the circuit under test be $y_1, y_2, \ldots, y_m$. Let the set of the delay faults induced by crosstalk effects be $\mathbf{D} = \{d_1, d_2, \ldots, d_m\}$, where the m is the number of faults. The implementation procedure of the test generation method consists of following steps.

Algorithm 1

Step 1: Set up the initial value of the parameter n, i.e., k: = 0.
Step 2: Make the k: = k + 1, and select the delay fault d_k from the set **D**.
Step 3: Activate the crosstalk fault that consists of the aggressor line a_k and victim line v_k, i.e., set the values of the aggressor lines and the victim lines to the given transitions (0–1, or 1–0) respectively.
Step 4: For the sub-circuit, which consists of the signal lines that can produce effects on the values of aggressor line a_k, carry out the Boolean satisfiability and backtrack tracing operations to get a set of values of primary inputs, such that the value of aggressor line a_k is 1 (or 0), where the set is named as T_{10} (or T_{11}).
Step 5: For this sub-circuit, which consists of the signal lines that can produce effects on the values of victim line v_k, perform the Boolean satisfiability and backtrack tracing operations to get a set of values of primary inputs, such that the value of victim line v_k is 0 (or 1), where the set is named as T_{20} (or T_{21}).
Step 6: Select a path to propagate the delay on the victim line to at least one primary output. For the path selected, utilize the given values 1 (or 0) of aggressor line and 0 (or 1) of victim lines, and make use of the Boolean satisfiability and backtrack tracing operations to get a set of values of primary inputs, where the set is named as T_{30} (or T_{31}).

Step 7: Derive the intersection set α of the T_{10}, T_{20} and T_{30}, the $\alpha = T_{10} \cap T_{20} \cap T_{30}$. Derive the intersection set β of the T_{11}, T_{21} and T_{31}, the $\alpha = T_{11} \cap T_{21} \cap T_{31}$.

Step 8: If both the sets α and β are not empty, then the test vectors of delay fault exist. Let vector $V = (V_1, V_2)$, the $V_1 \in \alpha$ and $V_2 \in \beta$. The vector V is a test vector of the delay faults induced by crosstalk effects in aggressor line a_k and victim line v_k.

Step 9: If the $k < m$ is true, then go to the Step 2, else terminate the algorithm.

The Algorithm 1 is used to the test generation of the delay fault being produced by crosstalk slowdown, i.e., the aggressor line a_k and victim line v_k transit in the opposite direction, the line a_k transits from 1 to 0, and the line v_k transits from 0 to 1. The similar test generation can be performed for the crosstalk speedup.

Besides, in the Step 6 of the Algorithm 1, the path selected should propagate the delay to at least one primary output. If the path chosen can not satisfy the requirement, then we need to select another path. If there are not such paths, then the delay fault on the victim line is not testable, i.e., the test vectors of the delay faults do not exist, afterwards, directly go to the Step 9.

The Boolean satisfiability (SAT) approach is used in the Step 4, Step 5 and Step 6, it implementation is as follows. In the SAT approach, the conjunctive normal form (CNF) formula of the circuit under test is constructed, and the satisfying assignments for the CNF formula are found by explicitly enumerating every satisfying pattern for the CNF formula and checking the consistency. Let the inputs of the basic gates be $x_1, x_2, \ldots, x_n$, and the output be z. The CNF formulas of the basic gates AND, NAND, OR, and NOR are given respectively as follows.

$$\left[\prod_{i=1}^{n}(x_i+\bar{z})\right]\cdot\left(\sum_{i=1}^{n}\bar{x}_i+z\right),\quad \left[\prod_{i=1}^{n}(x_n+z)\right]\cdot\left(\sum_{i=1}^{n}\bar{x}_i+\bar{z}\right),$$
$$\left[\prod_{i=1}^{n}(\bar{x}_i+z)\right]\cdot\left(\sum_{i=1}^{n}x_i+\bar{z}\right),\quad \left[\prod_{i=1}^{n}(\bar{x}_i+\bar{z})\right]\cdot\left(\sum_{i=1}^{n}x_i+z\right),$$

The CNF formula of the XOR gate with the output z and two inputs x_1 and x_2 is

$$(\bar{x}_1+x_2+z)\cdot(x_1+\bar{x}_2+z)\cdot(x_1+x_2+\bar{z})\cdot(\bar{x}_1+\bar{x}_2+\bar{z})$$

The CNF formula of the NOT gate with an input x_1 and the output z is

$$(x_1+z)\cdot(\bar{x}_1+\bar{z})$$

The CNF formula of a circuit can be obtained by using the combination of the CNF formulas of every basic gate in the circuit. We take the circuit shown in Fig. 25.2a as an example.

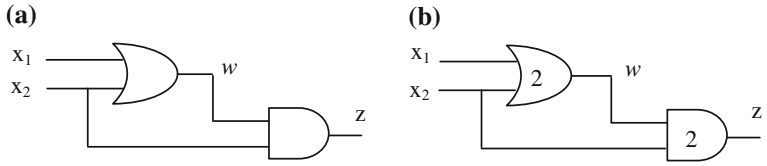


Fig. 25.2 An example of circuit **a** the lines x_1 and x_2 are the inputs of the circuit, the w is the output of the OR gate in the circuit, the z is the output of the circuit **b** the delays are 2 respectively for the lines x_1 to w, x_2 to w, x_2 to z, and w to z

The CNF formulas of the gates OR and AND in Fig. 25.2a are

$$(\bar{x}_1 + w) \cdot (\bar{x}_2 + w) \cdot (x_1 + x_2 + \bar{w}) \text{ and } (w + \bar{z}) \cdot (x_2 + \bar{z}) \cdot (\bar{w} + \bar{x}_2 + z)$$

Thus, the CNF formula of the circuit in Fig. 25.2a is given by

$$(\bar{x}_1 + w) \cdot (\bar{x}_2 + w) \cdot (x_1 + x_2 + \bar{w}) \cdot (w + \bar{z}) \cdot (x_2 + \bar{z}) \cdot (\bar{w} + \bar{x}_2 + z)$$

If the timed Boolean variables are used, the CNF formula will be the timed formula, where for the test generation of the delay faults induced by crosstalk effects, it is necessary the timed CNF formula. For example, for the circuit in Fig. 25.2b, suppose the delays between the lines x_1 to w and the x_2 to w be 2. The delays between the lines w to z and the x_2 to z also be 2. Where the $z(t = 4)$ is evaluated, the $z(t = 4)$ is abbreviated as $z(4)$. The timed CNF formulas of the gates OR and AND in Fig. 25.2b are

$$(\bar{x}_1(0) + w(2)) \cdot (\bar{x}_2(0) + w(2)) \cdot (x_1(0) + x_2(0) + \bar{w}(2)) \text{ and}$$
$$(w(2) + \bar{z}(4)) \cdot (x_2(0) + \bar{z}(4)) \cdot (w(2) + \bar{x}_2(0) + z(4))$$

Therefore, the timed CNF formula of the circuit in Fig. 25.2b is given by

$$(\bar{x}_1(0) + w(2)) \cdot (\bar{x}_2(0) + w(2)) \cdot (x_1(0) + x_2(0) + \bar{w}(2))$$
$$\cdot (w(2) + \bar{z}(4)) \cdot (x_2(0) + \bar{z}(4)) \cdot (\bar{w}(2) + \bar{x}_2(0) + z(4))$$

An assignment q satisfies a formula P if $P(q) = 1$, where the assignment is a satisfying truth assignment. A formula P is unsatisfiable if there is no satisfying truth assignment for the formula. For a formula P, the task of the Boolean satisfiability approach is to find the assignments to satisfy the formula P, or to indicate that there are not such assignments.

In this paper, in order to determine the values of the signal lines in the circuits, the implementation steps of satisfiability approach consists of following steps. First of all, we construct the timed CNF formula of the circuit under test. Second, we count the number of each variable (or its complement) that appearing in the CNF formula and name the number μ. Third, the assignment of the variable with maximal number μ is carried out. Afterwards, if a variable has its bigger number μ, then

the assignment of the variable have precedence over the other variable with small number μ. This procedure is executed repeatedly until the assignments of all variables have finished.

25.4 Experimental Results

The test generation method in this paper for the delay faults induced by crosstalk effects has been implemented in C++ language, and run them on a personal computer with 3.06 GHz and 1G main memory. We have carried out a lot of experiments for ISCAS85 benchmark circuits. The Table 25.1 shows the features of the benchmark circuits.

In Table 25.1, the column "Circuits" shows the name of a benchmark circuit. The column "PI" gives the number of primary inputs in a benchmark circuit. The column "PO" denotes the number of primary outputs in a circuit. The column "Gates" shows the total number of the gates in a circuit. The column "Lines" gives the total number of signal lines in a circuit.

For these benchmark circuits, we have performed the experiments for the test generation of delay faults induced by crosstalk effects. In these experiments, we randomly choose 80 crosstalk delay faults from these circuits. The test vectors of the crosstalk delay faults are generated by using the Algorithm 1.

The total time needed to obtain the test vectors for all 80 crosstalk faults in a benchmark circuit is given in Table 25.1. The column "Test time" shows the total time needed, the time unit is minute. The test time needed for the circuits C880, C1355 and C1908 is less than 0.1 min, for the circuits C2670 and C3540 is less than 0.3 min, for the circuit C5315 is less than 0.5 min, and for the circuits C6288 and C7552 is less than 0.6 min.

These experimental results demonstrate that for the testable delay faults, the test vectors of delay faults can be generated rapidly by using the method in this paper. For the not testable delay faults, i.e., there are not the test vectors for the faults, the method in this paper can also tackle this case by finding the conflicts of value assignments of the variables in the satisfiability operations, and stop the execution

Table 25.1 The ISCAS85 benchmark circuits

Circuits	PI	PO	Gates	Lines	Delay faults	Test time
C880	60	26	383	880	80	0.1
C1355	41	32	546	1355	80	0.1
C1908	33	25	880	1908	80	0.1
C2670	233	140	1193	2670	80	0.3
C3540	50	22	1669	3540	80	0.3
C5315	178	123	2307	5315	80	0.5
C6288	32	32	2416	6288	80	0.6
C7552	207	108	3512	7552	80	0.6

of the method. The experimental results also show that the most of the time shown in the column "Test time" in Table 25.1 is spent on the processing for the not testable delay faults.

25.5 Conclusions

The delay fault is one of main faults induced by crosstalk effects. The test generation of the delay faults is related to the processing for the aggressor lines and victim lines, the total circuit structure, and the value assignments of signal lines in the circuits. The test generation method in this paper utilizes the Boolean satisfiability to get the test vectors of the delay faults. There are some factors that can take impact on the performance of the method, for example, the size of the CNF formula of the circuit under test, the activations of the aggressor lines and victim lines, and the value assignments for the primary input lines. Therefore, the reduction technique for the CNF formula of a circuit should be investigated intensively in the future.

Acknowledgments This work was supported by Guangdong Provincial Natural Science Foundation of China (No. 2014A030313441), Guangdong Province Science and Technology Project (No. 2013B090600063, No. 2014B090901005), Guangzhou Science and Technology Project (No. 201510010169), and National Natural Science Foundation of China (No. 61072028).

References

1. Wahab M, Alam M (2014) Implications of electrical crosstalk for high density aligned array of single-wall carbon nanotubes. IEEE Trans Electron Devices 61(12):4273–4281
2. Nikdast M, Jiang X, Duong LH (2015) Fat-tree-based optical interconnection networks under crosstalk noise constraint. IEEE Trans VLSI 23(1):156–169
3. Nazarian S, Das D (2013) An efficient current-based logic cell model for crosstalk delay analysis. Int J Electron 100(4):439–467
4. Sayil S, Yuan L (2015) Modeling single event crosstalk speedup in nanometer technologies. Microelectron J 46(5):343–350
5. Dhiman R, Chandel R (2015) Dynamic crosstalk analysis in coupled interconnects for ultra-low power applications. Circ Syst Sig Process 34(1):21–40
6. Shang C, Cheng J, Ke HT, Shin LD (2014) Constructions of memoryless crosstalk avoidance codes via C-transform. IEEE Trans VLSI 22(9):2030–2033
7. Qin T, Zjajo A, Berkelaarand M, Meijs ND (2014) Considering crosstalk effects in statistical timing analysis. IEEE Trans CAD 33(2):318–322
8. Shirmohammadi Z, Miremadi SG (2013) Crosstalk avoidance coding for reliable data transmission of network on chips. International symposium on system on chip, pp 1–4
9. Yiyuan X, Nikdast M, Jiang X (2013) Formal worst-case analysis of crosstalk noise in mesh-based optical networks-on-chip. IEEE Trans VLSI 21(10):1823–1836
10. Peng K, Yilmaz M, Chakrabarty K (2013) Crosstalk and process variations-aware high quality tests for small-delay defects. IEEE Trans VLSI 21(6):1129–1142

11. Jayanthy S, Bhuvaneswari MC, Prabhu M (2013) Simulation-based ATPG for low power testing of crosstalk delay faults in asynchronous circuits. Int J Comput Appl Technol 48 (3):241–245
12. Sinha A, Gupta SK, Breuer MA (2008) A multi-valued algebra for capacitance induced crosstalk delay faults. IEEE 17th Asian test symposium, pp 89–96
13. Garbolino T, Gucwa K, Kopec M (2007) Crosstalk-insensitive method for testing of delay faults in interconnects between cores in SoCs. In: International conference on mixed design of integrated circuits and systems, pp 496–500
14. Gope D, Walker DM (2012) Maximizing crosstalk-induced slowdown during path delay test. In: IEEE 30th international conference on computer design, pp 159–166
15. Huan C, Silva JM (2013) A two-variable model for SAT-based ATPG. IEEE Trans CAD 32 (12):1943–1956

Chapter 26
A Fitting Method for Wheel Profile Line Based on Lagrange Multiplier

Lu Wang, Xiaoqing Cheng, Zongyi Xing, Yong Zhang and Yong Qin

Abstract Lagrange multiplier method is proposed to extract wheel profile line accurately. First, the data measured by 2D laser sensors is divided into several sections. Second, the least squares fitting equations are established for the data of each section. Third, the least squares regression model is built by the fitting equation. Then, taking the third-order continuous derivative at breakpoints as constraint conditions, Lagrange multiplier method is applied to derive the parameters of least squares solution. Finally, the field experiments are conducted. The result shows that the profile line is smooth and continuous. Compared with the actual data points, the maximum offset amplitude is within 0.1 % which meets the accuracy requirements of measuring wheel size.

Keywords Wheel profile line · Piecewise profile line fitting · Lagrange multiplier method

26.1 Introduction

With the rapid development of urban rail transportation in China, the security issues of train operation has become increasingly prominent. The wheel is subject to wear due to the interaction with the rail. When the wear is beyond the limitation, the derailment and other traffic accidents may happen [1]. So it is very important to extract the wheel profile line accurately.

The wheel profile line shown by discrete points could be measured by laser CCD [2] or laser rangefinder [3]. Generally, the profile line is fitted by the discrete points.

L. Wang · Z. Xing (✉) · Y. Zhang
School of Automation, Nanjing University of Science and Technology,
No. 200 Xiaolingwei, Xuanwu, Nanjing, Jiangsu, China
e-mail: xingzongyi@163.com

X. Cheng · Y. Qin
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiao Tong University,
Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_26

The piecewise curve fitting method achieves better results compared with using only one kind of curve due to the complexity of wheel profile [4]. A lot of researches about piecewise curve fitting have been made recently. Yan [5] used the least squares method based on interpolation model to get the piecewise fitting curve with small volatility. Mashinchi [6] proposed the least squares method based on global multi-objective programming model to get the polynomial fitting curve infinitely approaching measured value. Zhang [7] used a fitting method of piecewise cubic curve based on least squares to get the polynomial fitting curve.

The traditional global fitting method fails to take the additional oscillation caused by high-order polynomial into consideration. Besides, the curve is not smooth at breakpoints. The accuracy of curve fitting is reduced by only using the interpolation method. So, a piecewise curve fitting method based on Lagrange multiplier method is proposed. First, the data measured from sensors are preprocessed, which includes covering and fusing the data. Second, the data are divided into several sections. Third, the polynomial equations are established to fit the piecewise profile lines. Then taking the third-order continuous derivative at breakpoints as constraint conditions, the least squares model, which has the overall minimum fitting error and the third derivative at breakpoints, is built. Finally, Lagrange multiplier method is applied to derive the parameters.

26.2 The Principle and Device of Wheel Profile Detection

The wheelset consists of two identical wheels and an axle. The tread is the contacting zone of the rail and wheel, and the wheel section is shown in Fig. 26.1. The inside of the wheel tread along the circumference of the wheel is called rim. The base point is the point on the tread, 70 mm distant from the base line around the inside rim. The rim height is the horizontal height between the base point and the top of rim. The rim thickness is the rim horizontal thickness which is 10 mm higher than base point [8]. During wheels maintenance, the changes of rim thickness and tread wear are the key parameters, so it is important that the profile line should be accurately extracted.

The wheel profile detection system contains two 2D laser sensors which are symmetrically installed beside the track. The relative distances between each sensor

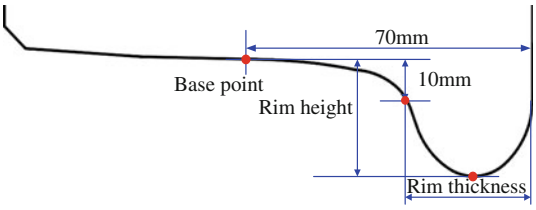


Fig. 26.1 The wheel sectional schematic

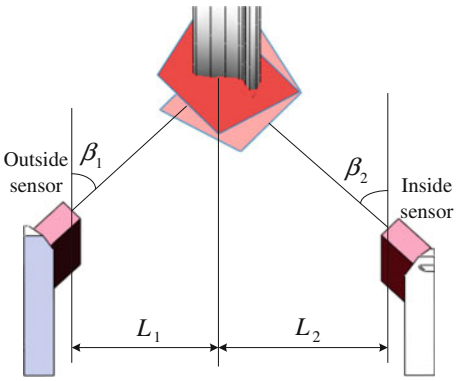


Fig. 26.2 The sensors installation diagram

and track are L1, L2 which are within the range from 100 to 450 mm. The laser surface of the two sensors should be in the same plane. The angles between each sensor and track are β_1, β_2, which are both 45° as shown in Fig. 26.2.

26.3 The Principle of Profile Fitting Algorithm

The Fig. 26.3 shows the flow diagram of the profile fitting algorithm. There are two parts: the first one is to preprocess data. The second one is to gain the global smooth and continuous profile line.

26.3.1 Data Preprocessing

The coordinates of output points detected by the 2D laser displacement sensors are transformed and rotated into same coordinate. The new coordinates can be calculated according to the formulas:

$$u_n^{(1)} = x_n^{(1)} \cos\beta_1 + y_n^{(1)} \sin\beta_1 \quad v_n^{(1)} = y_n^{(1)} \cos\beta_1 - x_n^{(1)} \sin\beta_1$$
$$u_n^{(2)} = x_n^{(2)} \cos\beta_2 - y_n^{(2)} \sin\beta_2 \quad v_n^{(2)} = y_n^{(2)} \cos\beta_2 + x_n^{(2)} \sin\beta_2$$

where $(x_n^{(1)}, y_n^{(1)})$ and $(x_n^{(2)}, y_n^{(2)})$ are the coordinate value of any output point in the initial coordinate system, respectively. $(u_n^{(1)}, v_n^{(1)})$ and $(u_n^{(2)}, v_n^{(2)})$ are the coordinate

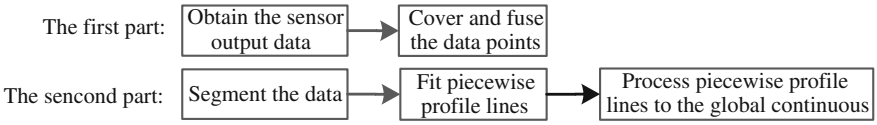


Fig. 26.3 The flowchart of algorithm

value of any output point in the new coordinate system, respectively. β_1 and β_2 are the angles between the laser and longitudinal vertical line, respectively.

The two sets of data after transformed are fused into a same system based on the following formula:

$$u_n^{(0)} = u_n^{(1)} + a, v_n^{(0)} = v_n^{(1)} + b \quad u_n^{(0)} = u_n^{(2)}, v_n^{(0)} = v_n^{(2)}$$

The coordinate value of the origin point of the outside sensor is (a, b) in the transformed coordinate system of inside sensor.

26.3.2 Fitting Profile Line

The valid data of tread is stored in a form of discrete point after data preprocessing. In order to accurately calculate the parameters of wheel, it is necessary to fitting profile line to reconstruct the wheel profile line.

The each segment of standard profile line is composed of linear and circular, so the standard profile line could be divided into several sections, and each section could be fit with the specified function [9].

The coordinates of data points after preprocessing are (x_i, y_i), i = 1, 2, …, n. The data points would be divided into K groups, and the coordinates are the following:

$$\begin{aligned}
&(x_{11}, y_{11}), (x_{12}, y_{12}), \ldots, (x_{1N_1}, y_{1N_1}) \\
&(x_{21}, y_{21}), (x_{22}, y_{22}), \ldots, (x_{2N_2}, y_{2N_2}) \\
&\cdots \\
&(x_{K1}, y_{K1}), (x_{K2}, y_{K2}), \ldots, (x_{KN_K}, y_{KN_K})
\end{aligned}$$

where N_1, N_2, …, N_k,…N_K are the numbers of data points in each interval, and they meet the following equation:

$$N_1 + N_2 + \cdots + N_k + \cdots + N_K = n$$

The fitting equation $F(x)$ is established according to data points in each interval. As the continuous function can be approached to polynomial arbitrarily, the fitting equation is generally polynomial. The $F(x)$ can be shown by the following:

$$F(x) = \begin{cases} F_1(x) = \sum_{j=1}^{m_1} \alpha_{1j} f_{1j}(x), x \le x_{o1} \\ F_2(x) = \sum_{j=1}^{m_2} \alpha_{2j} f_{2j}(x), x_{o1} \le x \le x_{o2} \\ \vdots \\ F_K(x) = \sum_{j=1}^{m_K} \alpha_{Kj} f_{Kj}(x), x_{o(K-1)} \le x \end{cases} \tag{26.1}$$

where α is the undetermined coefficient, $f(x)$ is the linearly independent basis functions $\{1, x, x^2, \ldots, x^n\}$, and m_k is the number of the primary functions in corresponding interval.

The least squares model, a widely used method, is built to fit the piecewise profile line. The least squares model, which has the overall minimum fitting error and the third derivative at the segment point, is shown by the following formula:

$$S = \min \sum_{k=1}^{K} \sum_{j=1}^{N_k} [F_k(x_{kj}) - y_{kj}]^2 \tag{26.2}$$

where, S is the least squares estimator.

To ensure the continuity from $F_{k-1}(x)$ to $F_k(x)$ at the segment point x_{ok}, the function values of the $F_{k-1}(x)$ and $F_k(x)$ at the segment point x_{ok} should be equal, so the constraints to end points are described by Eq. (26.3):

$$F_{k-1}(x_{ok}) = F_k(x_{ok}) \tag{26.3}$$

At the same time, to ensure the smoothness from $F_{k-1}(x)$ to $F_k(x)$ at the segment point x_{ok} and the curve with the better geometric analysis performance, the third derivative of the two curves forward and backward should be successive at the segment point x_{ok}, so the constraints to end points are described by Eq. (26.4):

$$F'_{k-1}(x_{ok}) = F'_k(x_{ok}), F''_{k-1}(x_{ok}) = F''_k(x_{ok}), F'''_{k-1}(x_{ok}) = F'''_k(x_{ok}) \tag{26.4}$$

The Lagrange multiplier method is applied to transform the conditional extremum problem into the unconditional [10].

The following equations are constructed according to the four constraint conditions in the previous section.

$$H_{k1} = F_{k-1}(x_{ok}) - F_k(x_{ok}) = 0 \quad H_{k2} = F'_{k-1}(x_{ok}) - F'_k(x_{ok}) = 0$$
$$H_{k3} = F''_{k-1}(x_{ok}) - F''_k(x_{ok}) = 0 \quad H_{k4} = F'''_{k-1}(x_{ok}) - F'''_k(x_{ok}) = 0 \tag{26.5}$$

The Lagrange multipliers λ are introduced, and the Lagrange function is constructed according to Eq. (26.5) as follows:

$$\begin{aligned} L_k(\alpha, \lambda) &= L_k(\alpha_{k1}, \alpha_{k2}, \ldots, \alpha_{km_k}, \lambda_{k1}, \lambda_{k2}, \lambda_{k3}, \lambda_{k4}) \\ &= F_k + \lambda_{k1} H_{k1} + \lambda_{k2} H_{k2} + \lambda_{k3} H_{k3} + \lambda_{k4} H_{k4} \end{aligned} \tag{26.6}$$

According to the necessary condition for obtaining extremum of multivariate function, the partial derivations of function $L_k(\alpha, \lambda)$ with respect to variables α and λ are the following:

$$\begin{cases} \frac{\partial L_k(\alpha,\lambda)}{\partial \alpha_{kj}} = 0, j = 1, 2, \ldots, m_k \\ \frac{\partial L_k(\alpha,\lambda)}{\partial \lambda_{ki}} = 0, i = 1, 2, 3, 4 \end{cases} \tag{26.7}$$

The variables α are obtained by solving the above equations.

26.4 Case Study

The actual detection system is installed in a certain subway depot. And the train goes through the detection system with the speed of 3 km/h.

The two 2D laser sensors are installed on both sides of the track. And the angles between each sensor and vertical are β_1 and β_2 that are both 45°. When the train goes through the system, the data points are recorded by sensors. Then, the coordinates of these data points are transformed according to the formula:

$$u_n^{(1)} = x_n^{(1)} \cos 45° + y_n^{(1)} \sin 45° \quad u_n^{(2)} = x_n^{(2)} \cos 45° - y_n^{(2)} \sin 45°$$
$$v_n^{(1)} = y_n^{(1)} \cos 45° - x_n^{(1)} \sin 45° \quad v_n^{(2)} = y_n^{(2)} \cos 45° + x_n^{(2)} \sin 45°$$

Obtain the horizontal ordinate $X = -268.6155$ mm of the right end face of the tread by data processing as shown in Fig. 26.4. Section the valid data of the tread based on these 6 number sets: $(X - 130, X - 100)$, $[X - 100, X - 60)$, $[X - 60, X - 46)$, $[X - 46, 25)$, $[X - 25, X - 6)$, $[X - 6, X)$. Therefore, the horizontal ordinate of segment points x_{ok} are 100, 60, 46, 25, 6.

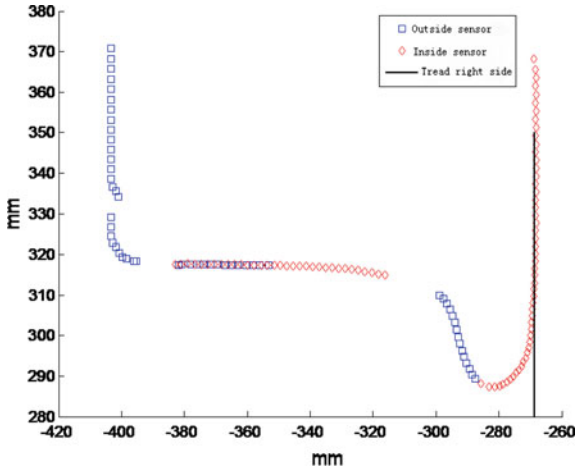


Fig. 26.4 The effective data points of the wheel profile line

According to the principle of least squares, the least squares regression model which has the global minimum fitting error and is successive at segment points can be set up by the following formulas:

$$S = \min \sum_{k=1}^{6} \sum_{j=1}^{N_k} [F_k(x_{kj}) - y_{kj}]^2 \tag{26.9}$$

$$\begin{aligned} \text{s.t.} \quad & F_{k-1}(x_{ok}) = F_k(x_{ok}) \quad F'_{k-1}(x_{ok}) = F'_k(x_{ok}) \\ & F''_{k-1}(x_{ok}) = F''_k(x_{ok}) \quad F'''_{k-1}(x_{ok}) = F'''_k(x_{ok}) \end{aligned} \tag{26.10}$$

where, the value of k is from 2 to 6 in Eq. (26.10). The Lagrange function is established based on Lagrange multiplier method. The partial derivations of function $L_k(\alpha, \lambda)$ with respect to variables α and λ are established according to the necessary condition for obtaining extremum of multivariate function, and the results of the partial derivations are defined zero to establish the equations. Then, the equations are solved in MATLAB platform. The fitted wheel profile line is shown in Fig. 26.5.

From Fig. 26.5, the smoothness at the interaction between the wheel and the rail is unsatisfied, which may be caused by two reasons. One is the actual wear shown in Fig. 26.5; the other is the lack of the detection data at the interaction between the wheel and the rail. The errors between the data from detection and detected data and the data from fitting profile line are shown in Fig. 26.6.

From Fig. 26.6, the errors in tread area are relatively small within ±0.1 mm; the errors in rim area are relatively large and within ±0.4 mm. After analyzing the data at the rim, the results are shown in Table 26.1.

Compared the original points (x, y) with the corresponding points (x, y') on the fitting profile line, the maximum error R is not beyond ±0.39 mm, and the offset amplitude A ($A = R/(x^2 + y^2)^{\wedge 1/2\%}$) [11] is less than 0.096 %. The wheel rim

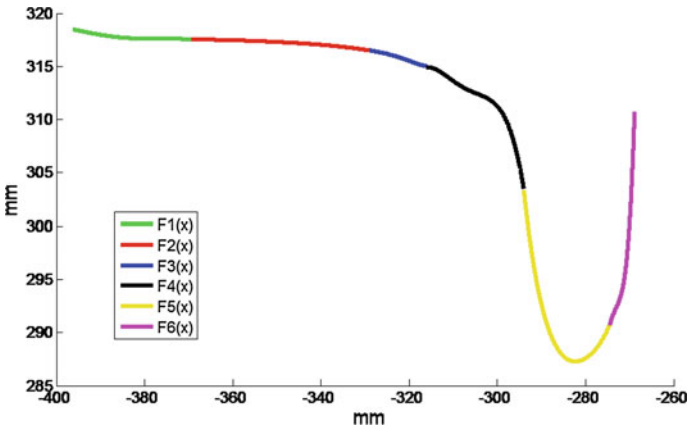


Fig. 26.5 The fitting wheel profile line

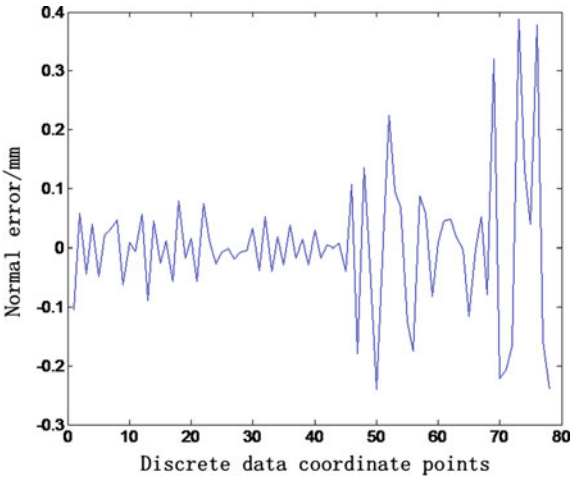


Fig. 26.6 The errors between the data from actual detection and the data from fitting profile line

Table 26.1 The analysis results

System output value (x, y)	Fitting value (x, y')	Error R	System output value (x, y)	Fitting value (x, y')	Error R
−288.47, 290.35	−288.47, 290.43	−0.09	−273.70, 291.56	−273.70, 291.48	0.08
−287.43, 289.28	−287.43, 289.34	−0.06	−272.84, 292.44	−272.84, 292.76	−0.32
−285.64, 288.17	−285.64, 288.09	0.08	−272.21, 293.52	−272.21, 293.30	0.22
−283.09, 287.32	−283.09, 287.33	−0.01	−271.49, 294.51	−271.49, 294.31	0.20
−281.35, 287.29	−281.35, 287.34	−0.05	−271.00, 295.76	−271.00, 295.60	0.16
−279.89, 287.55	−279.89, 287.60	−0.05	−270.54, 297.03	−270.54, 297.42	−0.39
−278.64, 287.99	−278.64, 288.01	−0.02	−270.30, 298.54	−270.30, 298.68	−0.14
−277.49, 288.54	−277.49, 288.54	0.00	−270.07, 300.06	−270.07, 300.10	−0.04
−276.49, 289.26	−276.49, 289.14	0.12	−269.80, 301.58	−269.80, 301.96	−0.38
−275.45, 289.92	−275.45, 289.92	0.00	−269.66, 303.23	−269.66, 303.07	0.16
−274.53, 290.71	−274.53, 290.76	−0.05	−269.48, 304.83	−269.48, 304.59	0.24

thickness and height are generally about 30 mm, so the error of offset amplitude for wheel set is:

$$30 \text{ mm} \times 0.096\ \% = 0.0288 \text{ mm}$$

This value of error shows that the error barely influences the detection accuracy, and this fitting method can satisfy the practical measurement requirements.

26.5 Conclusions

Lagrange multiplier method is proposed to fit wheel profile line. The field test shows that wheel profile line can be exactly fitted by this method, and the maximum offset amplitude between the fitting profile line and the actual detected points is less than 0.096 %. This result shows that the overall continuity and higher order smoothness of the profile line can be obtained, which can lay the foundation for the following geometric wear analysis.

Acknowledgments This study was supported by the Fundamental Research Funds for the Central Universities (AE89454). The author gratefully acknowledge the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially.

References

1. Marshall MB, Rovira A, Roda A et al (2011) Experimental and numerical modelling of wheel–rail contact and wear. J Wear 271(5):911–924
2. Jing C, Jianwei Y, Zhizhong T (2014) On dynamic testing system of metro vehicle wheelset. J Urban Mass Transit 17(7):82–84 (in Chinese)
3. Xin G (2014) The design of wheel profile measuring device based on laser distance sensor. Southwest Jiaotong University, Chengdu (in Chinese)
4. de Pater AD, Fujie X (1998) Piecewise curve-fitting method on measured profile. J China Railway Soc 20(1):119–122 (in Chinese)
5. Shiying Y, Xiangchang S, Yuanzhong L et al (2012) Application of curve fitting and interpolation model in deformation forecast of a mining area. J Nonferrous Met (Mine Section) 64(1):66–68 (in Chinese)
6. Mashinchi M, Pourkarimi L, Yaghoobi MA (2011) Efficient curve fitting: an application of multiobjective programming. J Appl Math Model 35(1):346–365
7. Hao Z, Huihong C, Shan C et al (2007) Research of piecewise cubic curve-fitting method based on least-square principle. J Sci Technol Eng 7(3):352–355 (in Chinese)
8. CN-TB (2003) Wheel profile for locomotive and car S 2003 (in Chinese)
9. Weiping T (2006) Math methods applied for measuring train's wheel figure with laser sensor. J Comput Measur Control 14(9):1233–1235 (in Chinese)
10. Eryan Z, Xiaofeng Z (2013) Study on the Lagrange multiplier method for solving constrained linear regression. J Beijing Inst Graph Commun 21(2):76–78 (in Chinese)
11. Ke W, Xingwei S, Zhonghui T (2013) Fairing processing of curve point cloud data based on Lagrange multiplier method. J Shenyang Univ Technol 35(3):309–313 (in Chinese)

Chapter 27
The Voice Conversion Method Based on Sparse Convolutive Non-negative Matrix Factorization

Qianmin Zhang, Liang Tao, Jian Zhou and Huabin Wang

Abstract We propose a voice conversion method based on sparse convolutive non-negative matrix factorization. The method utilizes the Itakura–Saito distance as the objective cost function, making the smaller matrix element with a smaller reconstruction error due to the property of scale invariant of the cost function. The time–frequency basis of the source and target were extracted during the training phase, and the speech is converted through time–frequency basis substitution. The effect of whisper-to-normal speech conversion experiment is also conducted. Experimental results show that the proposed voice conversion method outperforms the method based on the conventional convolutive non-negative matrix factorization and the method based on the Kullback–Leibler (K-L) cost function in the aspects of speech intelligibility.

Keywords Speech intelligibility · Voice conversion · Itakura–Saito distance · Time–frequency basis

27.1 Introduction

The aim of voice conversion is to convert one speaker's voice to another speaker's voice with the same pronunciation characteristics, while keeping the speech content unchanged. As a new branch of speech signal processing, voice conversion has broad application prospects, such as speech synthesis, speech recognition, multimedia entertainment, etc. The research of voice conversion has more and more important theoretical value.

In the 1980s, Abe is the earliest to study on the voice conversion among different speakers, and in the early 1990s [1]. The existing voice conversion methods mainly include Gaussian mixture model [2, 3], hidden Markov model [4, 5], artificial

Q. Zhang (✉) · L. Tao · J. Zhou · H. Wang
School of Computer Science and Technology, Anhui University,
No. 111 Jiulong Road, Shushan District, Hefei City, Anhui Province, China
e-mail: qmzhang002@sina.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_27

neural network [6], and so forth. However, there are still some disadvantages in these methods. For example, some methods assume that the inter-frame of each speech was independent, ignoring the temporal sequence correlation, thus leading to the discontinuity of the converted speech [7].

In order to take the correlation between adjacent speech frames into consideration in the process of conversion, Toda et al. and Zen, respectively, proposed the voice conversion methods based on the frequency spectrum parameters track of GMM [8] and HMM models [9]. Sun et al. proposed a method using convolutive non-negative matrix factorization [10] to deal with the voice conversion problem. Although it gets much better conversion performance, it fails to fully demonstrate the certain characteristics of speaker's speech signal, which affect the conversion effects.

To tackle these problems, Ma et al. proposed the voice conversion method based on sparse convolutive non-negative matrix factorization (SCNMF) [11]. The decomposed overcomplete time–frequency basis can better carry both the speaker's characteristic information and the speech inter-frame correlation, and greatly improving the effectiveness of voice conversion. However, this method cannot make the converted speech signal have good intelligibility.

In order to solve this problem, this paper proposes a voice conversion method based on sparse convolutive non-negative matrix decomposition. The method utilizes Itakura–Saito distance [12] as the objective cost function to measure the errors between an original matrix and its reconstruction version, making the smaller matrix element with a smaller reconstruction error, due to the property of scale invariant of the Itakura–Saito distance function.

27.2 Sparse Convolutive Non-negative Matrix Factorization

27.2.1 Non-negative Matrix Factorization

Generally speaking, basis decomposition is a common method to conduct signal analysis, and NMF is a matrix decomposition method in the constraint that all the elements in the factorized matrix are non-negative. Specifically, for a given non-negative matrix V, the NMF algorithm decomposes it to a non-negative matrix W and another non-negative matrix H, such that

$$V = W * H \tag{27.1}$$

In (27.1), V is $m * n$ matrix, W consists of r column vectors. W is called the basis matrix, H is called the coefficient matrix. (27.1) can also be written as follows:

$$v_{ij} \approx (\boldsymbol{W}*\boldsymbol{H})_{ij} = \sum_{r=1}^{R} w_{ir} h_{rj} \tag{27.2}$$

v_{ij} is the element of i-th row and j-th column in the matrix V. Generally speaking, the value of r is much smaller than n, so the relationship between r, m and n is as follows:

$$(m+n)*r \leq m*n, \tag{27.3}$$

27.2.2 Sparse Convolutive Non-negative Matrix Factorization

Smaragdis proposed a convolutive NMF for speech signal processing, using time–frequency basis to replace basis vectors, which effectively carries the partial correlation between successive frames of speech signal [13]. The convolutive non-negative matrix factorization of V can be formalized as follows:

$$\boldsymbol{V} \approx \sum_{t=0}^{T_0-1} \boldsymbol{W}_t \overrightarrow{\boldsymbol{H}}^{t}, \tag{27.4}$$

where T_0 is the length of each spectrum sequence, W_t is the t-th time–frequency basis matrix, and the matrix consisting of all the j-th columns of W_t along the t axis time spectrum is called the j-th object of the speech spectrum. $\overrightarrow{(*)}^{t}$ is the right operator of column shifting. Specifically, assuming that the value of H is

$$\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_i, \ldots, \boldsymbol{h}_N], \tag{27.5}$$

$\mathbf{h}_i$ is the i-th column vector of the coefficient matrix H, i = 1, 2,…, N,

$$\overrightarrow{\boldsymbol{H}}^{t} = [\boldsymbol{0}, \ldots, \boldsymbol{0}, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_i, \ldots, \boldsymbol{h}_{N-t}], \tag{27.6}$$

The objective cost function directly influences the efficiency of the algorithm. Lee et al. proposed two kinds of cost functions [14, 15], that is, the error minimization method based on Euclidean distance and that based on the K-L divergence. However, both the Euclidean distance and the K-L divergence have the disadvantages of scale variability, this paper utilizes Itakura–Saito distance as the error cost function:

$$\mathrm{D}(\boldsymbol{V} \parallel \boldsymbol{\Lambda}) = \sum_{ik} \left(\frac{v_{ik}}{[\boldsymbol{\Lambda}]_{ik}} - \log \frac{v_{ik}}{[\boldsymbol{\Lambda}]_{ik}} - 1 \right), \tag{27.7}$$

where matrix is the estimate of matrix V, $\boldsymbol{\Lambda} = \sum_{t=0}^{T_0-1} \boldsymbol{W}_t \overrightarrow{\boldsymbol{H}}^{t}$, $[\boldsymbol{\Lambda}]_{ik} = \sum_{t=0}^{T_0-1} \sum_{j=1}^{R} w_{ijt} (\overrightarrow{h}_{jk}^{t})$. As can be seen, (27.7) has the property of scale invariant.

(27.4) can be regarded as the sum of T_0-th non-negative matrix factorization operations. Because sparse convolutive non-negative matrix factorization algorithm can show the physical properties of the speech much more intuitively, and the algorithm has the superiority of fast convergence and small storage space [16], this paper combines the objective function of convolutive non-negative matrix with sparse constraint, the new objective cost function is obtained as follows:

$$\mathrm{D}(\boldsymbol{V} \parallel \boldsymbol{\Lambda}) = \sum_{ik} \left(\frac{v_{ik}}{[\boldsymbol{\Lambda}]_{ik}} - \log \frac{v_{ik}}{[\boldsymbol{\Lambda}]_{ik}} - 1 \right) + \lambda \sum_{jk} h_{jk}, \tag{27.8}$$

27.3 The Proposed Speech Conversion Method

Voice conversion is to convert one speaker's voice to another speaker and keep the speech content unchanged. In order to complete the voice conversion of a speaker, two phases are generally included: training phase and conversion phase. Figure 27.1 diagrams the voice conversion process [11].

As can be seen in Fig. 27.1, the voice conversion includes two stages: training phase and conversion phase. In the training stage, there already exist literatures

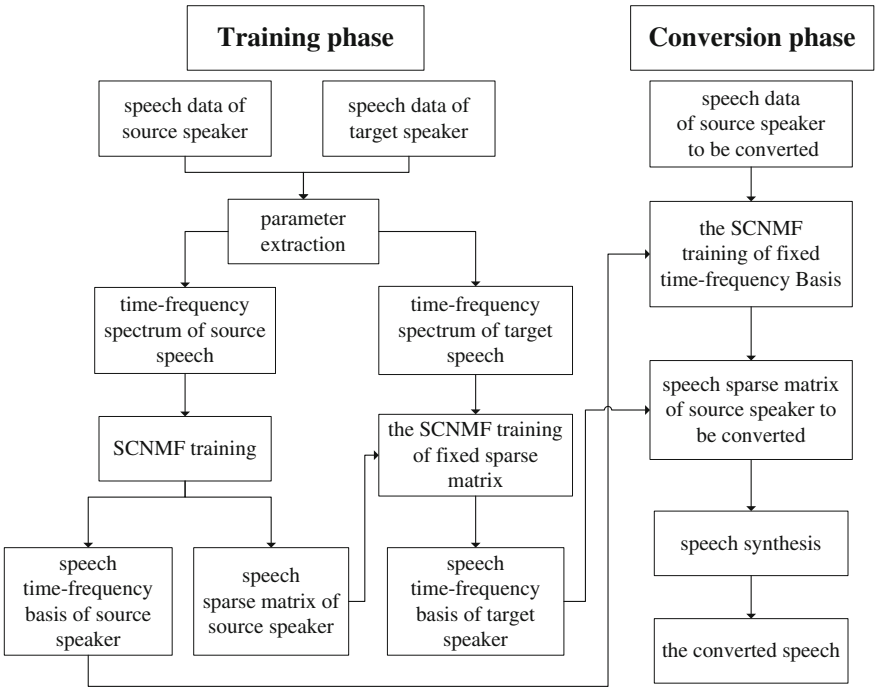


Fig. 27.1 Voice conversion process

showing that the sparse convolutive non-negative matrix factorization algorithm based on Itakura–Saito distance gains larger reconstruction result for speech ingredients than the sparse convolutive non-negative matrix factorization algorithm based on the K-L divergence, and the reconstructed speech signal has higher intelligibility. In this paper, we go a step further to evaluate its performance on the voice conversion.

27.4 Experiment Results and Disscussion

In order to evaluate the performance of the proposed speech conversion method, we convert the speech signal using the proposed voice conversion algorithm, the number of basis vectors R is set to 40, the sparse parameter λ is set to 0.1, and the iteration is set to 20. Figure 27.2a plots the spectrum of source speech signal in the training phase, Fig. 27.2b plots the spectrum of target speech signal in the training phase, and Fig. 27.2c is the spectrum of converted speech signal. As can be seen from Fig. 27.2, the spectrum of converted speech signal is similar to the spectrum of source speech signal and the spectrum of target speech signal.

In order to show that the proposed voice conversion algorithm is much better than the existing conversion algorithm, two more comparative experiments were conducted, that is, the voice conversion algorithm of sparse convolutive non-negative matrix factorization is based on the K-L divergence where the sparse degree λ is 0.1 (K-L-0.1) and the voice conversion algorithm of sparse convolutive non-negative matrix based on the Itakura–Saito distance with the sparse degree λ is set to 0 (IS-0).

Thirty pairs of corresponding female and male speeches were a part of the experiment, where the female speech (source speech) is converted into the corresponding male speech (target speech). We use short-time objective intelligibility (STOI) to evaluate conversion performance. Compared with the source speech, Fig. 27.3 gives the average value of the STOI of the converted speech using these three methods. The horizontal axis is the number of speech sentences, and the ordinate axis is the average value of the STOI. Compared with the target speech, Fig. 27.4 gives the average value of the STOI of the converted speech using these three methods. The horizontal axis is the number of speech sentences, and the ordinate is the average value of the STOI. Through the objective evaluation results, we can find that the proposed voice conversion algorithm gains more intelligibility improvement than the other two algorithms.

Since the proposed objective cost function of the algorithm can make the smaller matrix element with a smaller reconstruction error, we have also made the experiment on whispered conversion to the corresponding normal speech. We select 10 pairs of whispered (source speech) and corresponding normal speech (target speech). Because IS-0 algorithm can hardly convert into corresponding normal speech, we only utilize K-L-0.1 algorithm as comparative experiment, and the experimental process is the same as above. Compared with the source speech,

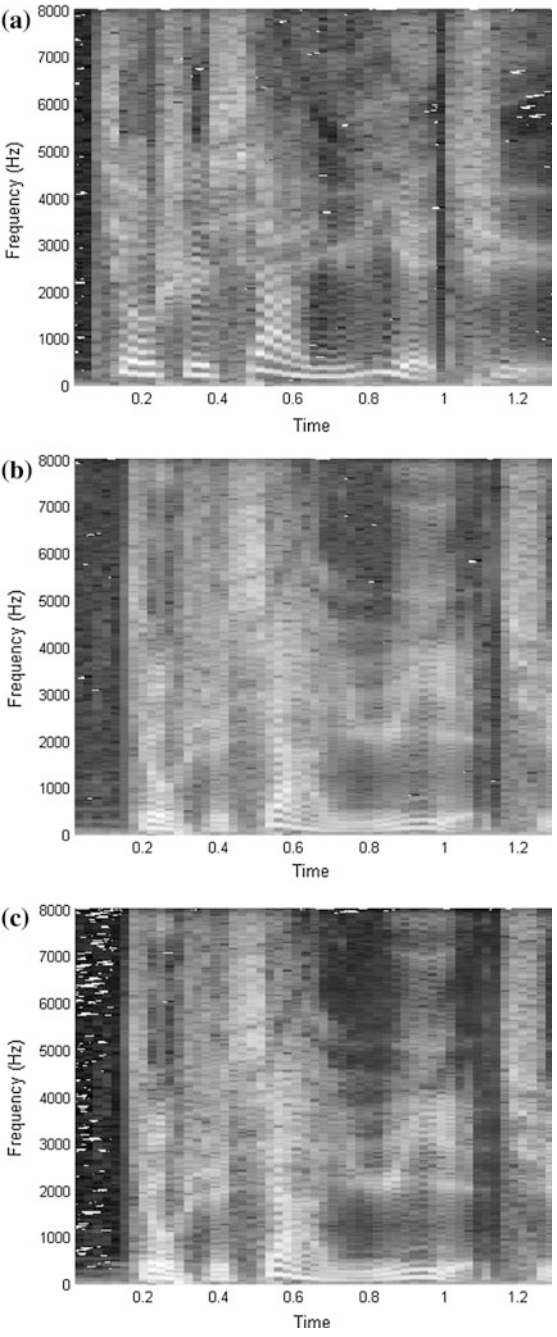


Fig. 27.2 Example of phoneme basis learning: (**a**) is the spectrum of source speech signal in the training phase, (**b**) is the spectrum of target speech signal in the training phase, (**c**) is the spectrum of converted speech signal

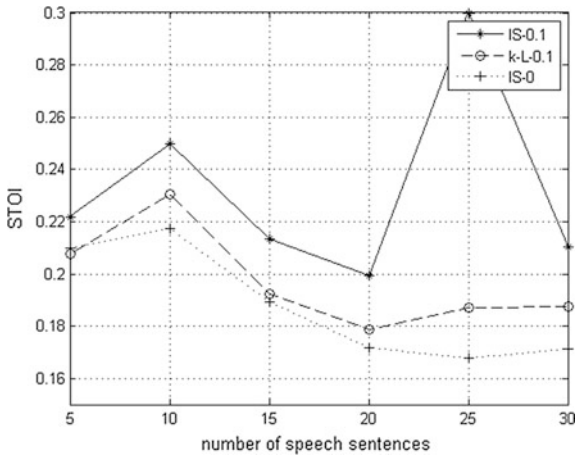


Fig. 27.3 The average value of the STOI compared with the source speech

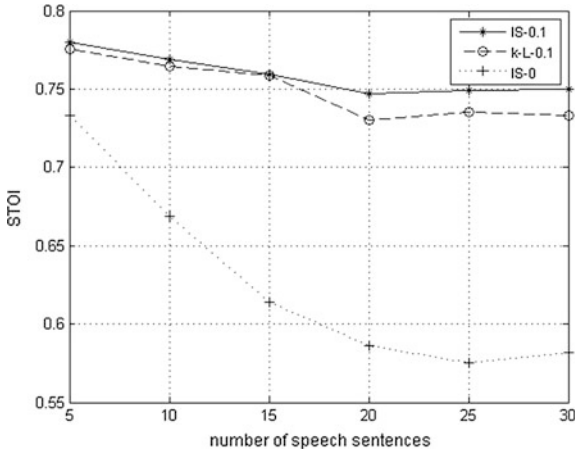


Fig. 27.4 The average value of the STOI compared with the target speech

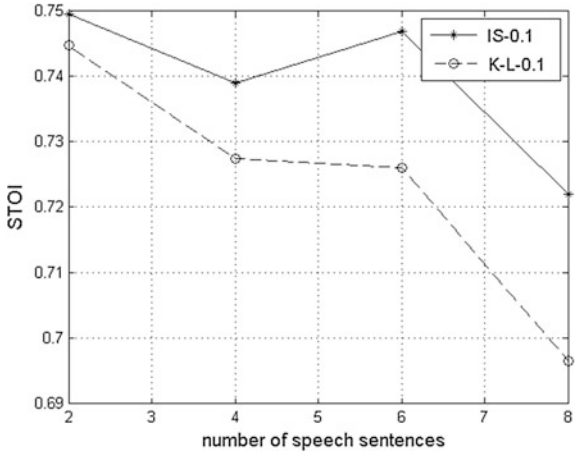


Fig. 27.5 The average value of the STOI compared with the source speech

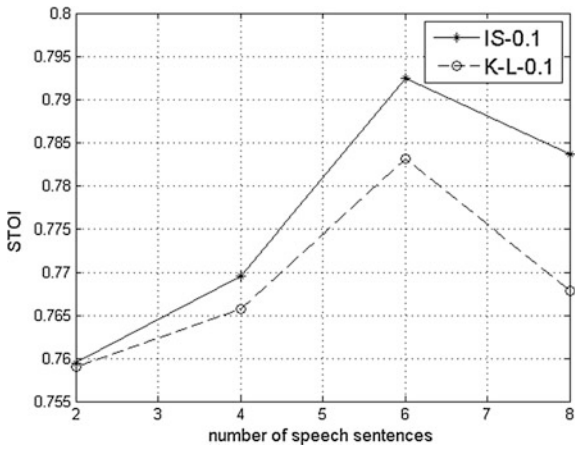


Fig. 27.6 The average value of the STOI compared with the target speech

Table 27.1 Subjective evaluation results of the three algorithms

	Compared to the source speech		Compared to the target speech	
	MOS (Female-to-male)	MOS (Whisper-to-normal)	MOS (Female-to-male)	MOS (Whisper-to-normal)
IS-0	2.13		3.98	
K-L-0.1	2.34	3.84	4.12	3.85
IS-0.1	2.56	3.97	4.33	4.06

Fig. 27.5 gives the average value of the STOI of the converted speech using these two methods. Compared with the target speech, Fig. 27.6 gives the average value of the STOI of the converted speech using these two methods. From the experimental results, we can know that the proposed voice conversion algorithm in this paper has much better effect than another algorithm.

In order to further compare the conversion effect, we utilize mean opinion score (MOS) to evaluate quality of converted speech. Table 27.1 is the subjective evaluation results of the three algorithms. According to the objective evaluation result and subjective evaluation result, we can find that the proposed voice conversion algorithm outperforms other two methods in the aspect of speech quality.

27.5 Conclusion

This paper proposes a voice conversion algorithm based on SCNMF. This method utilizes Itakura–Saito distance as the objective cost function, making the smaller matrix element with a smaller reconstruction error, and at the same time the cost function has the property of scale invariance. Based on this characteristic, we use

time–frequency basis to realize the voice conversion. Experimental results show that the proposed conversion method in this paper outperforms the other two methods, and at the same time the converted whisper gains larger intelligibility.

Acknowledgments This work is supported by the Natural Science Foundation of China (No. 61372137, No. 61301295) and the Anhui Natural Science Foundation (No. 1308085QF100).

References

1. Abe M, Nakamura S, Shikano K, Kuwabara H et al (1988) Voice conversion through vector quantization. In: IEEE international conference on acoustics, speech, and signal processing. IEEE Press, Washington, DC, pp 655–658
2. Stylianou Y, Cappe O, Moulines E (1998) Continuous probabilistic transform for voice conversion. IEEE Trans Speech Audio Process 6(2):131–142
3. Yue Z, Zou X, Wang H (2009) Voice conversion with the combination of HMM and GMM. J Data Acquisition Process 24(3):285–289
4. Yamagishi J, Kobayashi T, Nakano Y et al (2009) Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm. IEEE Trans Audio Speech Lang Process 17(1):66–83
5. Watts O, Yamagishi J, King S (2010) Synthesis of child speech with HMM adaptation and voice conversion. IEEE Trans Audio Speech Lang Process 18(5):1005–1016
6. Desai S, Black AW, Yegnanarayana B et al (2010) Spectral mapping using artificial neural networks for voice conversion. IEEE Trans Audio Speech Lang Process 18(5):954–964
7. Duxans H, Bonafonte A, Kain A et al (2004) Including dynamic and phonetic information in voice conversion systems. In: 8th international conference on spoken language processing. Jeju Island, Korea, pp 5–8
8. Toda T, Black AW, Tokuda K (2007) Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory. IEEE Trans Audio Speech Lang Process 15(8):2222–2235
9. Zen H, Nankaku Y, Tokuda K (2010) Continuous stochastic feature mapping based on trajectory HMMs. IEEE Trans Audio Speech Lang Process 19(2):417–430
10. Sun J, Zhang X, Cao T et al (2013) Voice conversion based on convolutive nonnegative matrix factorization. J Data Acquisition Process 28(2):141–148
11. Ma Z, Zhang X, Yang J et al (2013) Voice conversion based on sparse convolutive nonnegative matrix factorization. J Data Acquisition Process 34(2):1–7
12. Fumitada I, Shuji S (1970) A statistical method for estimation of speech spectral density and formant frequencies. Electron Commun 53(A):36–43
13. Smaragdis P (2007) Convolutive speech bases and their application to supervised speech separation. IEEE Trans Audio Speech Lang Process 15(1):1–12
14. Lee D, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791
15. Lee D, Seung HS (2001) Algorithms for non-negative matrix factorization. In: Advances in neural information processing systems. MIT Press, Cambridge, Mass, USA, pp 556–562
16. Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. J Mach Learn Res 5:1457–1469

Chapter 28
Research of Monitoring Method for the Protection of Nitrogen-Based Atmosphere of Heat Treatment Furnaces Based on DPCA

Dakuo He, Kai Zhang, Zhen Zhang and Shouxin Sun

Abstract Aviation structural material is a kind of special material for aircraft components. This material is usually produced in nitrogen-based atmosphere to protect furnace. To ensure the production operation safety and to guarantee product quality, process monitoring technology has been paid essential attention in this field. By analyzing the heating process, details including input monitoring variables, process parameters, and data pretreatment method are determined. Then DPCA method is used to deal with problems occurred in the nitrogen-based atmosphere to protect furnace heating process. Simulation experiment based on data collected from production field is conducted. The DPCA monitoring method is verified to be effective by the experiment results, and is proved with better performance in comparison with results of MPCA.

Keywords Nitrogen-based atmosphere protect furnace · Process monitoring · DPCA · Simulation research

28.1 Introduction

Special materials are usually produced in atmosphere of heat treatment furnace, which is mainly used for heating without oxygen. As a controllable atmosphere of heat treatment furnace, nitrogen-based atmosphere of heat treatment furnace is widely applied to produce special steel with extraordinary properties especial for aviation structural material. This material is mainly used to produce important components like aircraft girders, engine cradle, and others. Such furnace helps to make chemical heat treatment more effective, and reduce flaws in this process. Besides, nitrogen-based atmosphere of heat treatment furnace has superior advantages in

D. He (✉) · K. Zhang · Z. Zhang · S. Sun
College of Information Science and Engineering,
Northeastern University, Shenyang, China
e-mail: hedakuo@ise.neu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_28

saving fuel, lower cost, better security, and less pollution than traditional heat treatment furnace in [1].

Relative researches for heat treatment furnace have been conducted in modeling, control methods, and monitoring methods. Zhang summarized the development of advanced control techniques for reheating furnace in [2]. Rong introduced in his article new schemes of process control for reheating furnace, including intelligent PID control and advanced intelligent control in [3]. Lin discussed the comparison of modeling methods for reheating furnace in [4]. Chai put forward the optimal setting model of reheat furnace temperature to reduce the consumption of production in [5]. Traditional monitoring methods using the temperature monitored by infrared and laser are also mentioned in [6].

In this article, by analyzing the features of data collected from production field, PCA method is used to monitor product process. Because of the batch production process in nitrogen-based atmosphere of heat treatment furnace, traditional PCA method is not conducive to deal with problems inside. Considering the information between batch and batch, the DPCA method is used to handle this issue. By simulation experiment, this method is verified effective, and is proved to have better performance in the comparison with MPCA method.

The organization of this article is as follows. The production process is introduced at Sect. 28.2. Regular fault conditions are also mentioned in this part. Then the monitoring method is briefly reviewed in Sect. 28.3 including the DPCA method and the process monitoring statistics. While Sect. 28.4 illustrates the method proposed to be applied in the field production process. Relative simulation experiment is conducted to compare results of DPCA with those of MPCA. The conclusion remarks are given in Sect. 28.5.

28.2 Production Process Description

The protection of nitrogen-based atmosphere of heat treatment furnaces can be used to protect quenching, normalizing, and carburizing processes. It consists of several components including main heating furnace, nitrate slot, unloading vehicles, and oil tank. The production process has several procedures like loading materials, loading furnace, heating, unloading furnace, transforming, and quenching.

In order to produce and operate on various types of products with the same suit of furnace, batch production is used in this process. This method is widely used in aviation manufacturing for its advantages in flexibility and variability. Different from continuous product process, batch product process is the combination of many steady states, usually changing from one steady state to another. Besides, variables changing with time and the uncertainties of running time in different batches added the difficulties to guarantee product qualities. This may easily lead to quality problems and even industrial accidents in production line.

In order to get the material properties needed, constant temperature is needed for steel quenching in nitrogen-based atmosphere. Otherwise faults like carbon

increasing or decreasing at the surface and shape deformation can be caused. According to the standards for nitrogen-based atmosphere of heat treatment, the decarburization layer after heat process should be no more than 0.075 mm. Production with more or less quantities of carbon and nitrogen is useless and should be abandoned. So the carbon potential influences most on the product quality. Because of the difficulties in measuring carbon potential in the furnace, oxygen potential is taken as monitoring variable to reflect its changes. Oxygen potential monitoring in the production process is usually affected by carbon accumulation in oxygen sensor and shows abnormal oxygen potential recoveries, including recovering too slow when heating and failing to maintain at heating preservation period. Judging by this, the oxygen potential recovery time in the main heating furnace is chosen to monitor and reflect changes in the furnace.

28.3 Monitoring Methods

28.3.1 PCA Method

PCA method is one of the most important methods for fault detection in online monitoring of process industry in [7]. Since its appearance, contribution chart, multi-block approach, and other technology have been combined with the PCA method in different situations in [8]. Besides, subspace approach based on fault reconstruction is effective for fault isolation and diagnosis in [9].

The principal components extract process in PCA is usually by the spectral decomposition of covariance matrix for the matrix X in [10]. Image the pending dataset is $X(m \times n)$, sample number is m, and variable number is n. The standardization makes the sample dataset center coincides with that of coordinate origin, eliminating the pseudo-variation information caused by different dimension, and improves the analysis result rationality. This process can be described by

$$X_k = \frac{X_k - E(X_k)}{\sqrt{\mathrm{Var}(X_k)}} \tag{28.1}$$

While the decomposition of covariance matrix for the matrix X is

$$V = \frac{1}{n} X^{\mathrm{T}} X \tag{28.2}$$

$$V = P \Lambda P^T \tag{28.3}$$

The Λ stands for the diagonal matrix of covariance matrix V, including the nonnegative eigenvalues $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0)$ of decreasing amplitude, P is orthogonal matrix, the unitization eigenvectors corresponding to eigenvectors.

In this article, the principal component number is determined by the accumulative contribution rate of principal component variance, shown as Eq. 28.4

$$\sum_{i=1}^{d} \lambda_i \Big/ \sum_{i=1}^{m} \lambda_i \geq 85\,\% \tag{28.4}$$

When the accumulative variance of the first d principal components is more than the threshold set like 0.85, the first d principal components are chosen as the integrated indicators, and the m-dimensional space is decreased to n-dimensional space, $d \leq m$.

To maximize the data change information and minimize random noise effect on PCA, the first d lines of loading matrix P is chosen to correspond to loading vector related to the first d eigenvalues. The projection to low-dimensional space from X is contained in score matrix T, which can be got by Eq. 28.5

$$T = XP \tag{28.5}$$

28.3.2 DPCA Method

The PCA method has advantage to extract the most of dynamic data change information without accurate process mechanical model in [11]. When PCA is applied to dynamic process, the information of the auto-correlations is not taken into account. Considering the relationship between current monitor values and that of the past, DPCA method is needed to solve this problem in [12]. By expanding data matrix with delay data, PCA capture dynamic data connection on the augmented matrix, to build a multivariate autoregressive (AR) model.

A process can be represented by multi-element second-order AR model

$$X_k = A_1 X_{k-1} + A_2 X_{k-2} + \xi_k \tag{28.6}$$

The process variables $X_k \in R^n$, coefficient matrix $A_1, A_2 \in R^{n\times n}$, white noise $\xi_k \in R^n$; intercept m ($m > n$) sample data from this model are to constitute PCA data matrix $X^{d0} \in R^{m\times n}$, where superscript "$d0$" means 0 delay.

For process with s delay, the data matrix is

$$X^{ds} = \begin{bmatrix} X_k^{\mathrm{T}} & X_{k-1}^{\mathrm{T}} & \cdots & X_{k-\mathrm{s}}^{\mathrm{T}} \\ X_{k+1}^{\mathrm{T}} & X_k^{\mathrm{T}} & \cdots & X_{k-\mathrm{s}+1}^{\mathrm{T}} \\ \vdots & \vdots & \vdots & \vdots \\ X_{k+m-1}^{\mathrm{T}} & X_{k+m-2}^{\mathrm{T}} & \cdots & X_{k-\mathrm{s}+\mathrm{m}-1}^{\mathrm{T}} \end{bmatrix} \tag{28.7}$$

The delay length can be determined by the following procedure, as is shown in Fig. 28.1.

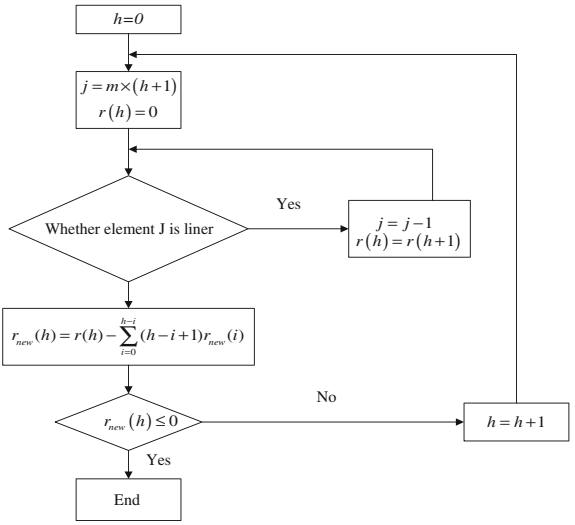


Fig. 28.1 Determination of delay length

By experience, the delay length is usually set to be 1 or 2 in DPCA method in [13]. The optimal order can be determined by the criterion that the number of new relationships will increase with the order until it is optimal in [14].

28.3.3 Process Monitoring Statistics

As process monitoring tool based on PCA model, statistics describes the characteristics of production process, mainly including T2 and SPE statics. The identification ability is improved by the index combining SPE and T2 statics in [15].

The model deviating degree of each principal component in change trend and attitude can be described by T2, to measure changes inside the model. T2 is defined as Eq. 28.8

$$\mathrm{T2} = tS^{-1}t^{\mathrm{T}} \tag{28.8}$$

The principal vector is represented by T, standing for the I line in matrix $T_A = [t_1, t_2, L, t_A]$. The diagonal matrix consisted of eigenvalues corresponding to the first A principal components is written by S.

Statics that represent the principal model deviating degree in actual value of input variables goes as SPE, as the measure of output variables changes, which is defined as Eq. 28.9

$$\mathrm{SPE} = ee^{\mathrm{T}} = (X - \hat{X})(X - \hat{X})^{\mathrm{T}} \tag{28.9}$$

In the equation above, e is the deviation between X and the reconstitution matrix, $\hat{X} = TP$ is the estimated value calculated by the reconstitution of principal component scores and load vectors.

The kernel density estimation method is used to calculate the corresponding monitoring index control limit, calculated by Eq. 28.10

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left\{\frac{\bar{x} - x_i}{h}\right\} \tag{28.10}$$

x_i is T2 and SPE statics to be calculated; $\bar{x}$ is the estimation center; h is bandwidth parameter; n is the data number to estimate; $K\{\cdot\}$ is kernel function, usually using Gaussian kernel function.

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \tag{28.11}$$

Calculate the upper control limit by golden search method while the confidence coefficient is α, the upper limit $Z_{d\alpha}$ is got by the following Eq. 28.12:

$$\frac{1-\alpha}{2} = \int_{-\infty}^{Z_{d\alpha}} \hat{f}(x)\mathrm{d}x \tag{28.12}$$

So as to calculate the lower limit $Z_{u\alpha}$:

$$1 - \frac{1-\alpha}{2} = \int_{-\infty}^{Z_{u\alpha}} \hat{f}(x)\mathrm{d}x \tag{3.13}$$

28.4 Simulation Research

After the preprocess treatment on redundant data, missing data and abnormal data, 320 batches of normal sample data are selected, with 200 sample time in each batch. 300 batches are chosen for modeling training, the rest 20 for online monitoring.

Figures 28.2 and 28.3 of normal data show that for MPCA and DPCA, T2 and SPE statistics are within control limit.

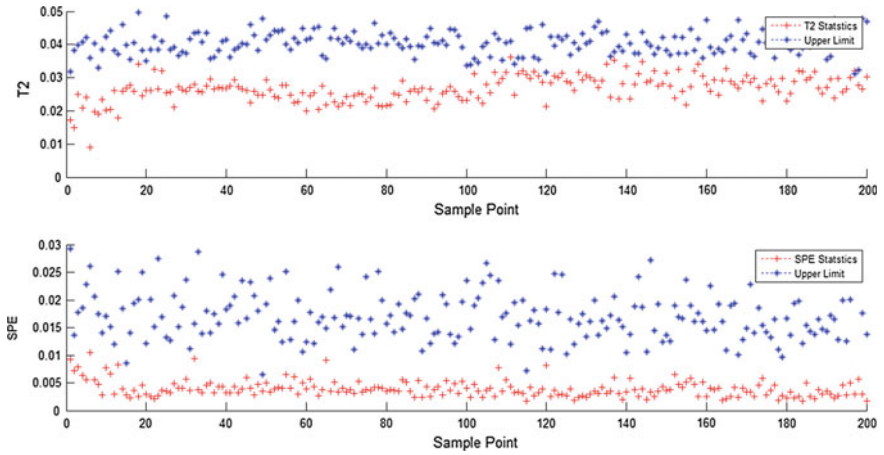


Fig. 28.2 T2 and SPE statistics of normal condition for MPCA

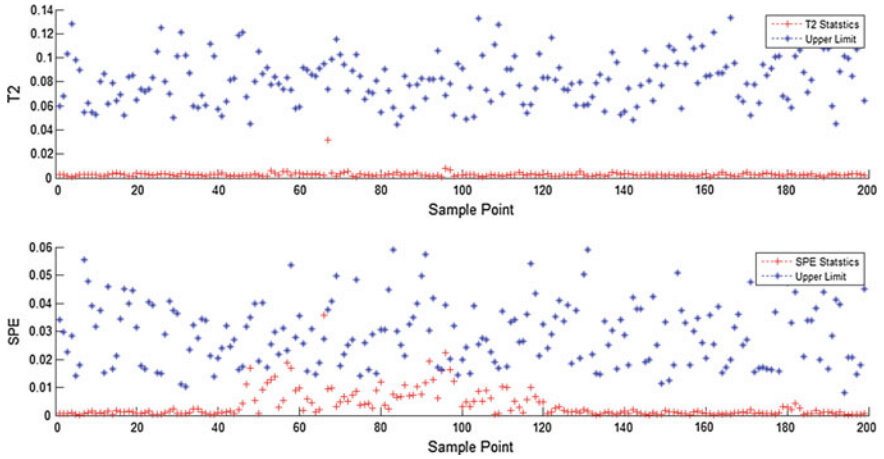


Fig. 28.3 T2 and SPE statistics of normal condition for MPCA

Then is the monitoring simulation for oxygen potential fault. Oxygen potential of abnormal condition recovers slower than that of normal condition. The T2 and SPE statics for fault condition are shown as Figs. 28.4 and 28.5.

From Figs. 28.4 and 28.5, it can be shown that both MPCA and DPCA can detect faults in carbon potential, and DPCA can detect the faults relatively earlier.

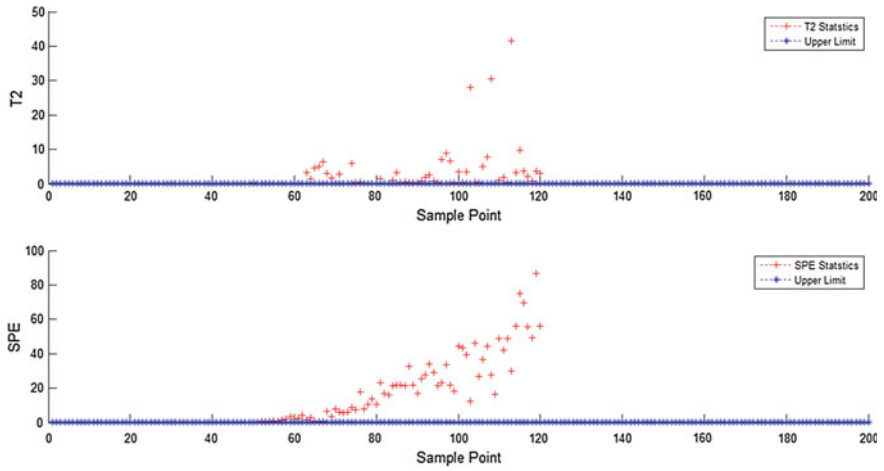


Fig. 28.4 T2 and SPE statistics of fault in carbon potential for MPCA

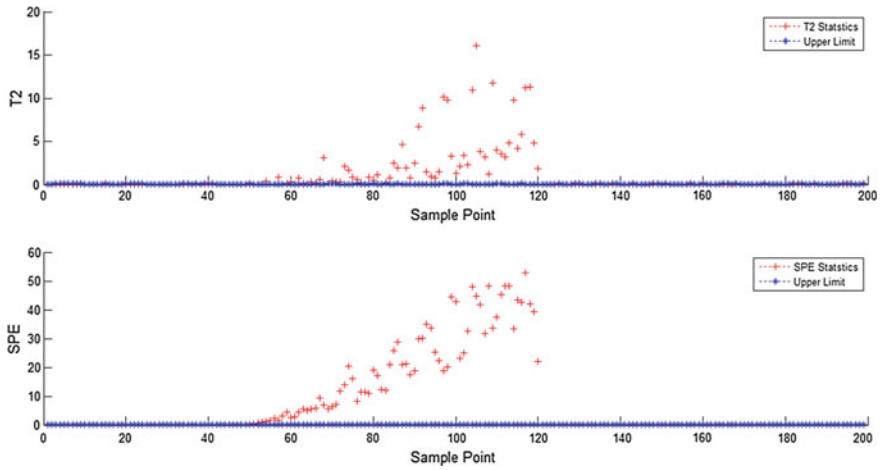


Fig. 28.5 T2 and SPE statistics of fault in carbon potential for DPCA

28.5 Conclusion

To detect abnormal conditions in the production process of the protection of nitrogen-based atmosphere of heat treatment, DPCA method is used as monitoring methods. Carbon potential is chosen as the variable to be monitored. By collecting normal data for DPCA modeling, calculating monitoring statics, and the control limit, the monitoring model were built to prove DPCA methods available to detect carbon potential faults with simulation data. Comparison with MPCA method

shows that DPCA method can detect the faults relatively earlier, with better performance in solving dynamic problems when data of different sample time is related.

Acknowledgments This work is supported by the National Nature Science Foundation of China under Grant 61374147 and 61004083 and 973 Project No. 2009CB320601 and 863 Project No. 2011AA060204 and the Fundamental Research Funds for the Central Universities N120404014.

References

1. Chen ZY (2005) Research progress in nitrogen based atmosphere. Powder Metall Technol 23 (4):291–295
2. Zhang KJ, Shao C (2003) Advanced control techniques for reheating furnace and their development in iron and steel industry. Metall Ind Autom 27(1):11–15
3. Rong L, Chai TY, Qian XL (2000) New scheme of process control for reheating furnace—intelligent control. Control Decis 15(3):269–273
4. Lin T, Duan WZ, Li CC, Zuo WH (2008) Comparison of modeling methods for reheating furnace. Comput Simul 25(3):95–99
5. Chai TY, Wang ZJ, Zhang L (2000) Optimal setting model of reheat furnace temperature. Acta Automatica Sinica 26(1):537–541
6. Dunia R, Qin SJ (1998) Joint diagnosis of process and sensor faults using principal component analysis. Control Eng Pract 6(4):457–469
7. Liang J, Qian JX (2003) Chinese. J Chem Eng 11:191
8. MacGregor JF, Jaeckle C, Kiparissides C, Koutoudi M (1994) Process monitoring and diagnosis by multiblock PLS methods. AIChE J 40:826
9. Dunia R, Qin SJ (1998) Subspace approach to multidimensional fault identification and reconstruction. AIChE J 44:1813
10. Wang QH, Yu SM (2004) Selection of number of principal components based on fault diagnosing performance optimization. J Chem Ind Eng 55(2):214–219
11. Li RY, Rong G (2006) Dynamic process fault isolation by partial DPCA. Chem Biochem Eng Q 20(1):69–77
12. Wold S (1987) Multi-way principal components and PLS analysis. J Chemom 1:41–56
13. Chiang LH, Russell EL, Braatz RD (2000) Fault detection and diagnosis in industrial systems. Springer, London
14. Wise BM, Gallagher NB (1996) The process chemometrics approach to process monitoring and fault detection. J Process Control 6:329
15. Yue H, Qin SJ (2001) Reconstruction-based fault identification using a combined index. Ind Eng Chem Res 40:4403

Chapter 29
Research of Process Monitoring and Fault Trace System for Annealing Furnace of Hot-Dip Galvanizing Line

Dakuo He, Kai Zhang, Da Li and Qihao Wu

Abstract Annealing furnace is the key equipment of hot galvanizing production line. Process monitoring and fault trace for annealing furnace is quite essential for its essential influence on the product quality. In view of the complex annealing process and the nonlinear relationship between the fault reasons and symptoms, nonlinear methods kernel principal component analysis (KPCA) is used to monitor the condition and support vector machine (SVM) is used for fault trace. Method based on data reconstruction is used to distinguish the fault variables. Depending on the monitoring method, simulation study for annealing furnace of galvanizing line is conducted. This method proposed is proved usable through results of the simulation experiments.

Keywords Hot-dip galvanize · Annealing furnace · KPCA · SVM · Process monitoring · Fault diagnosis · Fault trace

29.1 Introduction

As the key implement in hot galvanizing production line, annealing furnace condition directly affect quality of galvanized production. Faults in this procedure affect quality of galvanized sheets and manufacture of the whole production line. Otherwise great damage to the key equipment can be caused, threatening the safety of operators in production field. Process monitoring and faults diagnosis in annealing furnace helps to find and faults in the system. Study on this hot galvanizing production line is of great theoretical and practical significance.

Principal component analysis (PCA) is widely used in monitoring the process performance and fault diagnosis in [1]. This method is effective in dimensionality reduction, noise removal, and feature extraction from the original data set as a

D. He (✉) · K. Zhang · D. Li · Q. Wu
College of Information Science and Engineering, Northeastern University, Shenyang, China
e-mail: hedakuo@ise.neu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_29

preprocessing. As a linear method, PCA is inadequate in solving nonlinear condition in [2]. To solve this problem, a nonlinear PCA method based on auto-associative neural networks is proposed by Kramer in [3]. Tan and Mayrovouniotis developed a nonlinear principal component method based on input-training neural network in [4]. Dong and MaAvoy developed a nonlinear PCA approach based on principal curves and two three-layer neural networks in [5]. Saegusa et al. developed a hybrid neural network approach in [6]. Another approach based on transform space is kernel principal component analysis (KPCA) in [7, 8].

Support vector machines (SVM) were first suggested by Vapnik in [9] and have recently been used in a range of problems including pattern recognition in [10]. Bioinformatics in [11] and text categorization in [12] has also been discussed. Based on this, Huang and Wang proposed a genetic algorithm approach for feature selection and parameters optimization in [13]. This method is applied to software effort estimation by Adriano et al. in [8].

In this article, the production process is briefly introduced in the first part. In this part several faults as well as their behaviors in annealing furnace are mainly discussed. Then the monitoring method including KPCA method and SVM is introduced. Relative solution is proposed for the practical problems in the application of KPCA in the annealing process of hot-dip galvanizing production line. The monitoring statistic and limits are calculated with normal sampling data to establish monitoring model. SVM model is trained with the training data set. Cross-validation method is used to improve the generalization ability of the model. Simulation experiment is conducted to prove this method available.

29.2 Production Process Description

Steel strip continuous hot galvanizing is the process where cold-rolled steel strip is galvanized in continuous production line. During this process, the steel strip is decoiled by the decoiler, washed in washing section, annealed at the annealing furnace, and cooled down at the cooling tower. The whole production line can be divided into three parts: entrance section, production process and exit section. Procedure of annealing furnace production process is shown as Fig. 29.1.

Annealing furnace gradually deteriorate when used for long time. Faults in the furnace may lead to the whole furnace breaking down, and even cause serious industrial accidents. Ordinary faults are listed as follows:

1. Flaws in furnace body and furnace roller. Monitoring variables of this fault indicate that dew point at the leakage increases a lot and oxygen content inside the furnace is relatively higher.
2. Leakage of cooling water at furnace roller. Monitoring variables of this faults shows that dew point at the leakage suddenly increases while the oxygen content is relatively normal.

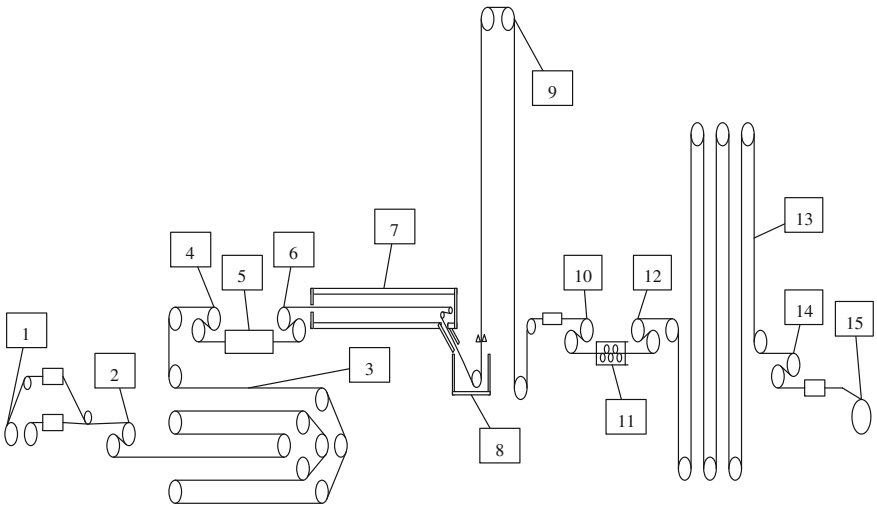


Fig. 29.1 The steel heat treatment process

3. Water loss of heat exchange at cooling section. Monitoring variables of this faults shows that dew point at rapid section suddenly increase while the oxygen content is relatively normal. When the water loss at heat exchanger is serious, dew point there is higher than −5° or even positive, at the same time with no regular fluctuation.
4. Leakage of heat exchanger at preheating section. Monitoring variables of this fault shows that dew point and oxygen at preheating section would suddenly rise.

29.3 Monitoring Methods

29.3.1 KPCA Method

A distribution with n data points $x_i \in Rm$ $i = 1, \ldots, n$ is discussed here. These data points are mapped into a higher dimensional feature space H, Φ: $x_i \rightarrow H$. Here $\Phi(x_i)$ is assumed already mean-centered and variance-scaled. Calculating the covariance matrix C^H in H, equation $C^H = \frac{1}{n}\sum_{\substack{i=1 \\ j=1}}^{n} \Phi(x_i)\Phi(x_i)^T$ to solve the following eigenvalue problem.

$$\lambda v = C^H v \tag{29.3.1}$$

here eigenvalue $\lambda \geq 0$ and $v \in F/\{0\}$. All solutions v with $v \neq 0$ must lie in the span of $\Phi(x_1), \ldots, \Phi(x_n)$, $v = \sum_{i=1}^{N} \alpha_i \Phi(x_i)$ Eq. (29.3.1) can be described as

$$n\lambda\alpha = K\alpha \quad \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n] \tag{29.3.2}$$

K is the Gram matrix and is defined by $[K]_{ij} = K_{ij} = \langle \Phi(x_i), \Phi(x_j) \rangle$, and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are normalized. The corresponding vectors in H are normalized like $\langle v_i, v_j \rangle = 1$, so $\lambda_i \langle \alpha^i, \alpha^j \rangle = 1$. The principal components t of a test vector x are then extracted by projecting $\Phi(x)$ onto eigenvectors v_k in H where

$$t_k = \langle v_k, \Phi(x) \rangle = \sum_{i=1}^{n} \alpha_i^k k(x_{i,}, x) \quad k = 1, \ldots, n \tag{29.3.3}$$

In KPCA method, T^2 statistic and the Q statistic in the feature space can be interpreted the way like PCA. A measure of variation within the KPCA model is given by Hotelling's T^2 statistic, standing for the sum of the normalized squared scores, and can be defined as

$$\mathrm{T}^2 = [t_1, \ldots, t_p] \Lambda^{-1} [t_1, \ldots, t_p]^{\mathrm{T}} \tag{29.3.4}$$

where t_k can be obtained from above and Λ^{-1} is the diagonal matrix of the inverse of the eigenvalue associated with the retained PCs. The confidence limit for T^2 is obtained by the F-distribution

$$\mathrm{T}^2_{p,N,\alpha} \sim \frac{p(n-1)}{n-p} F_{p,n-p,\alpha} \tag{29.3.5}$$

Here N means the number of samples in the model and p for the number of PCs. The measure of goodness of fit for a sample to the PCA model is the squared prediction error, which is known as the Q statistic. In the feature space, SPE is defined as follows:

$$\mathrm{SPE} = \left\| \phi(x) - \phi_p(x) \right\|^2 = \sum_{j=1}^{n} t_j^2 - \sum_{j=1}^{p} t_j^2 \tag{29.3.6}$$

where n is the number of nonzero eigenvalue generated from Eq. (29.3.2). The confidence limit for the SPE can be computed from its approximate distribution:

$$\mathrm{SPE}_\alpha \sim g\chi_h^2 \tag{29.3.7}$$

Here $g = b/2a$ and $h = 2a2/b$, a and b are the estimated mean and variance of the SPE in [14].

29.3.2 SVM Method for Fault Trace

SVM is widely applied in classification of remotely sensed data, but now has been extended from solving two-class problem to deal with Q class problem ($Q > 2$). It is an optimal statistical tool for numerous, heterogeneous variables, mixtures of qualitative and quantitative variables. According to Waske and Benediktsson, range of the most commonly used classifiers are compared for multisource classification, namely maximum likelihood, decision trees, boosted decision trees, and SVM in [15]. According to their results, SVM gives the best accuracy.

Figure 29.2 shows the basic principle of SVM, as a binary classifier for its first use. Within a multi-dimensional feature space, two classes of training samples are used by SVM to fit an optimal separating hyperplane. Vector components in each dimension are biophysical variables. Distance between the closest training samples, or support vectors, and the hyperplane itself is maximized.

To project data vectors into a high-dimensional feature space, Kernel function is used in SVM. Then the optimal hyperplane is fit and classes are separated by an optimization function. For a generic pattern x, the corresponding estimated label $\hat{y}$ is given by

$$\hat{y} = \text{sign}[f(x)] = \text{sign}\left(\sum_{i=1}^{N} a_i y_i K(x_i, x) + b\right) \tag{29.3.8}$$

where the number of training points is N, y stands for the label of the ith sample, while b is a bias parameter, $K(x_i, x)$ is the chosen kernel function and a denotes the Lagrangian multipliers. The radial basis function (RBF) is taken as the kernel

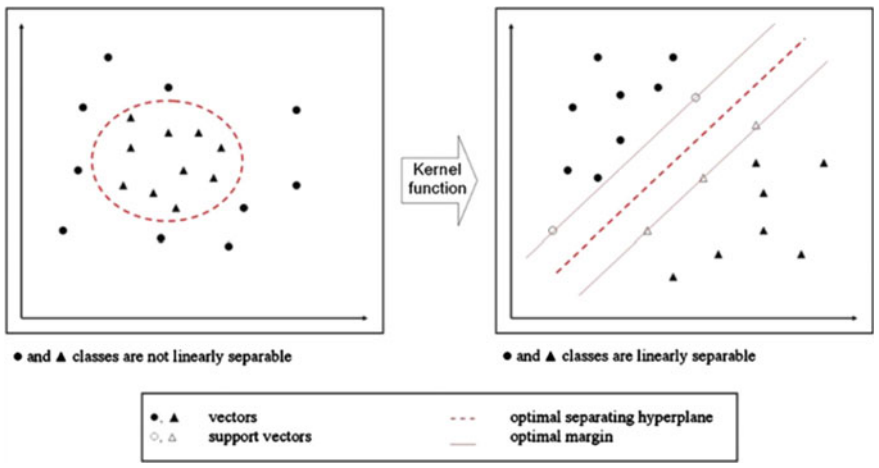


Fig. 29.2 SVM classifier schema

$$K(x_i, x) = \exp \frac{\|x_i x\|^2}{2\sigma^2} \tag{29.3.9}$$

Noise in the data can be accounted by defining a distance tolerating the data scattering. The decision constraint is then relaxed [15].

29.4 Simulation Research

To verify the availability of KPCA monitoring method on annealing furnace of hot-dip galvanizing line, monitoring variables data is selected in state monitoring for simulation research.

First, 1500 groups of normal data coming from the production line are selected. Then select 1000 points from 1500 normal data for KPCA modeling, the left 500 for monitoring test. Figure 29.3 is the T^2 and SPE statistics.

It can be inferred from the simulation result that statistics for normal data are within control limit. Because of the random noise in data, statistics of several moments are beyond control limit, and this can be seen as normal.

Data of fault 4 as example can be taken to verify KPCA method for abnormal condition. Fault 4 is introduced after the 50th group to see whether the fault can be diagnosed.

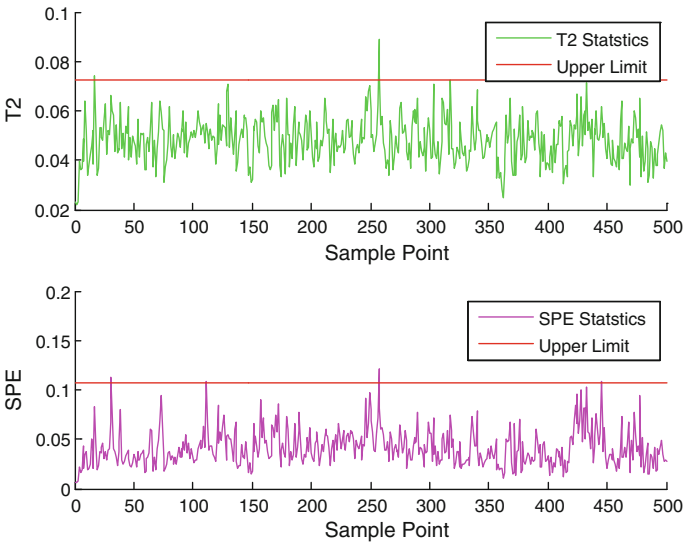


Fig. 29.3 T^2 and SPE statistics of normal condition

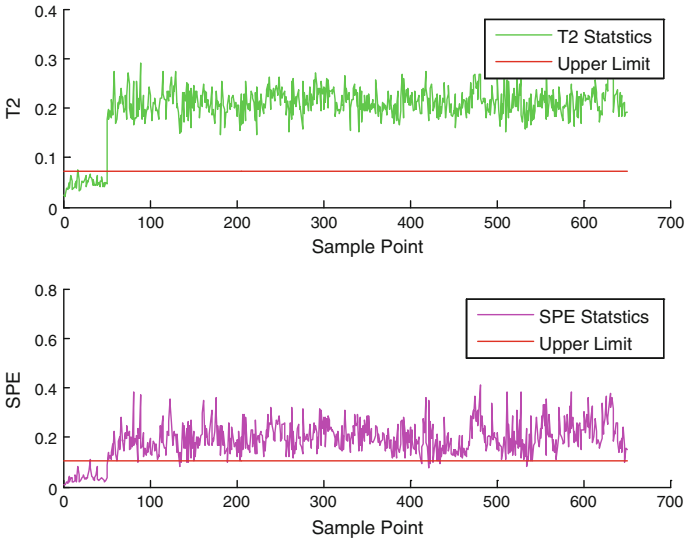


Fig. 29.4 T^2 and SPE statistics for abnormal condition

Figure 29.4 shows that T^2 and SPE statistics are both beyond control limit after introducing fault 4 data, and the validity of KPCA method on process monitoring is testified.

Simulation for oxygen potential fault is then monitored. Oxygen potential of abnormal condition recovers slower than that of normal condition. The meaning of SPE statistics is the change that has not been explained by principal component model in the initial data. Therefore, in this paper, only the fault variables trace of SPE statistics is onsidered. The SPE statistics for fault condition are shown in Fig. 29.5.

It can be seen that the 8th and 20th variable are abnormal when using fault 4 data. In reality the fault 4 is caused by leakage of heat exchanger at the preheating

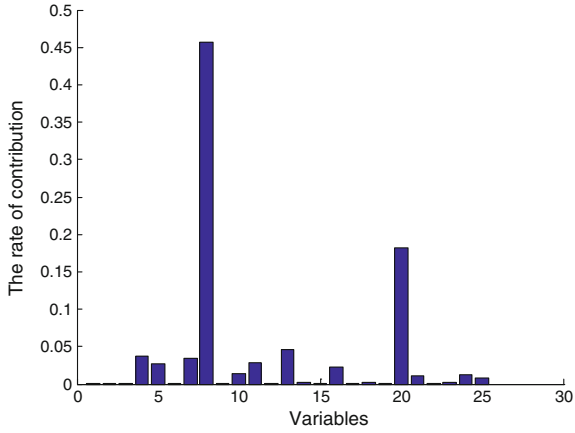


Fig. 29.5 Fault variables trace

Table 29.1 Results of SVM training test after optimizing parameters

Fault type	Accuracy rate of training (%)	Accuracy rate of testing (%)
Normal	100	98
Fault 1	100	98.75
Fault 2	100	99.75
Fault 3	100	100
Fault 4	100	95
Average accuracy	100	98.3

section, when the 20th variable oxygen content and the 8th variable dew point suddenly rise. This result corresponds to the real condition.

Train the SVM model with experiment data and test with testing dataset. The test result is shown in Table 29.1.

It can be seen from the result that SVM model trained by optimized parameters based on cross-validation and grid search methods is effective for normal data and four types of faults, which corresponds to the reality.

29.5 Conclusion

In this article, the KPCA method is used for state monitoring and SVM method for fault trace. The process monitoring system is designed based on KPCA method and SVM method for annealing furnace of hot-dip galvanizing line. By modeling and monitoring the selected process, the production process of decarburization and annealing with complex mechanism is conducted by simulation experiment. Simulation results show that faults can be diagnosed with alarm when production condition is abnormal, and relative fault variables can be detected. This method is verified usable to solve this problem.

Acknowledgments This work is supported by the National Nature Science Foundation of China under Grant 61374147 and 61004083 and 973 Project No. 2009CB320601 and 863 Project No. 2011AA060204 and the Fundamental Research Funds for the Central Universities N120404014.

References

1. Wold S (1978) Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 20(4):397–405
2. Palus M, Dvorak I (1992) Singular-value decomposition in attractor reconstruction: pitfalls and precautions. Physica D 55:221–234
3. Kramer MA (1991) Nonlinear principal component analysis using auto-associative neural networks. AIChE J 37(2):233–243

4. Tan S, Mayrovouniotis ML (1995) Reducing data dimensionality through optimizing neural network inputs. AIChE J 41(6):1471–1480
5. Dong D, MaAvoy TJ (1996) Nonlinear principal analysis based on principal cure and neural networks. Comput Chem Eng 20(1):65–78
6. Saegusa R, Sakano H, Hashimoto S (2004) Nonlinear principal component analysis to preserve the order of principal components. Neurocomputing 61:57–70
7. Mika S, Schölkopf B, Smola AJ, Muller KR, Scholz M, Ratsch G (1999) Kernel PCA and de-noising in feature spaces. Adv Neural Inf Proc Syst 11:536–542
8. Schölkopf B, Mika S, Burges CJC, Knirsch P, Muller KR, Ratsch G, Alexander JS (1999) Input space versus feature space in kernel-based methods. IEEE Trans Neural Netw 10 (5):1000–1016
9. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
10. Pontil M, Verri A (1998) Support vector machines for 3D object recognition. IEEE Trans Pattern Anal 20(6):637–646
11. Yu GX, Ostrouchov G, Geist A, Samatova NF (2003) An SVM based algorithm for identification of photo synthesis-specific genome features. In: Second IEEE computer society bioinformatics conference, CA, USA, pp 235–243
12. Joachims T (1998) Text categorization with support vector machines. In: Proceedings of European conference on machine learning (ECML), Chemintz, DE, pp 137–142
13. Huang CL, Wang CJ (2006) A GA-based feature selection and parameters optimization for support vector machines. Expert Syst Appl 31(2):231–240 (Original Research Article)
14. Pouteau R, Meyer J-Y, Stoll B (2011) A SVM-based model for predicting distribution of the invasive tree *Miconia calvescens* in tropical rainforests. Ecol Model 222(2011):2631–2641
15. Adriano LIO, Petronio LB, Ricardo MFL, Márcio LC (2010) GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation. Inf Softw Technol 52(11):1155–1166

Chapter 30
Cloud Tasks Scheduling Meeting with QoS

Xinxing Li, Xiaoqing Zhang, Qiongfen Qian, Zixuan Wang, Hang Shang, Manli Huang and Zhejin Lu

Abstract Meeting users' Quality of Service (QoS) requirements is a key problem of tasks scheduling in cloud computing. A cloud tasks scheduling algorithm CTS_QoS based on maximal QoS satisfaction and minimal QoS distance between tasks and resources is presented in this paper. Under meeting maximal QoS satisfaction of user's tasks, CTS_QoS can select the resources with minimal QoS distance to map. Experimental results show that though CTS_QoS cannot guarantee a high resource utilization, it can gain users' QoS satisfaction maximization.

Keywords Cloud computing · Tasks scheduling · QoS constraints

30.1 Introduction

As a new computing model and business model, cloud computing is attracting more interest from the academia and industry [1, 2]. Cloud computing technology is the further development of distributed computing, parallel computing, and grid computing. Cloud computing uses virtualization technology to collect computing resource, storage resource, and bandwidth resource into resource pool with scalability and dynamics, which is provided to different users on demand with Internet as a carrier.

Task scheduling is a very critical and complex issue in cloud computing technology [3–5]. The problem solving not only affects the users' satisfaction, but also

X. Li · X. Zhang (✉) · Z. Wang · H. Shang · M. Huang · Z. Lu
School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China
e-mail: zxqtzy@126.com

Q. Qian
Department 4, Air Force Early Warning Academy, Wuhan, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_30

greatly affects the overall performance of the system. Thus, how to conduct reasonably and efficiently cloud tasks scheduling and improve users' satisfaction have become a key and difficult point in cloud computing environment. Generally, the cloud tasks scheduling problem is an NP-complete problem.

Quality of Service (QoS) is a key issue which needs to be solved in cloud tasks scheduling [4, 6]. QoS in cloud denotes users' requirements when they use cloud computing system. QoS includes many different parameters, such as CPU, cost, access speed, distance, resource stability, and storage capacity. Because there are lots of QoS parameters in users' requirements, meeting different QoS requirements is a huge challenge in cloud tasks scheduling. At present, some researches on cloud tasks scheduling based on QoS constraints usually focus on one-dimensional QoS constraints, such as execution time of tasks or execution cost [7, 8]. The cloud tasks scheduling strategy based on multidimensional QoS constraints will be necessary for the development of cloud computing.

30.2 Model Description

The tasks set is T, $T = \{t_1, t_2,\ldots,t_n\}$, $n = |T|$, n is the number of users' tasks, t_i is ith task of users, $i \in [1, n]$. $t_i = \{\text{tID, tLen, tQoS, tSta}\}$

tID is the only identification of tasks;
tLen is the length of tasks, Unit: MI, Million Instruction;
tQoS = $\{QoS_1, QoS_2,\ldots,QoS_k\}$, which denote the multidimensional QoS requirments of tasks, k is the dimension of QoS;
tSta = {tAlloc, tExecu, tSucc}, which denote the status of user's tasks including three types;
tAlloc denotes the waiting scheduling status of tasks;
tExecu denotes the executing status of tasks;
tSucc denotes the completion status of tasks.

The physical hosts set is P, $P=\{p_1, p_2,\ldots,p_l\}$, $l = |P|$, l is the number of the physical hosts. $P_h = \{\text{pID, pType, pSta}\}$.

pID is the only identification of physical hosts;
pType is the type of physical hosts, such as workstation, mainframe or microcomputer;
pSta is the status of physical machine, pSta = {pFree, pRun}, pFre denotes the idle status of physical hosts, i.e., have not deployed VMs or VMs have not run, pRun denotes the working status of physical hosts.

Because tasks are executed on VMs finally, VMs are defined as resources.

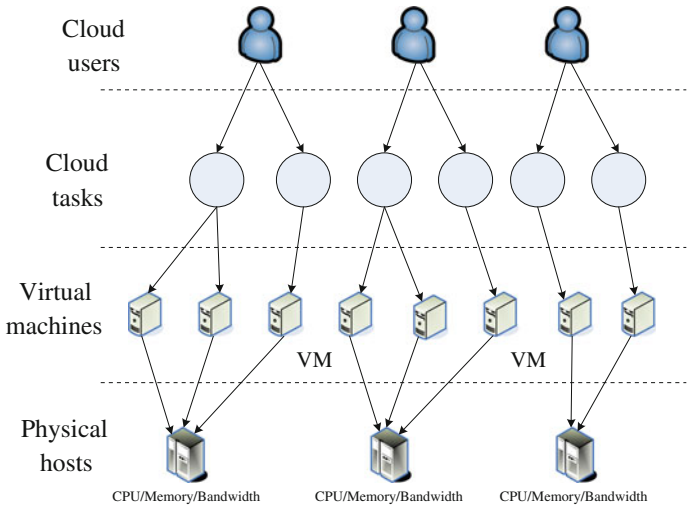


Fig. 30.1 Cloud computing task scheduling model

The resource set is R, $R=\{r_1, r_2,\ldots,r_m\}$, $m = |R|$, m is the number of resources provided by cloud data center. r_j is the jth resource, $j \in [1, m]$. r_j = {rID, rCap, rQoS, rSta, rLoc}.

rID is the only identification of resources;
rCap is the calculation capacity of resources. Unit: MIPS (Million Instruction Per Second);
rQoS = $\{QoS_1, QoS_2,\ldots,QoS_k\}$, which denote the service capacity of resources in multidimensional QoS;
rSta = {rRun, rFre}, which denote the status of resources including two types, rRun is the running status of resources and rFre is the idle status of resources;
rLoc is the physical host resources locate.

Based on the above three types of entities, the tasks scheduling model in cloud computing is shown in Fig. 30.1.

30.3 Method on Multidimensional QoS Constraints-CTS_QoS

30.3.1 Multidimensional QoS Standardization

The QoS dimension of resources is k, the number of resources is m, then, QoS that m resources provide can be defined as $m \times k$ matrix. Let n denote the number of tasks. k-dimension QoS requirement of n tasks can be expressed as $n \times k$ matrix,

$$< ![CDATA[\begin{aligned} \mathrm{QoS}^R_{m,k} &= \begin{bmatrix} \mathrm{qos}^R_{1,1}, & \mathrm{qos}^R_{1,2}, & \ldots, & \mathrm{qos}^R_{1,k} \\ \mathrm{qos}^R_{2,1}, & \mathrm{qos}^R_{2,2}, & \ldots, & \mathrm{qos}^R_{2,k} \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{qos}^R_{m,1}, & \mathrm{qos}^R_{m,2}, & \ldots, & \mathrm{qos}^R_{m,k} \end{bmatrix} \\ \mathrm{QoS}^T_{n,k} &= \begin{bmatrix} \mathrm{qos}^T_{1,1}, & \mathrm{qos}^T_{1,2}, & \ldots, & \mathrm{qos}^T_{1,k} \\ \mathrm{qos}^T_{2,1}, & \mathrm{qos}^T_{2,2}, & \ldots, & \mathrm{qos}^T_{2,k} \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{qos}^T_{n,1}, & \mathrm{qos}^T_{n,2}, & \ldots, & \mathrm{qos}^T_{n,k} \end{bmatrix} \end{aligned}]] > aligned \tag{30.1}$$

Because of the difference between QoS provision of resources and QoS requirement of tasks, the above two matrices should be standardized for evaluating the service capacity of QoS,

First, merge two matrices and form the simultaneous matrix QoS_{m+n}. In the process of standardization, QoS can be divided into positive measure and negative measure. The bigger the value of positive measure, the higher the QoS, such as reliability, security, and stability. The smaller the value of negative measure, the higher the QoS, such as service fee. The positive measure is standardized by Eq. (30.2) and the negative measure is standardized by Eq. (30.3),

$$\mathrm{qos}^R_{i,j} = \begin{cases} 1, & \mathrm{qos}^{\max}_j - \mathrm{qos}^{\min}_j = 0 \\ \frac{\mathrm{qos}_{i,j} - \mathrm{qos}^{\min}_j}{\mathrm{qos}^{\max}_j - \mathrm{qos}^{\min}_j}, & \mathrm{qos}^{\max}_j - \mathrm{qos}^{\min}_j \neq 0 \end{cases} \tag{30.2}$$

$$\mathrm{qos}^S_{i,j} = \begin{cases} 1, & \mathrm{qos}^{\max}_j - \mathrm{qos}^{\min}_j = 0 \\ \frac{\mathrm{qos}^{\max}_j - \mathrm{qos}_{i,j}}{\mathrm{qos}^{\max}_j - \mathrm{qos}^{\min}_j}, & \mathrm{qos}^{\max}_j - \mathrm{qos}^{\min}_j \neq 0 \end{cases} \tag{30.3}$$

where $\mathrm{qos}_{i,\mathrm{j}}$ is the QoS value of ith row and j-column in matrix QoS_{m+n}, $1 \le i \le \mathrm{m} + \mathrm{n}$, $1 \le m \le k$. $\mathrm{qos}^{\min}_j$ and $\mathrm{qos}^{\max}_j$, respectively denotes the minimum and maximum of QoS value of jth column in matrix QoS_{m+n}, $\mathrm{qos}^S_{i,j}$ denotes the QoS value of ith row and jth column after the standardization. The standardized matrix is,

$$\begin{aligned} \mathrm{QoS}^S_{m+n,k} &= \begin{bmatrix} \mathrm{qos}^S_{1,1}, & \mathrm{qos}^S_{1,2}, & \ldots, & \mathrm{qos}^S_{1,k} \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{qos}^S_{m,1}, & \mathrm{qos}^S_{m,2}, & \ldots, & \mathrm{qos}^S_{m,k} \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{qos}^S_{m+n,1}, & \mathrm{qos}^S_{m+n,2}, & \ldots, & \mathrm{qos}^S_{m+n,k} \end{bmatrix}, \\ \mathrm{qos}^S_{i,j} &\in [0,1], \quad 1 \le i \le m+n, \quad 1 \le j \le k \end{aligned} \tag{30.4}$$

Now, the standardized matrix $\mathrm{QoS}^{S}_{m+n,k}$ is separated into two matrices, the standardized resource QoS matrix and the standardized task QoS matrix are represented as $\mathrm{QoS}^{SR}_{m,k}$ and $\mathrm{QoS}^{T}_{n,k}$.

$$\mathrm{QoS}^{SR}_{m,k} = \begin{bmatrix} \mathrm{qos}^{SR}_{1,1} & \mathrm{qos}^{SR}_{1,2}, & \ldots, & \mathrm{qos}^{SR}_{1,k} \\ \mathrm{qos}^{SR}_{2,1} & \mathrm{qos}^{SR}_{2,2}, & \ldots, & \mathrm{qos}^{SR}_{2,k} \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{qos}^{SR}_{m,1} & \mathrm{qos}^{SR}_{m,2}, & \ldots, & \mathrm{qos}^{SR}_{m,k} \end{bmatrix}$$
$$\mathrm{QoS}^{ST}_{n,k} = \begin{bmatrix} \mathrm{qos}^{ST}_{1,1} & \mathrm{qos}^{ST}_{1,2}, & \ldots, & \mathrm{qos}^{ST}_{1,k} \\ \mathrm{qos}^{ST}_{2,1} & \mathrm{qos}^{ST}_{2,2}, & \ldots, & \mathrm{qos}^{ST}_{2,k} \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{qos}^{ST}_{n,1} & \mathrm{qos}^{ST}_{n,2}, & \ldots, & \mathrm{qos}^{ST}_{n,k} \end{bmatrix} \tag{30.5}$$

30.3.2 QoS Satisfaction

Tasks scheduling strategy should satisfy users' QoS requirements by maximization. For the positive measure and negative measure, the QoS satisfaction users receive is calculated by the following equations respectively:

$$\mathrm{QoS}^{\mathrm{Deg}}_{i,j} = \begin{cases} \dfrac{\mathrm{qos}^{SR}_{i,j}}{\mathrm{qos}^{ST}_{i,j}} & \mathrm{qos}^{SR}_{i,j} < \mathrm{qos}^{ST}_{i,j} \\ 1, & \mathrm{qos}^{SR}_{i,j} \geq \mathrm{qos}^{ST}_{i,j} \end{cases} \quad \mathrm{QoS}^{\mathrm{Deg}}_{i,j} = \begin{cases} \dfrac{\mathrm{qos}^{ST}_{i,j}}{\mathrm{qos}^{SR}_{i,j}} & \mathrm{qos}^{ST}_{i,j} < \mathrm{qos}^{SR}_{i,j} \\ 1, & \mathrm{qos}^{ST}_{i,j} \geq \mathrm{qos}^{SR}_{i,j} \end{cases} \tag{30.6}$$

The overall satisfaction of task t_i in all dimensions QoS is

$$\mathrm{QoS}^{\mathrm{Deg}}_{i} = \sum_{j=1}^{n} \mathrm{QoS}^{\mathrm{Deg}}_{i,j} \Big/ k \tag{30.7}$$

The average QoS satisfaction for all users' tasks:

$$\mathrm{QoS}^{\mathrm{Deg}} = \sum_{i=1}^{n} \mathrm{QoS}^{\mathrm{Deg}}_{i} \Big/ n \tag{30.8}$$

30.3.3 QoS Distance Measurement

In order to make tasks with low QoS requirements not occupy high QoS resources, tasks should be scheduled to the resources with similar QoS requirements as far as possible. The weighted distance is used to measure the QoS gap between tasks and resources,

$$D = \sqrt{\sum_{j=1}^{k} w_j \left(\mathrm{qos}_{i,j}^{ST} - \mathrm{qos}_{i,j}^{SR} \right)^2}$$
$$s.t. \quad w_j \geq 0, j = 1, 2, \ldots k, \sum_{j=1}^{k} w_j^2 = 1 \tag{30.9}$$

where w_j is the weight of j-dimension QoS when the distance is calculated. Because of the difference that single dimension QoS capacity resources provide, in order to measure the distance better, the deviation maximizing method is used to decide the value of w_j. The greater the capacity gap provided by jth dimension QoS on resource, the greater the influence of the parameter when the distance is measured, i.e., the greater the value of w_j. Otherwise, the smaller the value of w_j. For the QoS of jth dimension, $D_{i,j}(w)$ denotes the deviation of the QoS service capacity between resource r_i and other resource,

$$D_j = \sum_{i=1}^{m} D_{i,j}(w) = \sum_{i=1}^{m} \sum_{i'=1}^{m} \left| \mathrm{qos}_{i,j}^{SR} w_j - \mathrm{qos}_{i',j}^{SR} w_j \right|, \quad j = 1, 2, \ldots, k \tag{30.10}$$

where D_j denotes the total deviation of the QoS service capacity of jth dimension between each resource and other resources.

The following equation is used to calculate the value of the QoS integrated service capacity and the integrated QoS requirements after QoS parameters standardization,

$$U_{\mathrm{qos}}(w) = \sum_{j=1}^{k} \mathrm{qos}_{i,j} w_j \tag{30.11}$$

Obviously for each resource, the greater the value of $\mathrm{U_{qos}}(w)$, the better its integrated QoS service capability. So, the objective function is defined as

$$\max D(w) = \sum_{j=1}^{k} \sum_{i=1}^{m} \sum_{i'=1}^{m} \left| \mathrm{qos}_{i,j}^{SR} - \mathrm{qos}_{i',j}^{SR} \right| \times w_j$$
$$s.t. \quad w_j \geq 0, j = 1, 2, \ldots, k, \sum_{j=1}^{k} w_j^2 = 1 \tag{30.12}$$

The Lagrangian method is used to solve this objective function,

$$w_j = \frac{\sum_{i=1}^{m} \sum_{i'=1}^{m} \left| \mathrm{qos}_{i,j}^{SR} - \mathrm{qos}_{i',j}^{SR} \right|}{\sum_{j=1}^{k} \sum_{i=1}^{m} \sum_{i'=1}^{m} \left| \mathrm{qos}_{i,j}^{SR} - \mathrm{qos}_{i',j}^{SR} \right|}, \quad j = 1, 2, \ldots, k \tag{30.13}$$

The pseudocode of CTS_QoS is showed as in the following:

Input: Number of tasks, number of VMs, QoS requirements of tasks, QoS capacity of resourcesOutput: Solution of tasks scheduling strategy

Step 1. Standardize the QoS matrix of tasks requirements and resource capacity;
Step 2. Calculate the weight of each dimension QoS;
Step 3. Calculate the integrated QoS requirement value of each task and sort decreasingly all tasks according to QoS;
Step . For first task t_0 in T, calculate the QoS satisfaction on each resource and find the resource with maximal QoS satisfaction for forming set ListRe;
Step 5. Calculate the QoS distance between t_0 and each resource in ListRe, sort increasingly all QoS distance for forming ListDis;
Step 6. Schedule t_0 to the resource with the minimal QoS distance in ListDis;
Step 7. Delete t_0 in T. If T is not empty, return to Step 4; otherwise go to next step;
Step 8. If all tasks have been completed go to Step 10; otherwise, check whether there are idle resources;
Step 9. In the resource find in Step 8, check whether there are no tasks completed in ListRe and schedule it; otherwise go to Step 8);
Step 10. All tasks are scheduled and completed.

30.4 Experimental Results and Analysis

For evaluating the performance of our algorithm, some simulation experiments are conducted by our developed platform. The dimension of QoS constraint is 4 including the security QoS, the stability QoS, the cost QoS, and the success QoS, where only the cost QoS is the negative measure.

The parameters setting of tasks and resources are shown in Tables 30.1 and 30.2. Table 30.3 shows the scheduling results of our algorithm. We can see that only the VM2 and VM5 have tasks executed. This is because these two resources have the highest QoS satisfaction, which would be considered by all tasks.

MIN_MIN algorithm is selected as the compared algorithm, Fig. 30.2 shows the resource utilization and the QoS satisfaction. Obviously, the resource utilization is higher than CTS_QoS, because MIN_MIN prefers the minimal completion time which, as far as possible, allocation of the tasks to the fastest resources with the earliest availability for executing the task, not considering QoS constraints. While our algorithm CTS_QoS schedules all tasks to a small number of resources with higher QoS satisfaction, which would lead to a lower resource utilization. And, with the increment of the tasks number, these two metrics will increase because more tasks are scheduled.

Table 30.1 Parameter value of tasks

Task	MI	QoS1	QoS2	QoS3	QoS4
1	12406	100	36	6	89
2	1065	80	46	4	8
3	13498	24	29	6	28
4	873	10	76	8	19
5	18675	57	27	2	25
6	19756	92	42	6	6
7	6233	45	98	5	9
8	15267	82	6	4	47

Table 30.2 Resource setting

VM	MIPS	QoS1	QoS2	QoS3	QoS4
1	35	0.6	0.8	7	0.5
2	25	1	0.8	5	0.9
3	20	0.8	0.9	4	0.8
4	20	0.9	0.8	4	0.8
5	45	0.2	0.7	9	0.4

Table 30.3 Results of tasks scheduling

Task	VM	Scheduling sequence	QoS requirements	QoS satisfaction
1	2	4	0.44	0.261
2	5	2	0.71	0.282
3	2	6	0.425	0.276
4	2	7	0.144	0.289
5	5	0	0.988	0.145
6	2	5	0.431	0.295
7	2	3	0.567	0.283
8	5	1	0.713	0.29

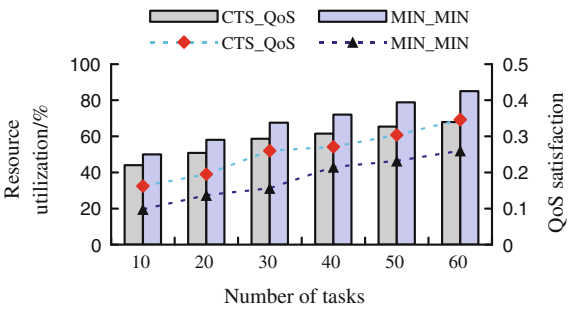


Fig. 30.2 Resource utilization and QoS satisfaction

30.5 Conclusions

A tasks scheduling algorithm based on multi-QoS constraints is presented. The algorithm considers synchronously the QoS satisfaction and the minimal QoS distance. The objective function reflecting the QoS service capacity is constructed and the Lagrangian method is used to solve the maximization function for obtaining the solution of the tasks scheduling. Experimental results show that our algorithm not only meets the maximal QoS satisfaction and improves the scheduling efficiency in multi-QoS constraints, but also ensures the resource utilization.

Acknowledgments The work was supported by Students' Scientific Research Project of WHPU (xsky2015033) and Innovation Training Project of WHPU (CXXL201510024).

References

1. William V, James B, Rajkumar B (2011) Introduction to cloud computing, cloud computing: principles and paradigms. Wiley Press, New York, pp 1–44
2. Anton B, Jemal A, Rajkumar B (2012) Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Future Gener Comput Syst 18(5):755–768
3. Jiangguang D, Yuelong Z, Huaqiang Y (2013) A service revenue-oriented task scheduling model of cloud computing. J Inf Comput Sci 10(10):3153–3161
4. Li B, Song M, Song J (2012) A distributed QoS-constraint task scheduling scheme in cloud computing environment: model and algorithm. Adv Inf Sci Serv Sci 4(5):283–291
5. Yang Z, Wang Q, Lv H (2014) Research on resource scheduling algorithm of cloud computing based on improved DAG diagram and task delay. Comput Meas Contr 22(2):499–502
6. Shen K, Hu D (2012) Research on task schedule based on cloud computing and improved discrete particle swarm. Comput Meas Contr 20(11):3070–3072
7. Feng L, Zhang T, Jia Z et al (2013) Task schedule algorithm based on improved particle swarm under cloud computing environment. Comput Eng 39(5):183–186
8. Wang W, Zeng G, Tang D et al (2012) Cloud-DLS: dynamic trusted scheduling for cloud computing. Expert Syst Appl 39(3):2321–2329

Chapter 31
A Method of Flange Size Measurement Based on Laser Displacement Sensor

Wen Huang, Xiaoqing Cheng, Zongyi Xing, Yong Zhang and Yong Qin

Abstract Aiming at solving the problem of measuring urban rail train flange wear, an online non-contact measuring method of flange size based on 2D laser displacement sensor is proposed to timely monitor flange size. First, the definition of the flange height and width for rail vehicle wheels is introduced. Then, a measurement technique of flange size is proposed to describe the tread profile line, including tread data segment, data preprocessing, and coordinate transformation data fusion. The flange size can be calculated by tread geometric relationship. Finally, the field experiments are conducted and the obtained results show that the detection accuracies of flange width and height are ±0.2 mm. The stability and reliability of the designed system can meet the actual measurement requirements.

Keywords Urban rail train · Flange wear · Non-contact measurement

31.1 Introduction

The wheels not only carry the full weight of the train, but also sustain the action force when the train is running. So it is a key part to ensure the safe operation of the train [1]. In the course of the train running, wheel wear occurs with the increase of operating mileage. The wear will cause a change in the relationship between wheel and rail, and reduce vehicle dynamic performance and the ride comfort [2]. Therefore, it is necessary to measure the parameters of the wheel to obtain the value of flange height and flange width regularly for making the repair decision whether the flange needs to be repaired or replaced [3].

W. Huang · Z. Xing (✉) · Y. Zhang
School of Automation, Nanjing University for Science and Technology,
No. 200 Xiaolingwei, Xuanwu, Nanjing, Jiangsu, People's Republic of China
e-mail: xingzongyi@163.com

X. Cheng · Y. Qin
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiao Tong University, Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_31

In the course of train operation, different wheels will have different flange wears, and the service time of different wheels also differs. So the large changes of the flanges make the measurement of the flanges become a difficult problem in wheel size measurement. The fourth checker [4] and the multi-link magnetic climbing tool [5] are the two main contact measurements in China. The measuring principle is easy but labor-intensive, and there are no methods to eliminate human error. In other countries, some factories developed a portable flange parameter measuring instrument based on laser displacement sensor. This instrument can get the full tread contour to realize the flange parameter calculation using the sensor to scan the full tread. PSD or CCD laser method is the main non-contact method [6–9]. They both use image recognition to obtain tread contour. However, this method has the disadvantage of complex structure, subject to environmental impact and poor reliability.

This paper provides a new online non-contact measurement method based on 2D laser displacement sensor. Wheel tread contour can be achieved by coordinate transformation, data integration, and other algorithm processing. The result of the field wheel test shows that the test error of the wheel size is ±0.1 mm, and the result of train test shows that the test error of the wheel size is ±0.2 mm. The consistency of the test results is higher than that of manual measurement. So this method can meet the size accuracy requirements of wheel repair.

31.2 Wheels Contour and Detection Scheme

31.2.1 Wheels Contour and Flange Size

The outline of wheel contour is shown in Fig. 31.1. The contact portion of the wheel rolling on the rail is tread. There is a raised portion along the circumference inside the wheel tread called flange. The base point is a point on the tread 70 mm away from inside baseline of the flange. The height between base point and flange is flange height; the width of flange which is 12 mm away from the base point is called flange width [10].

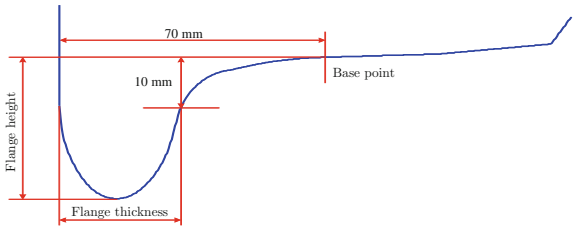


Fig. 31.1 Wheel contour diagram

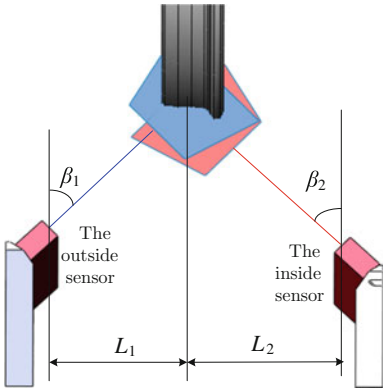


Fig. 31.2 Sensor installation diagram

31.2.2 Detection Scheme

There are two sets of 2D laser displacement sensors with certain geometric relationships mounted on both sides of the rail. The installation diagram is shown in Fig. 31.2.

31.3 Flange Size Detection Algorithm

Flange size measurement and data processing are shown in Fig. 31.3.

31.3.1 Data Segmentation

The laser sensor used in this paper has a specific distance interval of effective measurement range. The sensors will not output data if the distance is too far or too close. Therefore, this characteristic can be the principle to make data segmentation for raw data. Data output caused by other foreign matter is apparent inconsistency with wheels' data, so these data can be removed in data segmentation to realize the data preprocessing.

31.3.2 Coordinate Transformation

A part of the contour curves measured directly will exert distortion because the laser displacement sensor installation position will have a certain angle between the horizontal plane and rail direction. Therefore, it is necessary to correct deformed

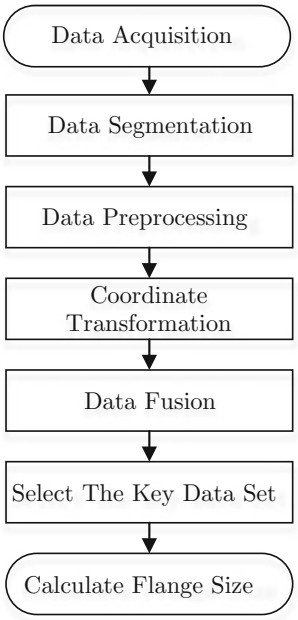


Fig. 31.3 Algorithm flowchart

curve by doing coordinate transformation of sensor output valid data. In this paper, the sensors' coordinates are called physical coordinates, coordinates after coordinate transformation are called intermediate coordinates, and coordinates after data fusion are called fusion coordinates.

The physical coordinates of the outside laser sensor 1 and the inside laser sensor 2 are $x^{(1)}o^{(1)}y^{(1)}$ and $x^{(2)}o^{(2)}y^{(2)}$. The relative positional relationship is shown in Fig. 31.4.

The intermediate coordinates of the outside laser sensor 1 and the inside laser sensor 2 are $u^{(1)}o^{(1)}v^{(1)}$ and $u^{(2)}o^{(2)}v^{(2)}$. So the coordinate transformation of the outside laser sensor 1 is shown in Fig. 31.5, and the coordinate transformation formulas are as follows:

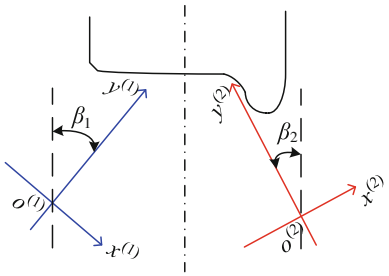


Fig. 31.4 Sensors' physical coordinate diagram

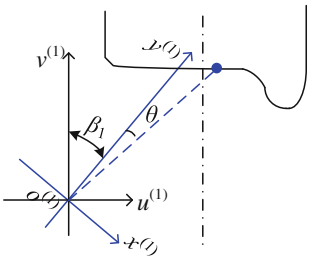


Fig. 31.5 Data coordinate transformation diagram of outside laser sensor

$$\begin{aligned} u_n^{(1)} &= \sqrt{x_n^{(1)^2} + y_n^{(1)^2}} \sin(\theta + \beta_1), \\ &= x_n^{(1)} \cos\beta_1 - y_n^{(1)} \sin\beta_1 \end{aligned} \tag{31.1}$$

$$\begin{aligned} v_n^{(1)} &= \sqrt{x_n^{(1)^2} + y_n^{(1)^2}} \cos(\theta + \beta_1), \\ &= y_n^{(1)} \cos\beta_1 - x_n^{(1)} \sin\beta_1 \end{aligned} \tag{31.2}$$

In the above formulas, $(x_n^{(1)}, y_n^{(1)})$ is an arbitrary point of physical coordinates $x^{(1)}o^{(1)}y^{(1)}$, θ is the angle between this point and $y^{(1)}$ axis, and $(u_n^{(1)}, v_n^{(1)})$ is the coordinate value in the intermediate coordinates $u^{(1)}o^{(1)}v^{(1)}$ of this point.

The coordinate transformation of the inside laser sensor 2 is shown in Fig. 31.6, and the coordinate transformation formulas are

$$\begin{aligned} u_n^{(2)} &= \sqrt{x_n^{(2)^2} + y_n^{(2)^2}} \sin(\theta' - \beta_2), \\ &= x_n^{(2)} \cos\beta_2 - y_n^{(2)} \sin\beta_2 \end{aligned} \tag{31.3}$$

$$\begin{aligned} v_n^{(2)} &= \sqrt{x_n^{(2)^2} + y_n^{(2)^2}} \cos(\theta' - \beta_2), \\ &= y_n^{(2)} \cos\beta_2 + x_n^{(2)} \sin\beta_2 \end{aligned} \tag{31.4}$$

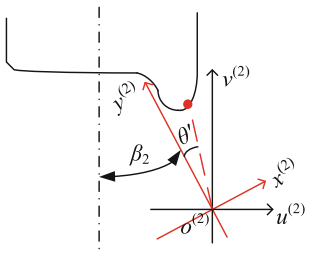


Fig. 31.6 Data coordinate transformation diagram of inside laser sensor

In these formulas, $(x_n^{(2)}, y_n^{(2)})$ is an arbitrary point of physical coordinates $x^{(2)}o^{(2)}y^{(2)}$, θ' is the angle between this point and $y^{(2)}$ axis, and $(u_n^{(2)}, v_n^{(2)})$ is the coordinate value in the intermediate coordinates $u^{(2)}o^{(2)}v^{(2)}$ of this point.

31.3.3 Data Fusion

As the single laser sensor can only get part of the wheel contours, it is necessary to make data fusion of the two sensors to get integral contours data. The relationship of fusion coordinates between $u^{(1)}o^{(1)}v^{(1)}$ and $u^{(2)}o^{(2)}v^{(2)}$ is shown in Fig. 31.7, and the data fusion formula is

$$\begin{aligned} u^{(0)} &= u_n^{(1)} + a \quad u^{(0)} = u_n^{(2)} + c \\ v^{(0)} &= v_n^{(1)} + b \quad v^{(0)} = v_n^{(2)} + d \end{aligned} \tag{31.5}$$

In this formula, a and b are respectively the horizontal and vertical coordinates of point $o^{(1)}$ in the fusion coordinates $u^{(0)}o^{(0)}v^{(0)}$, c and d are respectively the horizontal and vertical coordinates of point $o^{(2)}$ in the fusion coordinates $u^{(0)}o^{(0)}v^{(0)}$.

31.3.4 The Key Data Set Selection

When the wheel passes through the effective detection area of sensors, the laser sensors will output ranging number of multiple sets of the data. The measurement result of flange height is h when the laser line passes through the wheel center, and the measurement results are h' in the rest time. There are some deviations between h and h'. So the minimum flange height measured in different times is the key data set, and the final value of the wheel flange size should be measured by this data set.

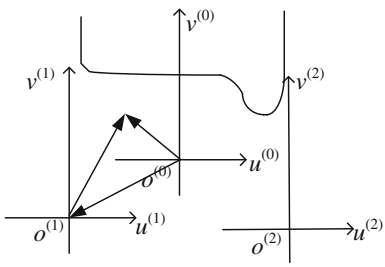


Fig. 31.7 Data fusion diagram of two laser sensors

31.4 Test and Analysis

In order to verify the validity and accuracy of the flange size detection proposed in this paper, the system is installed in the storage line of Guangzhou metro vehicle depot. Field test is carried out by a standard wheel-set test and a train with a speed of around 3 km/h passing through the detecting system.

In the wheel-set test, a standard wheel-set is placed on the rail and pushed manually to let the wheel-set pass through the detection system with a constant speed. First, the size of the flange is measured manually, and the results of flange high and flange width are 28.2 and 32.1 mm. This test is carried out for six times; therefore, the repeatability of the system can be obtained. Comparing with the artificial measurement results, the results of this system are shown in Table 31.1. From Table 31.1, the system measurement error is within ±0.1 mm.

Raw data collected by laser displacement sensors in the first time are shown in Fig. 31.8, and wheel contour in fusion coordinates after algorithm processing is shown in Fig. 31.9.

Vehicle experiment is to let a train repeatedly passing through the detecting system for 10 times. Then we can obtain the average value of every measurement results, and maximum or minimum results as well. In every test, artificial measurement can be obtained by skilled worker. The average values of every

Table 31.1 Test results of wheels/mm

Measurement times	1	2	3	4	5	6
Flange width	32.15	32.07	32.18	32.05	32.17	32.01
System error	0.05	−0.03	0.08	−0.05	0.07	−0.09
Flange height	28.26	28.27	28.27	28.18	28.27	28.15
System error	0.06	0.07	0.07	−0.02	0.07	−0.05

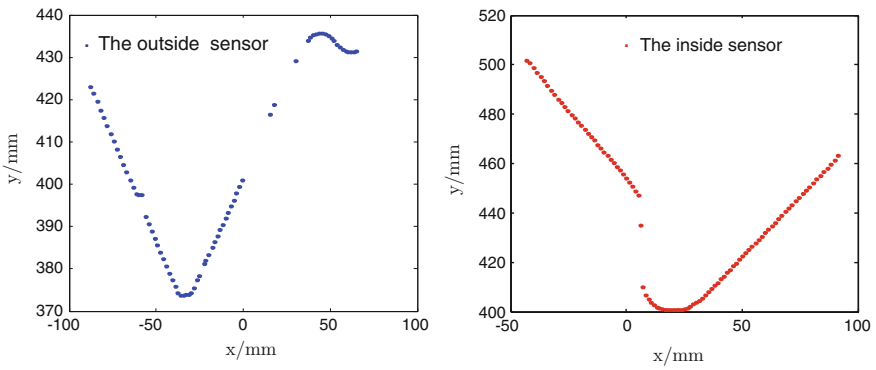


Fig. 31.8 Raw data collected by sensors

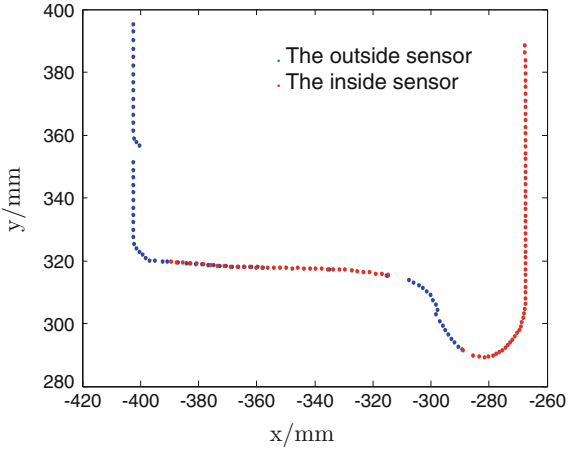


Fig. 31.9 Wheel contours in fusion coordinates

measurement results are shown in Fig. 31.10, and the deviation measurement values between manually measuring and systematic detecting are both within the range of 0.2 mm; the results show that the detection system has a high accuracy. The statistical results of maximum and minimum value of the wheel-set dimensions in 10 times are shown in Fig. 31.11.

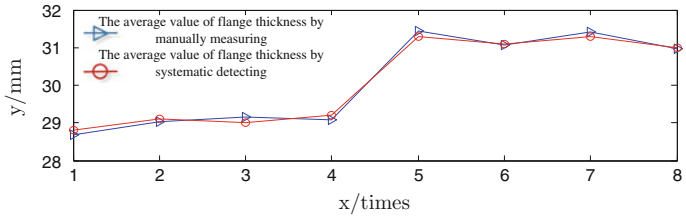


Fig. 31.10 Average value of flange width detected by system and people

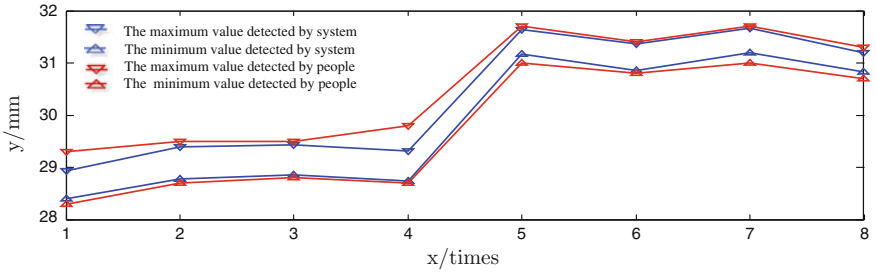


Fig. 31.11 Maximum and minimum results of flange width detected by system and people

31.5 Conclusions

This paper proposes a flange size measurement method based on laser displacement sensor to realize the online non-contact measurement of wheel-set dimensions. Through field test including wheel-set test and vehicle test, the measuring flange error is within ±0.2 mm, respectively. The consistency of the detection system is greater than the manual measurement, which means the reliability and stability of the system are both better. This system has established a foundation to grasp the flange size timely.

Acknowledgments This study was supported by the Key Technologies Research and Development Program of China (2011BAG01B05), the Fundamental Research Funds for the Central Universities (AE89454), and the Science and Technology Program of Guangzhou (20150810010). The author gratefully acknowledges the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially.

References

1. Hongbo Z, Haijian Y (2004) Researching of inspecting method for wheelsets abrasion based on image processing. J Mach 31(8):51–53 (in Chinese)
2. Yu F, Huadong M, Qi Z (2010) Analysis of abnormal wear of wheel tread in Shanghai metro line 3. J Electr Drive Locomot 2:45–46 (in Chinese)
3. Kaihua W, Jianhua Z, Zuohua H (2004) Using optoelectronic image detecting method to measure the wear of wheel set flange. J Opt Tech 30(5):535–537 (in Chinese)
4. Deli F, Xiaozheng W (1998) Development of the fourth checker LLJ-4. J Rolling Stock 36(2):42–44 (in Chinese)
5. Zhou W, Li D, Jiang X (2006) Study on the digital profilometer of railway wheels. J China Railway Soc 27(5):49–53 (in Chinese)
6. Jing C, Jianwei Y, Zhizhong T (2014) On dynamic testing system of metro vehicle wheelset. J Urban Mass Transit 7:82–84 (in Chinese)
7. Zhou W, Wu K, Cai X (2014) Research for multi-channel wheelset image acquisition trigger system. J Mech Electr Eng 31(6):800–804 (in Chinese)
8. Yongcheng Z (2011) Wheel device for dynamic testing. Southwest Jiaotong University, Chengdu (in Chinese)
9. Liyan W (2009) Metro wheel size online monitoring system, vol 2, pp 23–24 (in Chinese)
10. Yong Z, Zongde F, Kanwei W (2008) Field calibration method for wheelset profile non-contact measurement 35(12):135–137 (in Chinese)

Chapter 32
High-Speed Train System Reliability Modeling and Analysis

Jinchu Zheng, Yong Qin, Xiaoqing Cheng, Shuai Lin, Zhilong Zhang, Hengkui Li and Xiaojun Deng

Abstract At present, the international high-speed rail on clean and green and intelligent technology received unprecedented attention, especially the integration of security technology has become a core technology development trends. First, this paper does reliability analysis from the high-speed train system. Statistics of the subsystems of the failure rate is done. Then important subsystems bogie system reliability is analyzed combining with the maintenance on the high-speed train system reliability analysis. The parts distribution functions and associated parameters are fitted on software. Finally, we set up truck event tree model and reliability analysis of bogie system.

Keywords High-speed train · System · Reliability analyze

32.1 Introduction

High-speed rail with high speed, low pollution, fast, comfortable, secure, and on-time advantage has become one of the most sustainable mode of transportation. Reliability of high-speed train system refers to the system under the conditions laid down, within the time stated in the ability to perform specified functions [1, 2].

Y. Qin (✉) · X. Cheng
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
e-mail: qinyong2146@126.com

Y. Qin · X. Cheng
Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing 100044, China

J. Zheng · S. Lin
School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Z. Zhang · H. Li · X. Deng
CSR Qingdao Sifang Co., Ltd., No.88 Jinhongdong Road, Chengyang District, Qingdao 266111, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_32

The high reliability of the system shows that the likelihood of failure is smaller, perform functions more likely. China has become the world's largest high-speed rail network countries [3, 4]. China's high-speed rail faces from large-scale construction as the key to the strategic transition with a focus on safe and reliable operations [5, 6]. High-speed trains as a complex high-tech system, mechatronics, there liability study of genealogy work is necessary. Reliability focuses on support systems function well, achieving system objectives [7, 8].

32.2 Reliability Analysis of Vehicle System

Existing data mainly has the following problems: (1) recorded by hand, most of the descriptive language, and the large amount of data; (2) there will be some error in the recording process, despite staff of audit, but there are still some bad data exists; (3) properties of the data contain more, not all data can be used for analysis. As part of the data to a blank data, part of the data for this study does not make sense. Therefore, we first need to preprocess the data, excluding an entire column blank, data, and contents useless data; Modify the error data; Complete the missing data. Statistical analysis based on pretreatment data is the following [9].

The data show that CRH2 high-speed train failure consequences of the system continues to run, the temporary repair, broken, and outages, delays, no libraries, no influence, delay. Statistical analysis of fault type is shown in Table 32.1:

Fifteen subsystems of CRH2 high-speed train subsystems within 2 years of failure statistics is shown in Table 32.2.

As can be seen from Table 32.2, failure results in 2 years, air supply and the failure of the braking system with the highest frequency which followed by the water supply and sanitation systems, air conditioning and heating systems and traction drive system. In addition, it is evident as the running miles increase and high-speed train failure of each subsystem is also increasing.

Table 32.1 Fault statistics

Failure consequence	Number of occurrences
Continue to run	47
Temporary repair	102
Broken	39
Outage	26
Delayed	12
No libraries	1
No influence	729
Delayed	2

Table 32.2 Subsystem faults

Subsystem	Failure number	Kilometers	Failure rate(10^{-8})
Car-end connection system	24	38,724,954	0.0849
Body and body accessories	37	48,702,824	0.1041
Electrical and power strips	52	65,530,599	0.1087
Auxiliary power system	48	61,080,856	0.1076
High voltage power system	63	76,334,720	0.1131
Water and sanitation system	112	140,978,754	0.1088
Air supply and brake system	137	196,569,183	0.0955
Traffic safety equipment	18	23,279,418	0.1059
Air heating system	107	144,119,170	0.1017
Door and window system	77	88,189,913	0.1196
Built-in system	42	48,131,533	0.1195
Traction drive system	109	155,152,528	0.0962
Network and control system	19	26,361,314	0.0987
Television broadcast system	28	27,469,060	0.1396
Bogie suspension system	85	122,168,512	0.0953

Table 32.3 Failure distribution function of subsystems for high-speed train table

System	Distribution types	Parameters
Bogie	3-parameter Weibull	η: 1.268E + 04, β: 1.382, γ: 1.838E + 04
Electrical	3-parameter Weibull	η: 1.2321E + 04, β: 1.516, γ: 1.758E + 04
Brake	3-parameter Weibull	η: 1.551E + 04, β: 1.241, γ: 1.805E + 04
Ventilation cooling	2-parameter Weibull	η: 4.36E + 04, β: 4.483, γ: 0
Exterior	3-parameter Weibull	η: 1.082E + 04, β: 1.441, γ: 1.784E + 04
Built-in	3-parameter Weibull	η: 8132, β: 1.523, γ: 1.767E + 04
Body	3-parameter Weibull	η: 1.627E + 04, β: 1.315, γ: 1.732E + 04
Water sanitation	3-parameter Weibull	η: 1.05E + 04, β: 1.422, γ : 1.807E + 04

According to the functional characteristics of the high-speed train system, it generally can be divided into eight subsystems. These are the bogie system, electrical system, braking system, ventilation and cooling systems, vehicle systems, interior systems, body systems, water supply, and sanitation systems. Isograph software and analysis of failure distribution function of subsystems for high-speed train in sequence as shown in Table 32.3.

Table 32.3 shows that the bogie system follows the 3-parameter Weibull distribution. Its value is 0.9563. Variance is 0.0773. The Fig. 32.1 shows that early failure rate curves rose steadily, the medium-term accelerated, rising slowly in the late shows medium-term failures more likely. The continuous linear representations and downtime related to the failure rate, the distribution rate. This line failure is rate expectations in different time interval, namely regional rates.

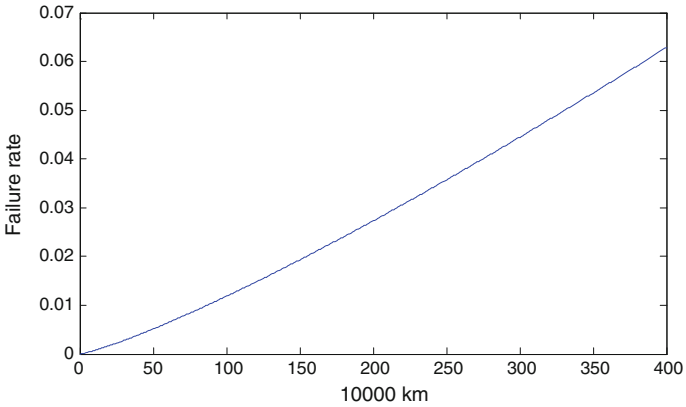


Fig. 32.1 Bogie system failure rate curves (color figure online)

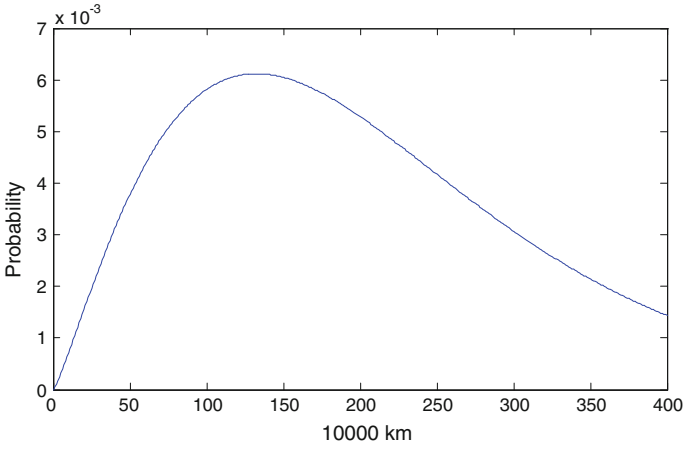


Fig. 32.2 Bogie cumulative probability distribution function (color figure online)

As shown in Fig. 32.2 and plot as a raw data, the line represents the cumulative probability curve fitting based on the original data. The line is the cumulative probability distribution of the bogie system. In order to facilitate the subsequent calculations, the cumulative probability distribution of the time takes the logarithm of this dot curve and this line. Probability density function corresponds to the highest quarters of time, indicates that bogie system failure probability is high.

32.3 Bogie System Reliability Analysis

Bogie is one of the most important components, high-speed running plays host buffer, traction, braking, steering, and so on, it directly determines the parameters of the high-speed vehicle stability and ride comfort [10]. Bogie structure and directly impact on the reliability design of the vehicles ride quality is dynamic performance and safety. Thus bogies on high-speed train system reliability and safety of the important status. Thinking bogie system here as an example, statistical processing and analysis of data is detailed.

High-speed train bogies can be divided into power and non-power steering bogie frame. Power steering rack is composed of six parts, each part is as follows:

Wheel as the vehicle and track system interfaces directly to the rail pass gravity, through the adhesion between wheel and rail traction or braking and wheel rotation of vehicles running on rails (translation); Axle box is the joints connect the frame assembly and wheel, in addition to ensure the wheel rotary motion, but can adapt wheel line is not smooth and other conditions, relative to the vertical, horizontal and vertical movement of the frame assembly.

Central suspension mainly formed by springs and dampers. Modern high-speed train vehicles use two suspension, a suspension at axle box and frame assembly, suspension frame assembly of the secondary suspension between secondary suspension is also known as the central suspension. The spring suspension device, used to balance the axle load distribution to ease line impact of the irregularity on the vehicle, ensure vehicles and smoothness, and ensure the vehicle through the rotary flexible curved bogie in relation to the car body.

Frame assembly based skeleton-bogie to install all components and parts of bogies, and accept and pass a variety of loads.

Basic brake the brake cylinder pressure increased a number of times to brake pads or brake shoes, which pressed brake disk (or wheel), for purposes of braking of vehicles.

Torque of the driving device will power effectively passed to the wheel, driving wheels.

Attachment device with the bogie on each device complete its function.

From existing data analysis, failure-prone components of high-speed train bogie system had 16, namely, brake caliper, brake pads, wheel-mounted disk brake, shaft, coupling, gear boxes, axles, traction motor box assembly. Statistics on the bogie system component failure based on Isograph software, respectively, received the parts distribution function is shown in Table 32.4.

Brake caliper failure follows the 3-parameter Weibull distribution. Mean value is 0.959. Variance is 0.08042. The failure rate function (as shown in Fig. 32.3) indicates that the brake caliper parts running about 1.5 year failure rate significantly increased. Describe the equipment started aging and progressive failure rate leveled

Table 32.4 Bogie system components distribution function

Parts	Distribution types	Parameters
Brake caliper	3-parameter Weibull	η: 1.008$E+04$, β: 1.305, γ: 1.85E + 04
Brake pads	2-parameter Weibull	η: 1.416E + 04, β: 2.666, γ: 0
Wheel-mounted disk brake	2-parameter Weibull	η: 5.68E + 04, β: 1.529, γ: 0
Axle box body	3-parameter Weibull	η: 1.674E + 04, β: 0.8512, γ: 1.959E + 04
Axle	3-parameter Weibull	η: 1.371E + 04, β: 1.397, γ: 1.863E + 04
Traction motors	3-parameter Weibull	η: 1.862E + 04, β: 0.9961, γ: 1.844E + 04
Coupling	2-parameter Weibull	η: 4.417E + 04, β: 3.221, γ: 0
Gear box assembly	3-parameter Weibull	η: 2.243E + 04, β: 1.132, γ: 1.742E + 04
Height adjustment device	2-parameter Weibull	η: 4.293E + 04, β: 3.566, γ: 0
Shock absorbers	2-parameter Weibull	η: 4.7E + 04, β: 1.167, γ: 0
Air spring	3-parameter Weibull	η: 5.429E + 04, β: 4.987, γ: −2366
Drawbar	Normal distribution	μ: 4.678E + 04, σ: 1.96E + 04
Transverse damping	Normal distribution	μ: 4.818E + 04, σ: 9011
Speed sensor	3-parameter Weibull	η: 3602, β: 0.5067, γ: 1.999E + 04
Tread cleaning device	3-parameter Weibull	η: 5064, β: 0.4869, γ: 2.109E + 04
Pressurized cylinder	3-parameter Weibull	η: 1.078E + 04, β: 1.3, γ: 1.86E + 04

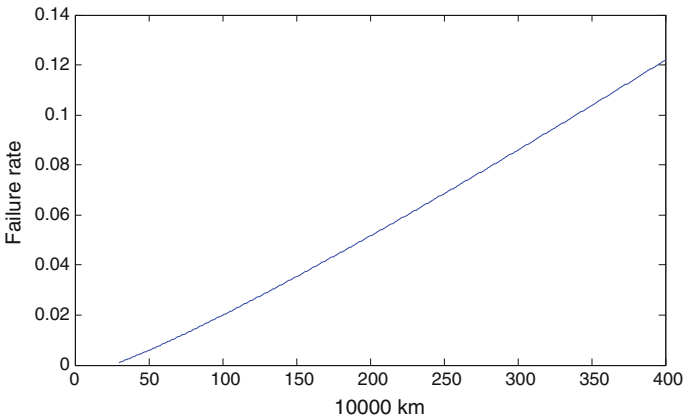


Fig. 32.3 Brake caliper cumulative probability distribution function

off after 3 years, running more stable. Fault density between 1.5 and 2.5 years, cumulative probability chart (as shown in Fig. 32.4), most of that fault in fault-intensive areas, a small number of fault distributions in densely populated areas outside.

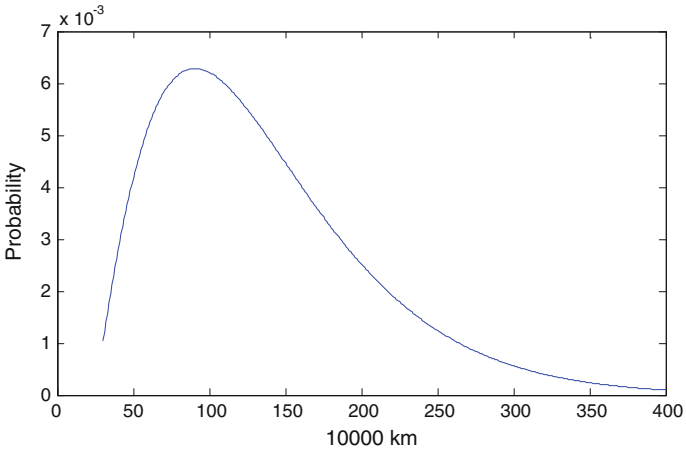


Fig. 32.4 Brake caliper failure distribution function

32.4 High-Speed Train Maintenance Analysis

According to the different models of high-speed train components failure probability of occurrence and different ways of working, different structure of maintenance period, China high-speed train system of repairs, high-speed train maintenance cycle is shown in Table 32.5.

Failure on the auxiliary power supply system for 1.2 million kilometers is four-level maintenance at about 900,000 km, meet the maintenance period. Television message broadcast system failure rate is high, running about 980,000 km

Table 32.5 Maintenance period

Maintenance level	Maintenance period		
	CRH1	CRH2	CRH5
Maintenance, routine inspection	0.25 month	End of each run, or 48 h	End of each run
Two-level maintenance, the focus of inspection	0.5 month	30,000 km or 1 month	600,000 km
Three-level maintenance, focusing on decomposition of checks	600,000 km	450,000 km or 1 year	1.2 million kilometer
Four-level maintenance, system check	1.2 million kilometer	900,000 kilometer or 3 years	2.4 million kilometer
Five-level maintenance, vehicle breakdown check	2.4 million kilometer	1.8 million kilometer or 6 years	4.8 million kilometer

when one fault appears, the adoption of reasonable maintenance plan or improving system design to avoid or reduce the incidence of failure. Other subsystem failure is basically similar to the auxiliary power supply system.

According to maintenance regulations of CRH2 high-speed train levels two stages, high-speed train wheel axle box, traction motors, air springs, axle box springs, hydraulic dampers, speed sensors, brake calipers, and other key components through levels after repair remain in a good state, guarantee the safe operation of 600,000 km; remaining after repair of parts under guarantee to a secondary inspection. Status update component should perform its warranty period. Statistical results of round three failure modes cracks, scratches, peeling failure rate in line with wheel consisting of three levels of maintenance procedures. Failure mode analysis of bogie parts of six parts meets the high-speed train grade maintenance regulations focus of the inspection.

32.5 Bogie System Events Tree

Event tree analysis is a method of logical deduction, which in the case of a given event in early, early analysis of the result of the sequence of events can lead to a variety of events and characteristics of qualitative and quantitative evaluation system, helping analysts get the right decision. It is frequently used to analyze accidents and reliability of the system.

Early bogie system event tree is to determine possible consequences of events, it can be used to describe truck sequence of events might occur in the system. Bogie system events to be able to quantify probabilities for each branch of a chain of events and to further assess the consequences of these events risk and reliability analysis of bogie system.

According to the previous data analysis, combined with access to the bogie system fault information using Isograph software, a gearbox is injured early for events, event tree model building bogie system, as shown in Fig. 32.5.

If case of gearbox for high-speed train attacked by a foreign object, the train depot is the probability of stopping 0.1568; trains stopped operating in situ maintenance of probability is 0.09433; the probability of running trains to slow down is 0.1761; local adjustments to maintain operation of probability is 0.2134–0.1653 running trains. Thus, foreign body injury train gear box needs to be adjusted in the context of the most likely, the normal operation of the probability is less than 20 %, so the bogies in the course of system maintenance, pay attention to check that the gearbox is injured by a foreign object.

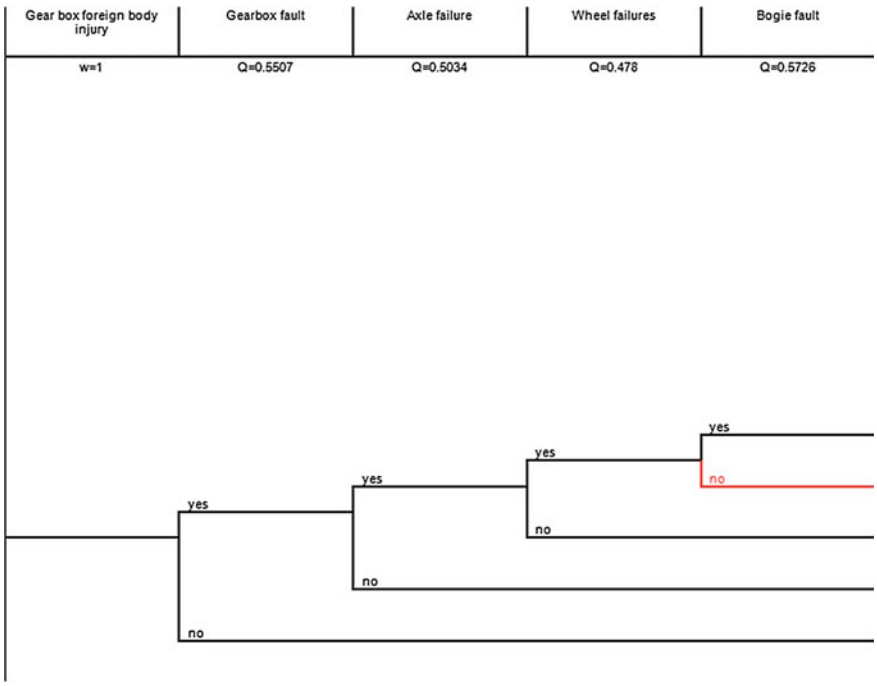


Fig. 32.5 Bogie system event tree model

32.6 Conclusion

In the high-speed train level and second level repair, bogie system is the focus of the overhaul, thus bogie system on the importance of high-speed train system, so we select the bogie system to analyze. In the bogie frame assembly of the high failure rate in the system, an average of 260,000 km one breakdown can occur, so the focus of the second-stage maintenance checks, two-level maintenance mainly by manual visual check bogies, drive the problems of maintenance in a timely manner. Based on the reliability and failure rate functions of time, reliability of the brake caliper in about half a year, 0.96 %, in a year or so of reliability at 0.925, three-level check is required. Wheel parts with the highest failure rate when probability 0.0132 failure occurs, in line with the three levels of maintenance for wheel set inspection and repair priorities. Bogie system reliability is about 0.93 in about 4 years which basically conform to the four-level maintenance for the decomposition of the system check.

Acknowledgment This research is sponsored by National 863 project (No. T13B200011), the Excellent Doctor Fund under Grant 2014YGS128, the Research Fund for the Doctoral Program of

Higher Education of China under Grand 20120009110035, the State Key Laboratory Program under Grant RCS2014ZT24 and Theory of mass transit train system reliability and safety assessment (No. I14K00451).

References

1. Seo SI, Park CS, Choi SH et al (2010) Reliability management and assessment for the electric traction system on the Korea High-Speed Train. Proc Inst Mech Eng Part F J Rail Rapid Transit 224(3):179–188
2. Zhang WJ, Lan N (2013) Research on the reliability growth management techniques of high-speed train for whole life cycle. In: The 19th international conference on industrial engineering and engineering management. Springer, Berlin, pp 529–539
3. Tan Q, Qi H, Li J et al (2012) The reliability modeling and analysis on brake system of medium-low speed maglev train. In: International conference on computer distributed control and intelligent environmental monitoring (CDCIEM). IEEE, pp 772–777
4. Flammini F, Mazzocca N, Iacono M et al Using Repairable Fault Trees for the evaluation of design choices for critical repairable systems. In: IEEE ninth international symposium on high-assurance systems engineering (HASE 2005). IEEE, pp 163–172
5. Zhang WJ, Lan N (2013) Research on the reliability growth management techniques of high-speed train for whole life cycle. In: The 19th international conference on industrial engineering and engineering management. Springer, Berlin, pp 529–539
6. Zhang B, Li X, Li J (2013) Research on emergency case ontology model based on ABC ontology. In: international conference on management science and engineering—annual conference, Taipei. IEEE Computer Society, pp 227–233
7. Li L, Wu W, Liu N (2013) Ontology model for situation awareness of city tunnel traffic[J]. Appl Mech Mater 47–350:715–717
8. Cui X, Yin G (2011) Web services discovery ontology model based on matching algorithm. Appl Res Comput 28(8):3010–3012
9. Zhang X, Tian T (2010) Logistics domain ontology model and its application. Adv Mater Res 108–111:1403–1408
10. Yu J, James A (2009) Requirements-oriented methodology for evaluating ontologies. Inf Syst 34:766–791

Chapter 33
Rail Train Door System Hidden Danger Identification Based on Extended Time and Probability Petri Net

Yunxiao Fu, Limin Jia, Yong Qin and Xiaoqing Cheng

Abstract This paper proposes a hidden danger inference approach to solve the rail train door fault predicting problem. Based on Extended Time and Probability Petri Net (ETPPN) the door action process model is built and the fault reasoning process is realized by TC-PPN which is the backwards inference of ETPPN. The time interval of every event occurrence and the history statistical probability are two information aspects for the search evidence of hidden fault. In terms of time interval, the hidden fault could be insulated with most other hidden faults, and the probability sequence could give the confidence level to search the relative hidden faults. And based on the train door system physical constitution, the hidden fault could be searched by TC-PPN step and the result satisfies the real maintenance strategy indeed. Overall, this algorithm based on ETPPN could give the accurate result for train door system hidden trouble identification, and is valuable to apply in train maintenance engineering.

Keywords Extended time and probability petri net · Discrete event system · Hidden fault · Train door system

Y. Fu · L. Jia (✉) · Y. Qin · X. Cheng
State Key Lab of Rail Traffic Control and Safety, Siyuan Building,
Beijing Jiaotong University, Haidian District, Beijing, China
e-mail: jialm@vip.sina.com

Y. Fu · L. Jia · Y. Qin · X. Cheng
Collaborative Innovation Center of Railway Traffic Safety,
Beijing Jiaotong University, Beijing 100044, China

Y. Fu · L. Jia · Y. Qin · X. Cheng
Beijing Research Center of Urban Traffic Information Sensing
and Service Technologies, Beijing Jiaotong University, Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_33

33.1 Introduction

The train door system which belongs to the key systems of train is complex mechatronic system with many spare parts. The multiple fault sources and high fault frequencies of door system have profound influence on the reliability and security of train operation. The research of active identification of hidden danger is beneficial for the reliability and security of train operation. The train door system is a complex system where considerable parts link in time domain and couple in space domain inseparably, and the operating faults occurrence is driven by the events information exchange. Diagnosis based on model of discrete event system (DES) is proposed by L. Sampath in 1995, which is used in chemical industry system [1, 2]. Fault diagnosis model oriented, problem definition oriented and the improvement and application of algorithm-oriented DES have been continuously brought followed and currently has formed developed directions owning multi-branches of research. fuzzy discrete event system (FDES) that is applied on the solution of uncertain problems under the certain conception is one of the significant research branches [3]. Furthermore, it can deduce the causal connection among the components inside the complicated nonlinear system as well [4].

Petri Net has good abilities of visible modeling and description of complicating, asynchronous, and discrete process. Many fields illustrate the applicability of Petri Net and many expand types of Petri Net are built [5, 6]. Petri Net is one of the mature approaches to model, reason, and analyze evolution of FDES. Document [7] designed a factory DES model to diagnose fault based on Petri Net, He [8, 9] and She [10] employ extensible time and fuzzy petri net (ETFPN) to locate and identify the fault in the electric power system. In the field of rail transit, Petri Net has been involved in security and risk analysis while rarely research on fault causal analysis of train pivotal systems used Petri Net. Document [11] impressed Petri Net on human factor engineering, which analyzes the relative problems with human factors in railway system. Document [12] applied automation Petri Net to analyze the security of interlock signal system. Above researches relate to train security field without analyzing fault effect relation aimed at real systems. Document [13] put forward a method to decouple and separate electric power system fault causal chains based on time-probability Petri Net model, which simplifies the fault diagnosis process of complex system and automatically analyze fault causal relationship. Similarly, this approach sufficiently matches to study train door system with frequently occurred and difficultly to diagnosed faults. However, this approach has no application case in rail transit field.

This paper employs Petri Net approach expanded with time series and probability information to analyze train door system fault. Because of this approach we study the train door system fault delivering and hidden trouble identification under operating condition. First of all, constructing the time sequence and probability Petri Net model of train door system aims at analyzing the fault causal chain of train door system object. Furthermore, the next step is analysis of time sequence relationship and probability information of backward inference in the same causal chain

model so that the process of searching fault source under the condition that the state set is incomplete and the interference exists could be realized. The result of experiment reveals that the Petri Net expanded time sequence and probability information could imitate the train door system fault causal relationship completely. Meanwhile, it could locate fault sources and infer hidden danger accurately. To conclude, it has better application popularized value.

33.2 Extended Time and Probability Petri Net Model

To define the time constraint probabilistic Petri net (TC-PPN), the graphic example is shown in Fig. 33.1.

Definition 1 A time probabilistic Petri nets (TPPN) structure is a eight-tuple here $\sum_{\mathrm{TP}} = (S, T, F, E, I, \delta, \tau, M_0)$, in which (S, T, F) is the basic Petri Net structure the basic feature of Petri Net is normal so that do not elaborate here. $E = \{e_l | l \in N^+\}$ is the finite set of fuzzy event. $I \subseteq (S \times T) \subset F$ is the finite set of inhibited arcs and $\forall i = (s,t) \in I: M_\sigma(S) > 0 \rightarrow M[\mathrm{t} \tilde{>} \, . \, M[t \tilde{>}$ denotes t could not be triggered under M state. The detailed description of $M_\sigma(s)$ is shown in Define 2. $\delta: E \rightarrow 2^{|T|}$ is the mapping of fuzzy events to fuzzy transitions. $\tau: T \rightarrow R_0 \times (R_0 \cup \{\infty\})$ exists, where $\tau(t_i)$ is the time interval associated with transition t_i. M_0 is the initial state marking.

In Fig. 33.1, the straight line with circle at the end part denotes inhibition of arc.

$M = [M(s_1), M(s_2), \ldots, M(s_n)]^T$ where $n = |S|$. Supposed $M = (M_\sigma, M_\upsilon)$, where M_σ is n dimensions vectors assemble by σ components of all $M(s_i)$ while M_v is n dimensions vectors assemble by υ components of all $M(s_i)$.

Definition 2 Suppose $\sum_{\mathrm{TP}}$ is a TC-PPN, fuzzy event e has a trigger matrix $B_e = [b_{ij}]_{m \times n}$ owning m rows and n columns, the element b_{ij} of which is defined as follows.

$$b_{ij} = \begin{cases} \pi\left(p(t_i^{\bullet} | {}^{\bullet}t_i)\right), & \text{if } t_i \in T_e \wedge (t_i, s_j) \in F \\ 0, & \text{else} \end{cases} \tag{33.1}$$

In which, $|S| = n$, $|T| = m$, $e \in E$, $i \in \{1, 2, \ldots, n\}$, $j \in \{1, 2, \ldots, m\}$. $T_e = \delta(e)$ is the subset of T, $P(t^{\bullet} | {}^{\bullet}t)$ is calculated by Eq. (33.3).

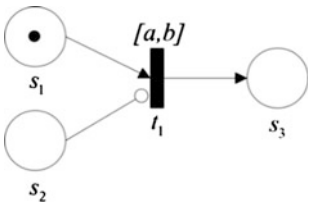


Fig. 33.1 Basic petri net construction

Definition 3 Suppose $\sum_{\mathrm{TP}}$ is a TC-PPN, the state transition regular could be defined as:

1. For $e \in E$, Te = $\delta(e)$ is the subset of T, if $(\exists \mathrm{t} \in \mathrm{Te})(\forall \mathrm{s} \in {}^{\bullet}t)$: $M\sigma(s) > 0$ then fuzzy event e is considered to possess happen right and trigger right, designated as $M[e>$.
2. If $M[e>$ exists and the new marking M' will be triggered by e, which could be described as $M[e> M'$, then M' could be obtained by followed equations.

$$\mathrm{M}'_{\sigma} = \mathrm{M}_{\sigma} \oplus M^{+}_{\sigma} \tag{33.2}$$

$$M^{+}_{\sigma} = M_{\sigma} \times (A^{-})^{T} \circ (A^{+} \cdot B_e \cdot \mathrm{BT}) \tag{33.3}$$

$$M'_{\upsilon}(s_i) = M_{\upsilon}(s_i) \cup \left(\bigcap_{s_j \in {}^{\bullet}t_e} \left(M(s_j) + \tau(t_e) \right) \right) \tag{33.4}$$

where, $t_e = \delta(e)$, A^{-} is the input matrix, A^{+} is the output matrix, B_e is the trigger matrix of fuzzy event e, and BT is the backward matrix of time interval. Operator $\cdot$ is the dot product operation of two matrixes, $\circ$ is the maximum product operation, $\odot$ is minimum subtraction operation, $\oplus$ is taking the maximum operation.

To make the analysis of concrete issue convenient, and reduce the difficulty of modeling, here an expended time and probability Petri net (ETPPN) is defined as Definition 4.

Definition 4 The (ETPPN) is an eight-tuple $\sum_{\mathrm{ET}} = \left(S', T', F', E', I', \delta', \tau', M'_0\right)$ where (S', T', F') is the original Petri net, satisfying $\forall s \in S$: $\{s\} \cap \overrightarrow{s^{\bullet}} = \emptyset$, $\forall t \in T'$: $|{}^{\bullet}t| \leq 1$. E' is a finite set of events. $I' \subset F'$ is the finite set of inhibited arcs. δ': $E' \rightarrow 2^{|T|}$ is the mapping from event to transition subset. Detailed ETPN-related reasoning description could refer to [10].

33.3 Train Door System Closing Process Analysis

The train door system is mechatronic system with high degree of complexity, which comprises electrical subsystem, mechanic subsystem and communicating subsystem. Electrical subsystem involves electronic door control unit (EDCU), power source, motor, and switch devices in basic electric circuits. Mechanic subsystem contains door and the surrounding mechanical parts, such as screw, door carried gantry, long, and short pillars. The communicating parts refers to communication equipment, one part of which is between Main EDCU (MDCU) and Local EDCU (LDCU) and the other part of which is between the driver chamber and the train door. The door closing process after EDCU receiving closing order consists of three steps.

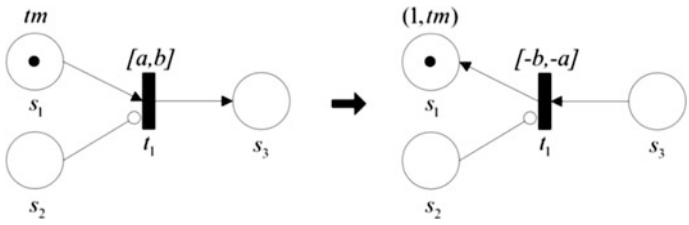


Fig. 33.2 Conversion of ETPPN to TC-PPN

The first step is EDCU controlling course. After EDCU receiving closing order and zero velocity signal, door locating information will deliver to EDCU from door location sensor. Program inside EDCU controls the yellow light alarm and the end of stroke switch S3 to unlock.

The second step is longitudinal move. The motor begins to drive the door device to move under the primary consideration that the motor has unlocked from the connecting mechanism of screw stud-Nuts. Under the drive of motor, the screw whirls, the screw nut drive drags the door carried gantry slipping on the long pillar. Simultaneously, the door frame rollers at the bottom of each door device trundle in the slide near the bottom of the door devices.

The third step is transverse swaying stage. In this period, the frame sways on the short pillars and drives the door carried gantry with the door device to sway together. Then stroke switch S4 in the safety interlock circuit will act that the normally open contacts close. EDCU will receive the action information of S4 and will finish the assignment that LS lock into each other. At the same moment, stroke switch S1 act, and under the precondition that the safety interlock circuit has been linked the door devices lock to each other. Meanwhile, at the bottom of the door device, the stopping pins return to insert places and the roller swing arms return to the original place. Part top mechanic structure of door device surrounding components is shown in Fig. 33.2. This part shown in Fig. 33.2 belongs to mechanism subsystem.

33.4 Experiment Analysis

33.4.1 Door Closing Process Petri Net Model

Based on the example of the door closing process, we reveal the simulate experiment procedure of hidden trouble excavation. Figure 33.3 shows the Petri Net model of the door closing process of rail transit train door system. Each mark in the figure has corresponding real action name shown in Table 33.1.

S1–S34 are the name marks of the component or the component action participating in the real door closing process. There are six virtual places which are

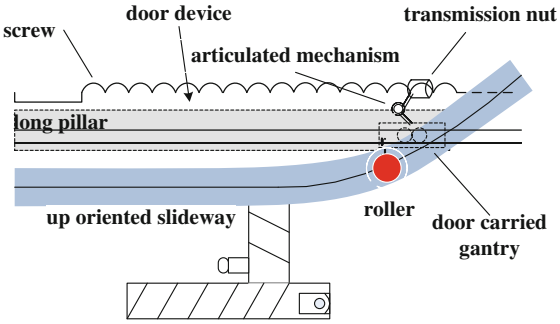


Fig. 33.3 Upper part of door device vertical view

expressed by gray place marks. Virtual places are middle places built up for modeling the door closing process and they have values for places state transiting without owning any real physic implication and representing any train door action

Table 33.1 Door closing process petri net place marks

Mark	Implication	Mark	Implication	Mark	Implication
S1	Pushing the button	S13	The end of S3 unlocking	S25	S4 open contact closed
S2	ATO switch at no automatic transmission	S14	Safety interlock circuit disconnecting	S26	LS interlocking
S3	Door closing signal	S15	S2 acting	S27	S1 acting
S4	EDCU device	S16	Motor deriving signal	S28	Safety interlock circuit connecting
S5	HMI in the driver cab	S17	The operation of motor	S29	The door device interlocking
S6	Automatic train operation (ATO) device	S18	The motor unlocking	S30	The stopping pins returning
S7	ATO switch at automatic transmission	S19	Longitudinal movement of the door device	S31	The roller returning to the original place
S8	VCU	S20	Longitudinal movement to the end position	S32	The door finishing closing process
S9	MVB normal	S21	Reducing the current	S33	The door location sensor
S10	Zero velocity signal	S22	Frame moving	S34	The door closed
S11	The largest open position of door devices	S23	Roller lipping in the slideway	Gray	Virtual place and middle place
S12	Yellow light flare	S24	S4 acting	–	–

information. The train door Petri net model which is constructed by the key components in door closing process and their correlated action built in Fig. 33.1 simplifies the practical action process, and ignores some stationary support non-critical components such as draught excluder strip of the door device.

33.4.2 Hidden Fault Identification in Door Closing Process

The fault mode set of door closing process mainly contains no door closing response condition, incomplete door closing condition, slow closing process, and noise increasing process. According to currently existing method such as expert system and empirical statistics analysis method the maintenance strategy could not get accurate reliable information to exhume fault source for the strategy is formulated after the human analysis which is based on more experience instead of scientific technology. Here, the integration of Petri net and time-probability information could calculate the sequence of hidden trouble confidence coefficient and then could support the strategies for train door maintenance. In order to explain the algorithm we take the door slower closing condition as example which is compared with expert system to verify the feasibility of Time-possible Petri Net algorithm.

1. To design the extensible time Petri net model $\sum_{\mathrm{ET}} = (S, T, F, E, I, \delta, \tau, M_0)$ of rail train door system door closing process and to invert it to get TC-PPN. Then assume that $S_{\mathrm{sr}} \subseteq S' : s \in S_{\mathrm{sr}} \rightarrow M_l(s) > \lambda \wedge s^* = \emptyset$ is the place set when σ is greater than some prepared value λ under the final state of $\sum_{\mathrm{TP}}$. Afterwards, the single fault causal chain could be separated out based on S_{sr}.
2. In accordance with the simplicial fault causal chain $\sum_{\mathrm{ET}}^{S} = (S^S, T^S, I^S, E^S, \delta, \tau, M^S)$ separated in (33.1), some hidden state or event of FDES is detected by fault reasoning while some itself belongs to a kind of hidden trouble which need further analysis to obtain its hidden state.

The door closing model of train door system is modeling as Fig. 33.1 to explain the critical technology. The monitor system catches the following information (unit: ms): motor operation $O(S_{17}) = 1077$, Nut locking device withdrawing the locking section with screw $O(S_{18}) = 1145$, the door carried gantry driving the door device longitudinal movement $O(S_{19}) = 2892$, the stopping pins bounding $O(S_{30}) = 2243$, the door closing action $O(S_{32}) = 2951$, in which $O(x)$ represents the earliest time that catches x.

Transfer $\sum_{\mathrm{ET}}$ to $\sum_{\mathrm{TP}}$, the backward inference Petri net is called as II Petri net. If $\sum_{\mathrm{II}} = \sum_{\mathrm{TP}}$, then $\sum_{\mathrm{II}} = (S_{\mathrm{II}}, T_{\mathrm{II}}, F_{\mathrm{II}}, I_{\mathrm{II}}, E_{\mathrm{II}}, \delta, \tau, M_O^{\mathrm{II}})$, where every element is arranged as follows:

1. $S_{\mathrm{II}} = \{S_{19}, S_{20}, S_{21}, S_{24}, S_{25}, S_{26}, S_{28}, S_{29}, S_{30}, S_{31}, S_{35}, S_{36}, S_{37}, S_{38}\}$;
2. $T_{\mathrm{II}} = \{T_{18}, T_{19}, T_{20}, T_{21}, T_{23}, T_{24}, T_{25}, T_{26}, T_{30}\}$; F_{II} is shown in Fig. 33.4, $I_{\mathrm{II}} = \{\}$;
3. $E_{\mathrm{II}} = \{e_1, \ldots, e_9\}$;

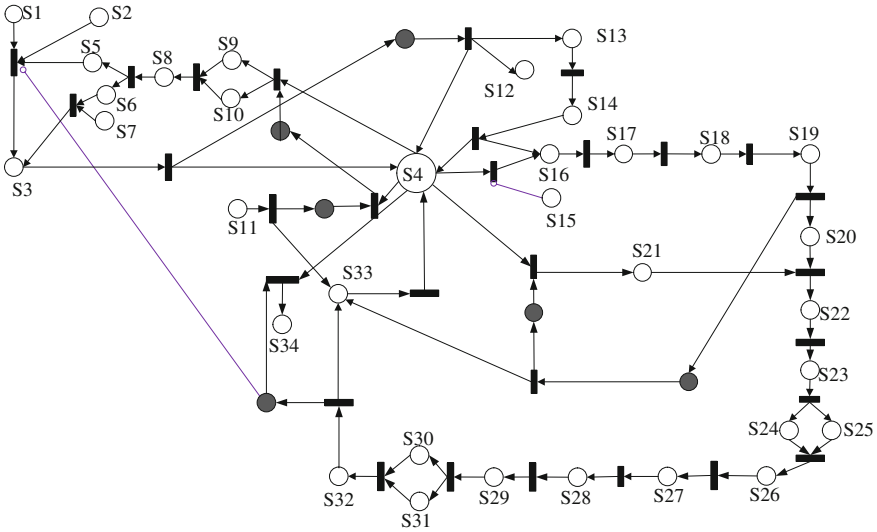


Fig. 33.4 Door closing petri net model

4. $\delta(e_2) = \{t_{19}\}$, $\delta(e_3) = \{t_{20}\}$, $\delta(e_4) = \{t_{21}\}$, $\delta(e_5) = \{t_{25}\}$, $\delta(e_6) = \{t_{24}\}$, $\delta(e_7) = \{t_{25}\}$, $\delta(e_8) = \{t_{26}\}$, $\delta(e_8) = \{t_{26}\}$, $\delta(e_9) = \{t_{30}\}$, $\delta(e_{10}) = \{t_{31}\}$;
5. $\tau(t_{18}) = \tau(t_{19}) = \tau(t_{26}) = \tau(t_{30}) = [10, 40]$, $\tau(t_{21}) = \tau(t_{23}) = \tau(t_{24}) = \tau(t_{25}) = [510, 540]$, $\tau(t_{20}) = [20, 40]$;
6. $M_{\sigma-0} = [1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1]$, $M_{v-0} = \{[1077, 1077], \varepsilon, \varepsilon, \varepsilon, [1145, 1145], \varepsilon, [2243, 2243], \varepsilon, \varepsilon, \varepsilon, [2892, 2892], \varepsilon, \varepsilon, [2951, 2951]\}$;
7. Build up the transpose of output matrix $(A^+)^T$ and the transpose of input matrix $(A^-)^T$;
8. Calculate based on Bayes posterior probability $p\ (S_{20} = 0|S_{19} = 1) = 0.00018$, $p\ (S_{24} = 0|S_{20} = 0) = 0.9952$, $p\ (S_{24} = 0|S_{19} = 0) = 0.9952$, $p\ (S_{25} = 0|S_{24} = 1) = 0$, $p\ (S_{26} = 1|S_{25} = 0) = 0.00012$, $p\ (S_{28} = 0|S_{26} = 1) = 0$, $p\ (S_{29} = 0|S_{26} = 0) = 0.9952$, $p\ (S_{30} = 1|S_{28} = 0) = 0$, $p\ (S_{31} = 0|S_{29} = 0) = 0.9952$, $p\ (S_{35} = 0|S_{30} = 1) = 0$, $p\ (S_{35} = 0|S_{31} = 1) = 0$.

Because of $Max\{p_i\} = 0.9952$, the corresponding possibility is transferred according to Eq. (33.4), construct $e_1 - e_9$ trigger matrix.

$M_0[e_{10} > M_1$: $M_{\sigma-1} = [1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0.9025, 0.9025, 1]$; $M_{v-1} = \{[1077, 1077], \varepsilon, \varepsilon, \varepsilon, [1145, 1145], \varepsilon, [2243, 2243], \varepsilon, \varepsilon, \varepsilon, [2892, 2892], [2905, 2935], [2905, 2935], [2951, 2951]\}$;

Suppose that a trigger sequence is $e_9 \rightarrow e_8 \rightarrow e_7 \rightarrow e_6 \rightarrow e_5 \rightarrow e_4 \rightarrow e_3 \rightarrow e_2 \rightarrow e_1$, then, the computing example is $\varepsilon, \varepsilon, [1145, 1145], \varepsilon, [2243, 2243], \varepsilon, \varepsilon, \varepsilon, [2892, 2892], [2905, 2935], [2905, 2935], [2951, 2951]\}$;

$M_1[e_9 > M_2$:$M_{\sigma-2} = [1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0.9025, 0.9025,1]$; $M_{v-2} = \{[1077, 1077], \varepsilon, \varepsilon, \varepsilon, [1145, 1145], \varepsilon, [2243, 2243], \varepsilon, \varepsilon, \varepsilon, [2880, 2900], [2905, 2935], [2905, 2935], [2951, 2951]\}$.

Table 33.2 Hidden fault inference results

Hidden fault name	Confidence degree	Hidden fault name	Confidence degree
Safety interlock circuit opening	0.9025	Door locking block invalid	0.6634
The roller rolling slowly in the slide way	0.9025	S1 turnoff	0.5987
Driving motor turnoff	0.6634	Door location sensor invalid	0.5404

Accordingly, event S_{14} and S_{23} failure possibility in $\sum_{TP}$ is 0.9025, the time interval [2905, 2935] is the fault possible occurring period. What is noticeable is $M_1[e_9 > M_2$ which demonstrates that state S_{22} which is induced by S_{19} and S_{23} has been verified the time interval is impartial for catching it from the perspective of fault spreading time trail (Table 33.2).

1. Under normal condition $S1$ switch action information should be delivered to EDCU but here the state of $S1$ is not been caught;
2. S_{19} the door carried gantry driving the door device longitudinal movement

According to the analyzed result, the sensor and state catching device related to stroke switch $S1$ and the interactive interface of EDCU controlling level and mechanic controlling level should be checked to avoid ignorance. The pin retainer and other mechanic device need to be maintaining so as to remove the hidden failure.

33.5 Conclusion

This paper researches an approach to model the fault causal chain with time and probability combining with Petri net. The ETPN and the TPPN gives the reasoning procedure based on time sequence and Bayes probability formula to analyze the FDES without integral information to support the inference. The hidden trouble reasoning method this paper proposes could give the hidden trouble possibility sequence with determinate evidence and save much economic and folk depletion. It has much more value to analyze the large-scale and complex system in the rail transit field such as the train door system. The research of TC-PPN predigestion continues to get a high efficiency. And next the method should focus on other system hidden fault analysis to prove its applicability.

Acknowledgement This work is supported by the Excellent Doctor Fund under Grant 2014YJS128 and the State Key Laboratory Program under Grant RCS2014ZT24. Thanks for the help.

References

1. Svärd C, Nyberg M, Frisk E (2011) A greedy approach for selection of residual generators. In: Proceedings of the 22nd international workshop on principles of diagnosis (DX-11), F. Citeseer
2. Cocquempot V, El Mezyani T, Staroswiecki M (2004) Fault detection and isolation for hybrid systems using structured parity residuals. In: Proceedings of the 5th Asian control conference, F 2004. IEEE, pp 1204–1212
3. Schmidt KW, Boutalis YS (2012) Fuzzy discrete event systems for multi-objective control: framework and application to mobile robot navigation. IEEE Trans Fuzzy Syst 20(5):910–922
4. Wei She, Yangdong Ye, Qian Chen (2014) Diagnosability analysis of fuzzy discrete event system based on extended fuzzy petri net. Comput Sci 41(7):62–67 (in Chinese)
5. Fraca Estíbaliz, Júlvez Jorge, Silva Manuel (2015) Hybrid and hybrid adaptive petri nets: on the computation of a reach ability graph. Nonlinear Anal Hybrid Syst 16:24–39
6. Basile F, Cabasino MP, Seatzu C (2015) State estimation and fault diagnosis of labeled time petri net systems with unobservable transitions. IEEE Trans Autom Control 60(4):997–1009
7. Basile F, Chiacchio P, de Tommasi G (2009) An efficient approach for online diagnosis of discrete event systems. IEEE Trans Autom Control 54(4):748–759
8. Wu S, He Z, Qian C etc (2011) Application of fuzzy petri net in fault diagnosis of traction power supply system for high-speed railway. Power Syst Technol 35(9):79–85 (in Chinese)
9. Jianwei Yang, Zhengyou He (2012) Study on recognition of fault transients using hybrid fuzzy petri net. Power Syst Technol 36(2):250–256 (in Chinese)
10. She W, Ye Y (2012) PPN: a probabilistic model for fault detection and diagnosis. In: IEEE 2012 12th international conference on intelligent systems design and applications
11. Suter J, Stoller N (2014) Use of simulators to investigate complex issues relating to human factors. In: FORMS/FORMAT 2014—10th symposium on formal methods for automation and safety in railway and automotive systems, pp 86–96
12. Malakar B, Roy BK (2014) Rail Way fail-safe signalization and interlocking design based on automation petri net. In: 2014 international conference on information communication and embedded systems, ICICES 2014, 5 Feb 2014
13. Nie M, Tan WW (2012) Theory of generalized fuzzy discrete event system. In: Proceedings of the 2012 IEEE international conference on fuzzy systems. IEEE, pp 1–7

Chapter 34
A Multi-flow Routing Strategy Based on Streamline Effect for Optical Burst Switching Networks Used in Rail Transportation

Tairan Zhang, Guotao Jiang and Jun Tang

Abstract We demonstrate a multi-flow routing strategy based on streamline effect for optical burst switching (OBS) networks used in rail transportation. The simulation results fully prove that the strategy can achieve fast speed and ideal burst loss ratio (BLR).

Keywords Optical burst switching · Streamline effect · Routing

34.1 Introduction

Because there will be massive burst data existing in the rail transportation networks in the future, optical burst switching (OBS) is becoming a promising technique to promote the network capability. In OBS networks, routing is a good way of controlling the burst loss ratio (BLR) since it can arrange rational paths for flows, which results in realizing load balancing as well as avoiding the contention and congestion effectively. Many routing techniques have been presented, but there are some disadvantages. The shortest path first (SPF) is too static to deal with the dynamic contention and congestion in the OBS networks. Each flow in the minimizing burst loss routing scheme [1, 2] usually focuses on how to minimize its burst loss selfishly, so some flows may conflict with each other on a same link, which finally makes the BLR of the whole OBS networks worse. Deflection routing [3, 4] is able to temporarily deflect the burst data flow to another node with less traffic for bypassing the node that lies on its original path and has encountered serious contention problem. However, it may introduce longer delay or endless loop, and even deteriorate the load unbalance in OBS networks. Some other routing strategies try to use self-learning [5, 6], integer linear programming [7] and network optimization theory

T. Zhang (✉) · G. Jiang · J. Tang
CSR Zhuzhou Institute Co. Ltd., Shi Dai Road, 169, Zhuzhou,
Hunan Province, China
e-mail: zhangtr@teg.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_34

[8] for improving network performances. However, because of not considering the streamline effect [9], their accuracy and effectiveness are questionable.

In this paper, we present four multi-flow routing options based on streamline effect for OBS networks and choose one of them as our proposed routing strategy. Since the strategy takes the streamline effect into consideration, it is more reliable than the strategies mentioned above. Moreover, the NS2 simulation results prove that the proposed strategy not only achieves fast speed, but also has ideal BLR.

34.2 Routing Strategy

Due to lack of optical memory in OBS networks, if some flows coming from different inputs will go to a same output simultaneously, the bursts of those flows have to contend for limited bandwidth resource, which leads to bursts loss inevitably. Once those bursts have been merged together, they form a new stream and there is no contention among them anymore. Hence those bursts just need to contend with the bursts from other streams in the next nodes, which is called as streamline effect by [9]. On the basis of function (34.1), where ρ_i is the load of input i, K is the number of data channels of an input and W is the number of inputs of a node, we can know the flow with more traffic tends to obtain less BLR. Thus if a flow just chooses a path for achieving its minimum BLR while routing, it is likely to bring about serious load unbalance in OBS networks according to streamline effect. Therefore, every flow should consider what influence it may generate on other flows while routing, which is named as the cost and shown in function (34.2) [10]. Loss_{ij}(Before) and Loss_{ij}(After) are the bit loss rates of the input j of node i before and after the flow going through that node, respectively. Hence when there are several alternative paths for choosing by a flow, the flow should select the path with the least cost. The cost is evaluated based on the real-time feedback information that can reveal the latest traffic situation of the OBS networks.

$$P_i = \frac{L(\rho_{\text{all}}) - L(\rho_i)}{1 - L(\rho_i)} \quad \rho_{\text{all}} = \sum_{i=1}^{W} \rho_i \quad L(\rho) = \frac{(\rho)^K / K!}{\sum_{i=0}^{K} (\rho)^i / i!} \tag{34.1}$$

$$C = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\text{Loss}_{ij}(\text{After}) - \text{Loss}_{ij}(\text{Before}) \right) \tag{34.2}$$

Supposing N new flows need to be routed in the OBS networks that have contained several existing background flows and each new flow has M alternative paths for selection. Here we design four routing options and finally use the one of them as our proposed routing strategy: serial routing, parallel routing, double parallel routing, and traversal routing. Serial routing selects the optimal path with the minimum cost from M prepared alternative paths for each flow in sequential, which means the ith flow cannot route until the former $i - 1$ flows have been

processed. Serial routing may achieve good BLR since the latter flow has a better knowledge of the network traffic situation than former flows, but the whole process of N flows is too long. On the contrary, the parallel routing is to rout N flows in parallel, which results in faster speed but higher BLR. In order to improve the BLR of parallel routing and keep rapid, double parallel routing routes N flows parallel in two times. First, the N flows evaluate the costs and choose their optimal paths together as parallel routing did. But the difference is that all new flows do not transmit their bursts at once after obtaining the optimal paths. Instead, they re-evaluate the costs according to first evaluating results. For example, if the optimal paths of flow 1 and 2 both include a same link l, flow 1 will take the sum of old background traffic and flow 2 as the new background traffic load of link l, and then flow 1 re-evaluates the cost of link l again. After re-evaluation, since flow 1 realizes that the traffic load of link l has become heavier due to the participation of flow 2, flow 1 may choose another path in the second parallel routing for avoiding the contention with flow 2. In fact, the key point of double parallel routing is trying to give the second routing chance to new flows, helping them do the proper selection after knowing the network traffic situation clearly. Therefore, its BLR performance can be guaranteed without lowering speed. The last option, traversal routing, is to obtain its optimal routing from M^N possibilities by exhaustive method. Obviously, the traversal routing is very time-consuming and unpractical due to its huge complexity and long delay since M^N may be a tremendous value.

34.3 Simulation Results

We construct a Nation Science Foundation Network (NSFNet) topology and do the simulations by using NS2 (network simulation software). Each link has eight data channels and the bandwidth of each channel is 2.5 Gbps. There are 400 random background flows in the NSFNet and their total traffic is 160 Gbps. Now N new random flows will be added into the NSFNet, and each of them has M random alternative paths. Three scenarios have been tested. (1) The load ratio of new flows to background flows increases from 0.05 to 0.45 when N is 10 and M is 3; (2) N varies from 8 to 14 when the load ratio is 0.3 and M is 3; (3) the load ratio is still 0.3 and N is 10, M changes from 2 to 6.

The results of Figs. 34.1, 34.2 and 34.3 correspond to the three scenarios listed above. From Fig. 34.1, we can see the BLRs of all routing strategies increase with the growth of load ratio. The BLR of SPF is the highest due to its static routing rule. The double parallel routing and traversal routing are better than serial routing and parallel routing. In Fig. 34.2, the BLRs from the lowest to highest are traversal routing, double parallel routing, serial routing, parallel routing, and SPF, which conforms to the theory analysis mentioned above. Since the overlapping degree of optimal paths increases and the average flow traffic decreases as the N becoming

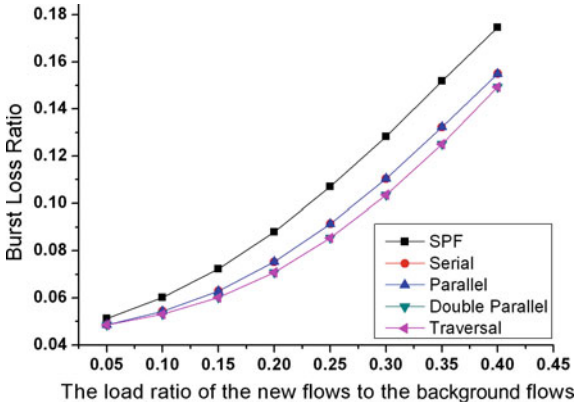


Fig. 34.1 The BLR comparison when the load ratios of the new flows to the background flows increase

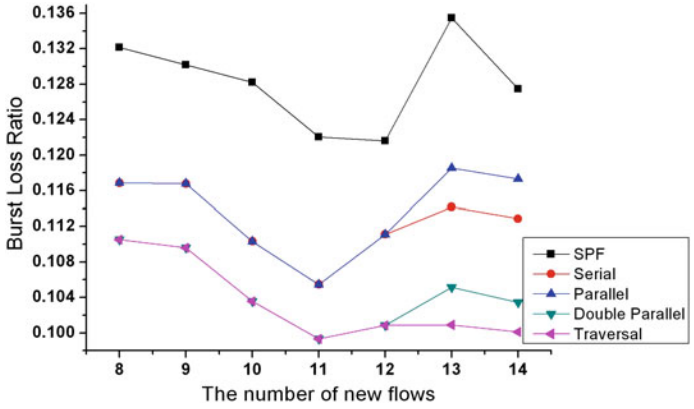


Fig. 34.2 The BLR comparison when the number of new flows increases

larger, there is a tradeoff between the overlapping degree of optimal paths and the average flow traffic, which makes the lowest BLRs appear when N is 9. As shown in Fig. 34.3, with the increment of M, each new flow has more chance to obtain a better routing, thus all of BLRs are reduced correspondingly. However, the BLR of double parallel routing is still lower than the BLRs of serial routing and parallel routing, although just a little more than the BLR of traversal routing. Because double parallel routing is faster than serial routing, acquiring better BLR than parallel routing and acting more practically than traversal routing, it is an ideal streamline effect based routing strategy for multiple flows in OBS networks.

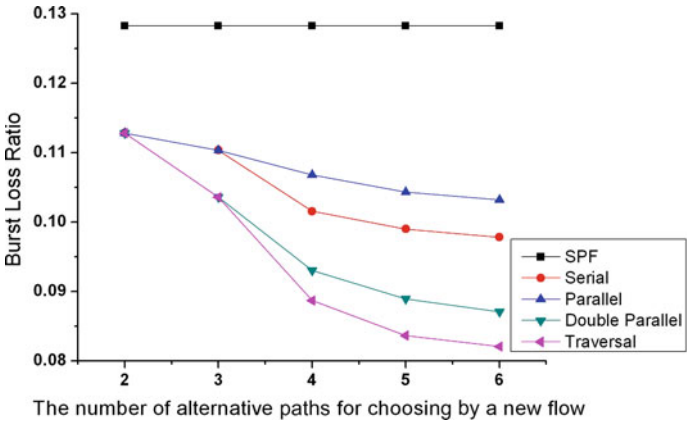


Fig. 34.3 The BLR comparison when the number of alterative paths increases

34.4 Conclusion

Although we have proposed four multi-flow routing options based on streamline effect for OBS networks in this paper, the double parallel routing is the best one since it gives the flow twice the chances to evaluate the cost of path. As a result, the double parallel routing has a full knowledge of the network traffic situation and is aware of what cost will be produced while routing, which makes it finally achieve the ideal BLR that is close to the optimal value and high speed due to its parallel operation.

References

1. Yang L, Rouskas GN (2006) Adaptive path selection in OBS networks. J Lightwave Technol 24(8):3002–3011
2. Lu J, Liu Y, Mohan G, Chua K (2006) Gradient projection based multi-path traffic routing in optical burst switching networks. Workshop on high performance switching and routing, pp 379–384
3. Wang X, Morikawa H, Aoyama T (2003) Burst optical deflection routing protocol for wavelength routing WDM networks. Optical networking and communication, pp 120–129
4. Zalesky A, Vu HL, Rosberg Z, Wong EM, Zukerman M (2003) Reduced load Erlang fixed point analysis of optical burst switched networks with deflection routing and wavelength reservation. 1st international workshop optical burst switching
5. Ishii D, Yamanaka N, Sasase I (2005) Self-learning route selection scheme using multi path searching packets in an OBS network. J. Opt Netw 4(7):432–445
6. Kiran Y, Venkatesh T, Murthy CSR (2007) A reinforcement learning framework for path selection and wavelength selection in optical burst switched networks. J Sel Areas Commun 25(9):18–26
7. Teng J, Rouskas G (2005) Routing path optimization in optical burst switched networks. In: 9th conference on optical network design and modeling, pp 1–10

8. Klinkowski M, Pioro M, Careglio D, Marciniak M, Sole-Pareta J (2007) Non-linear optimization for multi-path source routing in OBS networks. Commun Lett 11:1016–1018
9. Phung M, Shan D, Chua K, Gurusamy M (2006) Performance analysis of a bufferless OBS node considering the streamline effect. Commun Lett 10(4):293–295
10. Wu G, Dai W, Li X, Chen J (2014) A maximum-efficiency-first multi-path route selection strategy for optical burst switching networks. Int J Light Electron Opt 125(10):2229–2233

Chapter 35
Urban Railway Network Traffic Prediction with Spatiotemporal Correlations Matrix

Weijuan Shao and Man Li

Abstract Urban railway network traffic prediction is a fundamental capability method as smart transportation technologies. Urban railway traffic prediction is a need that traffic authorities have begun demanding with a rapid increase in the number of passenger flow. Contemporary smart transportation technologies require the traffic prediction capability to be both available and scalable to apply to urban networks. In this paper, spatiotemporal correlations matrix method will be presented to traffic prediction.

Keywords Urban railway network · Traffic prediction · Spatiotemporal correlations matrix

35.1 Introduction

With merits of high convenience, efficiency and reliability, urban railway traffic system is growing rapidly in recent years. Simultaneously, urban railway traffic prediction is a need that traffic authorities have begun demanding with a rapid increase in the number of passenger flow, as the current tools providing traffic

W. Shao (✉) · M. Li
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, No. 3 Shang Yuan Cun, Haidian District,
Beijing, China
e-mail: shaoweijuan521@163.com

W. Shao · M. Li
School of Traffic and Transportation, Beijing Jiaotong University,
Beijing, China

W. Shao · M. Li
Beijing Research Center of Urban Traffic Information Sensing and Service
Technologies, Beijing Jiaotong University, Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_35

control assistance as well as traveler information from historical data are not widely available to apply predictive information [1, 2]. The body of documents on traffic prediction methods has increased constantly in the past few decades [3–5]. Most of works focus on traffic prediction in single station, although the increasing complexity of urban network operations [6–8]. Hence, the drawback is that the approach using no or very few interactions between links may not be able to take relevant traffic elsewhere into account on the network. Therefore, it is of vital importance to predict passenger volume according to the requirements in terms of different time, different sizes of railway network in order to guide urban railway network performance.

In this paper, with the goal to enable a traffic prediction tool that can be used in network widely at any time, a prediction model of spatiotemporal correlations matrix is established in this paper, where the extension takes spatial and temporal interactions into account in a new way, specialized to the background of urban railway traffic. The next part describes the model and wavelet neural network (WNN) used by the model. In the last section we make use of data arriving with a frequency of 1 h to provide accurate forecasts the $t + 1$'st into the future of part of stations in Beijing subway. Finally, the paper presents conclusions and recommendations for further work.

35.2 Building Urban Railway Network Traffic Prediction Model Based on Spatiotemporal Correlations Matrix

35.2.1 Urban Railway Network Traffic Prediction Model

Assuming that urban railway network has station named and set for the study time; is stand for traffic of station in; is stand for traffic of station in spatiotemporal correlations matrix:

$$Q = \begin{pmatrix} Q_{A_1t_1} & Q_{A_2t_1} & Q_{A_3t_1} & \cdot & Q_{A_mt_1} \\ Q_{A_1t_2} & \cdot & \cdot & \cdot & \cdot \\ Q_{A_1t_3} & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & & \\ Q_{A_1t_n} & Q_{A_2t_n} & Q_{A_3t_n} & \cdot & Q_{A_mt_n} \end{pmatrix} \tag{35.1}$$

Each row of the matrix represents all stations traffic at a time in the network and each column represents one station's traffic at all time.

Table 35.1 The reference table of correlation coefficients

$\|R\|$	[0 0.3)	[0.3 0.5)	[0.5 0.8)	[0.8 1]
Value	No	Low	Medium	High

35.2.2 Correlation Analysis

Measuring the degree of linear correlation between two random variables correlation coefficients can be used to calculate the size of the traffic correlation between different stations.

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{35.2}$$

The value of R can be referred to Table 35.1 [9].

35.3 Building Passenger Flow Prediction Model Based on Urban Railway Network Traffic Prediction Model

35.3.1 Construction of Simulation Model

WNN is one kind of BP neural network whose error is backpropagation but signal is forward propagating. Hidden layer nodes in neural network are wavelet basis function. According to the characteristics of traffic in the urban rail transit network, in this paper we use a three-layer structure of WNN—input layer, hidden layer, and output layer.

When the degree of node is 4, input of the matrix Q is a sub-matrix $Q_p^{(k)}$ $(3 \times m)$ in $t_{j-3}, t_{j-2}, t_{j-1}$ (k is the number of iterations) whose every column contains of three traffic data is sub-input and the number of sub-input m is in total and output of the matrix Q is a row vector $(m \times 1)$ in t_j. In sum, it is to establish a WNN that has m sub-matrix. Each sub-WNN hidden layer contains n_j neuron. When the number of sub-input layer node is $n_{j(\text{in})}$ and the number of sub-input layer node is $n_{j(\text{out})}$ we can refer to the formula (35.3) to design the number of hidden nodes where a is constant between 1 and 10.

$$n_j = \sqrt{n_{j(\text{in})} + n_{j(\text{out})}} + a \tag{35.3}$$

Hidden layer of the transfer function is Morlet wavelet basis function

$$\varphi(x) = \cos(1.75x)\mathrm{e}^{x^2/2} \tag{35.4}$$

Output of hidden layer is the following formula:

$$\varphi(h) = \varphi_h\left[\frac{\sum_{j=4}^{m}\omega_{jh}Q_j - b_h}{a_h}\right] \tag{35.5}$$

where a_j is called the scale factor of wavelet function and b_j is called translation factor of wavelet function.

In sub-WNN:

$$a_{sh}^{(k)} = \sum_j w_{s,h}^2 Q_{sh}^{(k-1)} \tag{35.6}$$

where $a_{sh}^{(k)}$ indicates input of hidden layer neurons; $w_{s,h}^2$ denotes the weight from input layer to hidden layer.

$$b_{h0}^{(k)} = \varphi\left(a_{sh}^{(k)}\right) \tag{35.7}$$

where $b_{h0}^{(k)}$ indicates output of hidden layer neurons; $\varphi(x)$ is stands for hidden layer function.

$$Q_j^{(1)} = \sum_j b_{\mathrm{h0}}^{(k)} w_{h,o}^{(3)} \tag{35.8}$$

where $w_{h,o}^3$, denotes the weight from hidden layer to output layer.

In order to make the minimum error and obtain the optimal parameters of the network model, conjugate gradient descent method is used in this paper to evaluate the minimum value of error function.

35.3.2 Simulation Algorithm

Repeat the above steps to predict traffic of all correlative stations in time t_n. Operation scheduling plan and formula (35.9) are referred to modify forecast results.

35.4 Example Analysis

35.4.1 Test Results and Analysis

Applying the most recent actual passenger flow volume, the forecasting accuracy for volume is measured by

$$C = \frac{1}{2}\sum_{j=4}^{n-3} (\text{forecast}\, Q(j) - Q(j))^2 \tag{35.9}$$

where j is the number of iterations in the test data set.

Let us take one of Beijing subway transfer station of line 1 and line 2 Fuxingmen station as an example. Considering the scope of passenger flow influence between each other station, we select north and south adjacent two stations respectively: Fuchengmen (n1), Changchunjie (s1), Nanlitulu (w1), Xidan (e1), Chegongzhuang (n2), Xuanwumen (s2), and Muxidi (w2).

We group 1 month into five classes and for each class we introduce three modes: daytime, which we refer to as "morning peak" and is defined as (7:00 a.m.–8:00 a.m. 8:00 a.m.–9:00 a.m. 9:00 a.m.–10:00 a.m.) and "flat peak" and is defined as (12:00 p.m.–13:00 p.m. 13:00 p.m.–14:00 p.m. 14:00 p.m.–15:00 p.m.) and nighttime, called "off-peak" which we refer to as "evening peak" and is defined as (17:00 p.m.–18:00 p.m.; 18:00 p.m.–19:00 p.m.; 19:00 p.m.–20:00 p.m.). The five day-of-month classes are March 1, 8, 15, 22, and, 29, 2011. Therefore, we adopted four separate training templates and one test template in total. The training set consists of recorded data from the first four consecutive weeks and it was fitted through standard maximum likelihood estimation. In Table 35.2, we show four sample correlation coefficients for all 288 links.

The selected link parameters are shown in Table 35.2. The station can be access by each station in two time-steps. It has four neighboring stations from the extracted railway network (values are 0.9920, 0.9960, 0.9204, and 0.1488).

Interestingly, the table shows that some two time-steps neighboring stations have strong correlation with Fuxingmen station Particularly (Chegongzhuang station, Xuanwumen station, and Muxidi station). What is more, some of two time-steps neighboring stations have bigger correlation (larger coefficients) than the nearby neighboring stations (compare the third and fifth row). We will go through the results of the prediction to explain this phenomenon.

35.4.2 Prediction Performance

Let us choose the top seven highly relevant stations from Table 35.2 as the value of m (m = 7) and using peak hour passenger flow data of the first four Tuesdays in March 2011 (3.1, 3.8, 3.15, 3.1) as test sample data (j = 4) to establish

Table 35.2 Other stations' passenger flow volume correlation coefficient with Fuxingmen station for four Tuesdays, 7:00 a.m. till 20:00 p.m.

$\|R\|$	3.1	3.8	3.8	3.22	3.29	Rank
Fuxingmen	1	1	1	1	1	
Fuchengmen	0.9935	0.9940	0.9938	0.9868	0.9920	1
Chegongzhuang	0.9795	0.9819	0.9835	0.9826	0.9819	2
Xuanwumen	0.9823	0.9813	0.9813	0.9809	0.9815	3
Nanlitulu	0.9747	0.9582	0.9666	0.9645	0.9660	4
Changchunjie	0.9274	0.8961	0.9323	0.9256	0.9204	5
Muxidi	0.9823	0.8698	0.8939	0.8855	0.9079	6
Xidan	0.2783	−0.1370	0.2435	0.2102	0.1488	7
Tiananmenxi	−0.2200	−0.0416	0.0781	0.1821	−0.00035	8

spatiotemporal prediction matrix Q (11) of urban railway network to predict passenger flow data for the last Tuesday in March.

$$Q = \begin{bmatrix} Q_{A_1t_1} & Q_{A_2t_1} & Q_{A_3t_1} & Q_{A_4t_1} & Q_{A_5t_1} & Q_{A_6t_1} & Q_{A_7t_1} \\ Q_{A_1t_2} & Q_{A_2t_2} & Q_{A_3t_2} & Q_{A_4t_2} & Q_{A_5t_2} & Q_{A_6t_2} & Q_{A_7t_2} \\ Q_{A_1t_3} & Q_{A_2t_3} & Q_{A_3t_3} & Q_{A_4t_3} & Q_{A_5t_3} & Q_{A_6t_2} & Q_{A_7t_3} \\ Q_{A_1t_4} & Q_{A_2t_4} & Q_{A_3t_4} & Q_{A_4t_4} & Q_{A_5t_4} & Q_{A_6t_2} & Q_{A_7t_4} \end{bmatrix} \tag{35.10}$$

For any $Q_{A_it_j}(i = 1, 2, 3, \ldots, 7)(j = 1, 2, 3, 4)$ represents the day morning and evening passenger flow at different times within an hour.

Peak period is prone to traffic congestion, so accumulating unit time surplus forecast is very important.

$$\text{Cumulative residual traffic} = \text{ABS}(Q_{\text{in}} - Q_{\text{out}}) \tag{35.11}$$

where Q_{in} is passenger entering into the station within an hour; Q_{out} is passenger leaving the station.

35.4.3 Predicted Results Analysis

Take the first four days with number of 112 traffic flow data as the training sample to study training of network and with number of 28 data in the last day for the test sample within 1 h to forecast the traffic flow. The results are shown in Figs. 35.1, 35.2, 35.3 and 35.4.

Prediction performance metrics is shown in Table 35.3.

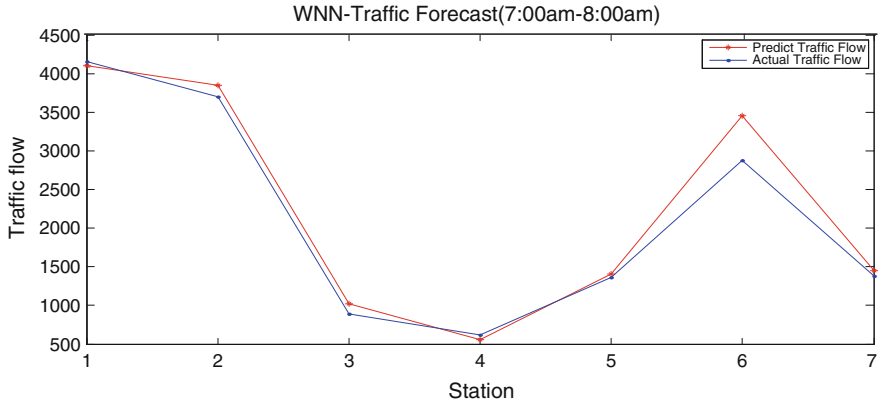


Fig. 35.1 WNN traffic forecast (7:00 a.m.–8:00 a.m.)

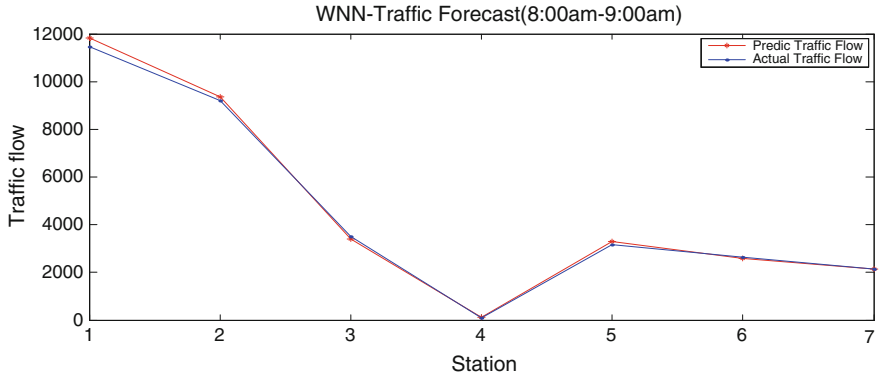


Fig. 35.2 WNN traffic forecast (8:00 a.m.–9:00 a.m.)

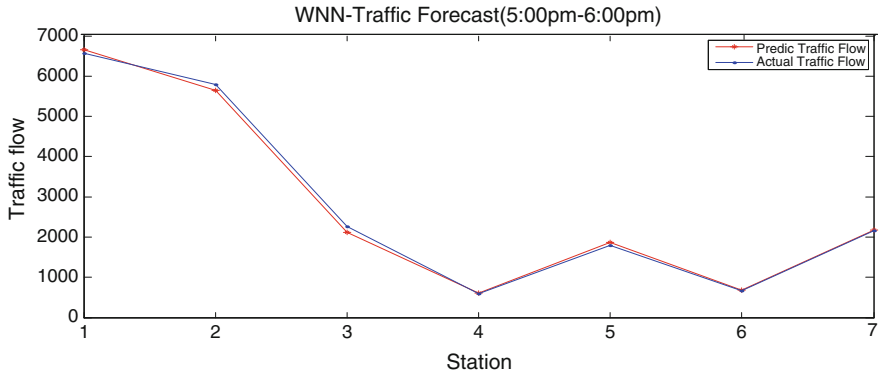


Fig. 35.3 WNN traffic forecast (5:00 p.m.–6:00 p.m.)

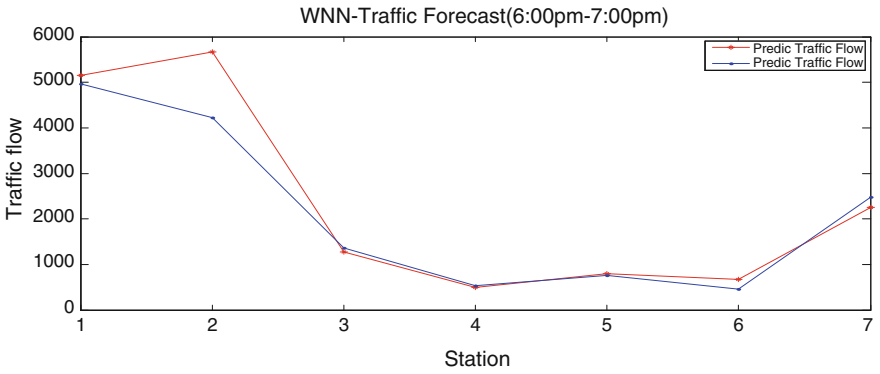


Fig. 35.4 WNN traffic forecast (5:00 p.m.–6:00 p.m.)

Table 35.3 WNN prediction performance metrics in weekday

Test number	MAPE	MAD	RMSE
1	0.080695	151.8571	13.3104
2	0.075982	119.1429	11.78983
3	0.026836	71.71429	9.146948
4	0.163704	321.7143	19.37352
Mean	0.086804	166.1071	13.56095

35.5 Conclusion

Traffic prediction with spatiotemporal correlations matrix can improve the efficiency of prediction. However, gap in different scales of station passenger flow is so huge, and sample data are not enough on the whole in this paper, and therefore the prediction performance is barely satisfactory and needs to be improved in the future studies.

Acknowledgment This work was supported in part by the China National Science and Technology Support Program under Grant (NO. I13B300020).

References

1. Al-Deek H, Ishak S, Wang M (2001) A new short-term traffic prediction and incident detection system on I-4, vol I. Final research report, transportation. Systems Institute (TSI), Department of Civil and Environmental Engineering, University of Central Florida
2. Kamarianakis Y, Prastacos P (2003) Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches. Transp Res Rec J Transp Res Board 1857:74–84

3. Vlahogianni EI, Karlaftis MG, Golias JC (2005) Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transp Res Part C: Emerg Techn 13(3):211–234
4. Kamarianakis Y, Kanas A, Prastacos P (2005) Modeling traffic volatility dynamics in an urban network. Transp Res Rec J Transp Res Board 1923:18–27
5. Giacomini R, Granger CWJ (2001) Aggregation of space-time processes. Manuscript, Department of Economics, University of California, San Diego
6. Min WL, Tsay R (2005) On canonical analysis of multivariate time series. Stat Sinica 15:303–323
7. Min W, Wynter L (2011) Real-time road traffic prediction with spatio-temporal correlations. Transp Res Part C Emerg Technol 19:606–616
8. Wei Y, Chen M-C (2012) Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transp Res Part C Emerg Technol 21:148–162
9. Knowles RD (2006) Transport shaping space: differential collapse in time-space. Science Direct

Chapter 36
Study on Factors Impacting the Boarding and Alighting Movement in Subway Station

Yafei Wang, Limin Jia and Man Li

Abstract For subway train stop (TS) time, this paper conducts field research in Beijing subway stations and make statistical analysis to determine the time and number of passengers of boarding and alighting; in order to improve the efficiency of passenger boarding and alighting movement, the paper analyzes the composition of factors impacting passenger boarding and alighting movement and the main factors affecting passenger boarding and alighting movement by interpretative structural model, and then it finds direct and deep reasons. Lastly, the article determines the impact of train-based and station-based factors on the rate of number of boarding passenger and number of alighting passenger and presents measures to improve the efficiency of passenger boarding and alighting based on both the train design and station facilities.

Keywords Subway station · Boarding and alighting · Factors · ISM

36.1 Introduction

With the growing of urban population and the increasing of residents travel, line volume in subway meets new demands, but it is decided by the station, and passenger boarding and alighting movement directly affects train stop (TS) time, which affects the train headway and impacts the size of the line volume. However, there are many factors influencing passenger boarding and alighting movement, which

Y. Wang (✉) · L. Jia · M. Li
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
e-mail: 13120966@bjtu.edu.cn; yafeiwang1126@126.com

Y. Wang · L. Jia · M. Li
School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Y. Wang · L. Jia · M. Li
Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_36

many scholars have studied. In terms of TS time, all the subway stations are facing the same problem that passenger boarding and alighting movement usually extends TS time, so the station becomes a bottleneck, which eventually leads to the decrease of line capacity. Improvement of solving this problem mainly includes train-based and platform-based solutions. The former measures include widening door (e.g., Mexico City), increasing doors per car (e.g., Hong Kong MTRC), reducing seats per car (e.g., Hong Kong MTRC); the latter measures include widening platform or extending platform (e.g., Moscow), separating platform for passenger boarding and alighting movement (e.g., Sao Paulo), writing the words "Step Aside" on the bright orange floor to separate the boarding and alighting passengers (e.g., New York) [1]. Similarly, shortening the gap between the train and platform can reduces TS time and movement time is reduced up to 0.1 s per passenger for each 10 cm decreasing in gap [2]. But some measures also play negative role in TS time, such as the use of some double-deck train on the rail transit, the research results show that although doors in these cars are wide enough, alighting time increased up to 0.3 s per passenger [1]. In addition, the passengers standing in the vestibule have a great influence on passenger boarding and alighting movement, especially those "sentries" who stand near the door [3]. However, the difficulty of this problem is that there are many uncontrollable environmental variables including the weather, types of passengers, behavior of passengers in rail transit operation, and the differences of pedestrian behavior between European and Asian were analyzed [4]. Therefore, the main work of this paper is to find factors that affect passenger boarding and alighting movement through field research of subway station and make analysis of the factors from three aspects of station-based, train-based, and passenger-based.

36.2 Field Research and Data Collection

This paper conducts field research in Beijing subway stations, and video recording method is used for the "key door" which means the busiest door (the door where passenger boarding and alighting movement needs the most time) in peak hours. The "key door" is determined by field observation research and inquiring subway coordinators, field research mainly means that investigators observe passenger boarding and alighting movement per door and make statistics on the number of boarding and alighting passengers by the manual counting method; while inquiring subway coordinators means that coordinators maintain the order of boarding and alighting passengers in peak hours for a long time and they know the number of passengers per door well. The paper identified the "key door" of Chongwenmen station on line 5 and video recording method was used for passenger boarding and alighting movement of "key door" in peak hours, finally data processing which includes time and numbers of boarding and alighting passengers of the video was taken.

The TS time includes six key time points in this paper, the TS, the door begins to open (BO), the door is fully opened (FO), the door begins to close (BC), the door is

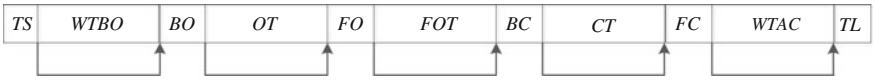


Fig. 36.1 Specific key time points of the train stop time

fully closed (FC), and the train leaves (TL). The time between TS and BO is called waiting time before opening (WTBO); the time between BO and FO is called opened time (OT); the time between FO and BC is called fully opened time (FOT); the time between BC and FC is called closing time (CT); the time between FC and TL is called waiting time after closing (WTAC), and the length of key time point is not considered in this paper. Specific time points are shown in Fig. 36.1 For the busy subway line, however, it is impossible to reduce the headway of trains in peak hours, so it needs to improve the efficiency of passenger boarding and alighting movement to decrease TS time. And in terms of TS time, WTBO and OT can be seen as invariants, which is suitable for CT and FOT, while passenger boarding and alighting movement needs to be completed during OT, besides the field survey finds that this phase includes passenger boarding and alighting time (BAT) and remaining time before closing door (RT).

It obtains specific time for each component part of TS time by making analysis of video's data statistics, and the changes of 20 consecutive cycles in morning peak hours are given in Fig. 36.2. It is shown in the figure that the TS time depends on passenger BAT and remaining time before closing door (RT) and BAT directly impact on RT. But the total time of boarding and alighting is determined by the number of boarding and alighting and how to complete boarding and alighting process, while boarding and alighting speed is decided by door width, vehicle interior layout, passenger density in the car and other factors.

Making statistical analysis of boarding and alighting passengers in Chongwenmen station on line 5, it is found that passenger boarding speed ranges from 0.375 to 1.875 person person/door/s. Passenger alighting speed varies between 0.5 and 2.375 person/door/s. And total number of boarding and alighting passengers exceeds 35 persons. While according to Figs. 36.2 and 36.3, on line 5 reaches 20 tph, so in order to meet train frequency, it is necessary to keep even increase boarding and alighting speed. Besides when headway is about 2 min, FOT varies between 20 and

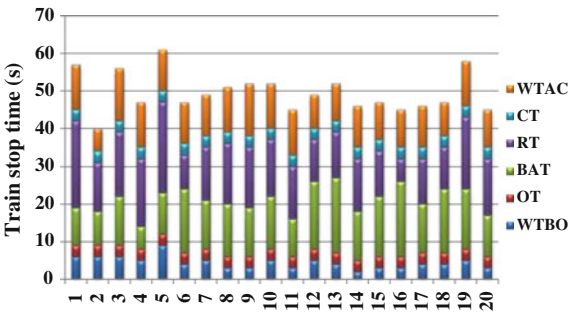


Fig. 36.2 Allocation map of each component part of train stop time

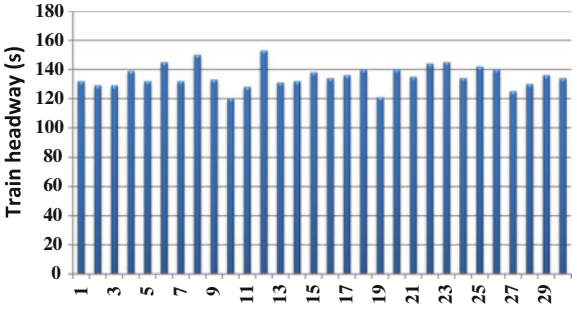


Fig. 36.3 Train headway

40 s, including BAT and RT which affects the reliability of operation, fluctuations of headway can be seen from Fig. 36.3, indicating that there are many uncertainties in the course of operations.

36.3 Analysis of Composition of Factors Impacting on Passenger Boarding and Alighting Movement

It has been able to determine the effect of various features on some stations through the study of various subway stations and the factors affecting passengers boarding and alighting movement fall into three aspects of station-based, train-based, and passenger-based in this part.

1. Station-based factors

Installation of platform screen doors (PSDs) has significant benefit; on the one hand it is advantageous to passengers' safety, avoiding casualties of falling into track; on the other hand it is effective to regulate air and reduce the noise and dust on the platform. However, installation of PSDs increases OT 1–2 s per train, and it can reduce the operating efficiency especially large flow lines.

Simultaneously, platform width and the distance between passengers and platform exit both impact passengers boarding and alighting movement. Platform width and the distance between passengers and platform exit directly affect the state of passenger flow on the platform or even impede the movement of getting into platform.

The platform management score is regarded as a factor in this paper, mainly considering management practices and facilities of the platform. And the Delphi method is used for detailed classification on a scale of 1–5. The detailed classification of the levels is shown in Table 36.1 [5].

2. Train-based factors

There is a gap of 10 cm between car and platform to prevent fractioning, considering the slight swing when the train runs, and the gap is called the horizontal

Table 36.1 Detailed classifications of platform management score

Level	Management practices and facilities of classification
1	No active management practices
2	Posters, advertisement, broadcasters, signs, and markings encouraging passengers to get off the train first
3	Queue signs and markings painted on the platform; announcements encouraging passengers to move along the platform and get off the train first; one coordinator on the platform during the peak hours
4	Facilities of controlling crowd; coordinator occasional announcements encouraging passengers to move along the platform and get off the train first; several coordinators on the platform during the peak hours
5	PSDs; coordinator frequent announcements encouraging passengers to move along the platform and get off the train first; many coordinators on the platform during the peak hours; platform entrances and exits are convenient

distance between the car and platform; In addition, the car floor is higher than platform 5–10 cm, and the altitude difference is called the vertical distance between the car and platform; so one can calculate the Pythagorean distance between the car and platform, which can impact on passengers boarding and alighting speed, stride, etc. Simultaneously, the capacity of vestibule hinders passengers boarding and alighting movement, especially those standing in the vestibule. Besides the capacity of vestibule has a relationship with door width, standback width, distance between door centerlines.

3. Passengers-based factors

Passengers standing in the vestibule can hinder boarding and alighting movement, especially those "sentries," and hence, the number of passengers through the vestibule reflects traffic of boarding and alighting and speed of passengers' boarding and alighting. Another variable is the ratio of the number of boarding passengers and the number of alighting passengers. When the sizes of two groups have larger differences, the passengers of dominant group board or alight easily, while the other group's passengers move with difficulty; when the sizes of two groups are similar, it can lead to a lot of movement conflicts.

In summary, the factors are shown in Table 36.2.

36.4 Analysis of Main Factors Affecting Passenger Boarding and Alighting Movement Based on ISM

In order to analyze the structural relationship, relevance and deep-seated reasons of the factors impacting on passengers boarding and alighting movement, it uses interpretative structural modeling (ISM) [6–8] for analytic hierarchy and forms concise and understandable hierarchical structure.

Table 36.2 Factors impacting on passengers boarding and alighting movement

Station-based factors	PSDs (S_1)
	Platform width (S_2)
	The distance between passengers and platform exit (S_3)
	Side platform
	Platform management score (S_4)
Train-based factors	Door width (S_5)
	The distance between door centerlines (S_6)
	Standback width (S_7)
	The Pythagorean distance between the car and platform (S_8)
	The ratio of longitudinal seats
	The capacity of vestibule (S_9)
Passengers-based factors	Number of passengers through the vestibule (S_{10})
	Number of boarding passengers/number of boarding passengers (S_{11})

1. Identify the main factors

Along with field research of Beijing subway station and the actual situation of the station, the paper identifies the main factors of ISM, including S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9, S_{10}, and S_{11}.

2. The establishment of correlation matrix

Correlation matrix R is established by using the Delphi method, and if it has a direct impact among factors, it is denoted as "1" in the matrix, otherwise "0", which can be seen in Eq. (36.1).

$$r_{ij} = \begin{cases} 1, & s_i \text{ directly influences } s_j \\ 0, & \text{others} \end{cases}, \quad (i,j = 1,2,\ldots 11) \tag{36.1}$$

3. Calculate reachable matrix

If the system A meets the requirement

$$(R+I)^{k-1} \neq (R+I)^k = (R+I)^{k+1} = M \tag{36.2}$$

Then matrix M is called the reachable matrix of system A. And the reachable matrix can be obtained by the correlation matrix

$$M = (R+I)^n \tag{36.3}$$

where I is 11 order identity matrices. And it gets M by MATLAB software.

$$M = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

4. Decomposition of reachable matrix

The article gets the reachable set $R(S_i)$, antecedent set $A(S_i)$, and intersection set of reachable set and antecedent set $R(S_i) \cap A(S_i)$ by decomposition of reachable matrix. And the reachable set of reachable matrix can be seen in Table 36.3, so is antecedent set and intersection set of reachable set and antecedent set.

Table 36.3 shows $L(1) = \{10\}$. And remove row 10 and column 10 in M. Besides the paper can get $L(2) = \{11\}$, $L(3) = \{4, 9\}$, $L(4) = \{1, 3, 5, 6, 7\}$ and $L(5) = \{2, 8\}$ with the same method.

5. Hierarchical structure analysis

Based on the above analysis, hierarchical structure of the main factors impacting on passengers boarding and alighting movement can be seen in Fig. 36.4.

Figure 36.4 shows that the main factors are divided into five levels by using ISM. The first level is the number of boarding passengers/number of boarding passengers, and the second level is the number of passengers through the vestibule; while the third level includes platform management score and the capacity of

Table 36.3 Level 1 reachable set and antecedent set

S_i	$R(S_i)$	$A(S_i)$	$R(S_i) \cap A(S_i)$
1	1, 4, 10, 11	1	1
2	2, 3, 4, 10, 11	2	2
3	3, 4, 10, 11	2, 3, 8	3
4	4, 10, 11	1, 2, 3, 4, 8	4
5	5, 9, 10, 11	5	5
6	6, 9, 10, 11	6	6
7	7, 9, 10, 11	7	7
8	3, 4, 8, 10, 11	8	8
9	9, 10, 11	5, 6, 7, 9	9
10	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	10
11	10, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 11	11

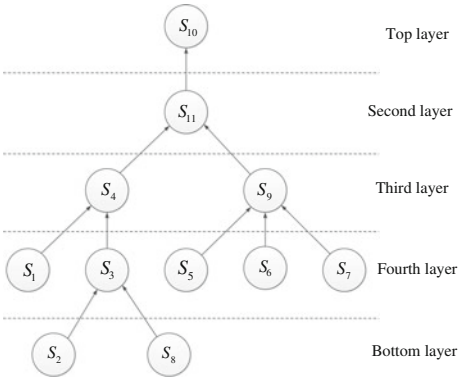


Fig. 36.4 Main factors of interpretative structural modeling

vestibule. On the one hand platform management score influences the number of boarding and alighting passengers by guidance and management of passenger flow, on the other hand the capacity of vestibule is limited and it is advantageous to control the number of boarding and alighting passengers.

PSDs and the distance between passengers and platform exit in the fourth level have direct influences on platform management score. Moreover, door width and the Pythagorean distance between the car and platform in the fifth level directly impacting on the distance between passengers and platform exit. Besides door width, the distance between door centerlines and standback width in the fourth level directly affect the capacity of vestibule. Therefore, it can be seen that these factors mainly fall into two aspects of train-based and station-based to affect the number of passengers in the vestibule, simultaneously restricting the efficiency of passengers boarding and alighting movement.

36.5 Conclusions and Future Work

The paper conducts field research to analyze the composition of factors impacting passenger boarding and alighting movement and the main factors affecting passenger boarding and alighting by ISM, then it finds direct and deep reasons influencing passenger boarding and alighting. Moreover, the article determines the impact of train-based and station-based factors on the rate of number of boarding passenger and number of alighting passenger, and it presents measures to improve the efficiency of passenger boarding and alighting movement based on both the train design and station facilities. As a result, the future work is to quantify these factors and try to find a function of the number of boarding passengers/number of boarding passengers and the factors. Finally it can determine the influence level of factors.

Acknowledgments This work is supported by National Key Technology R&D Program (NO. I13B300020). Yafei Wang is the corresponding author.

References

1. Harris NG, Anderson RJ (2007) An international comparison of urban rail boarding and alighting rates. Proc Inst Mech Eng, Part F J Rail Rapid Transit 221(4):521–526
2. Railway Consultancy. The impact of stepping distance on train boarding and alighting times. Report for cross-rail project, 1996
3. Rosser J (2000) Alighting and boarding time algorithm for London stations. Technical note 94, Thameslink 2000 project
4. Daamen W (2004) Modelling passenger flows in public transport facilities. Delft University of Technology, TU Delft
5. Harris NG, Graham DJ, Anderson RJ et al (2014) The impact of urban rail boarding and alighting factors. TRB 9nd3 2014 annual meeting
6. Xu T (2012) Study on risk assessment theory and methods on urban rail transit network operation. Beijing Jiaotong University, Beijing (in Chinese)
7. Yin H, Xu L (2010) Research on influencing factors of road networks vulnerability based on interpretive structural model. Soft Sci 14(10):122–128 (in Chinese)
8. Sun H, Zhou Y (2010) An analysis of the influencing factors of public transportation passenger flow on the base of interpretive structural model. J Beijing Inst Technol (Soc Sci Ed) 12(1): 29–34 (in Chinese)

Chapter 37
A Prediction Model of Security Situation Based on EMD-PSO-SVM

Xiaowen Yao, Fuge Wang and Yong Zhang

Abstract A security situation prediction model is proposed to forecast security situation of rail station based on Empirical Mode Decomposition (EMD), Support Vector Machine (SVM), and Particle Swarm Optimization (PSO). First, the basic concepts of EMD, SVM, and PSO are introduced. Second, EMD is used to decompose the original security situations data into several IMFs. Then, PSO-SVM model is used to forecast each IMF in which PSO is used to optimize the parameters of SVM. Finally, the sum of each IMF's forecasting result is the final prediction. The experimental results show the efficiency of the presented method in security situation prediction.

Keywords SVM · EMD · PSO · Security situation prediction

37.1 Introduction

The prediction of security situation about rail transit station is a new security technology that can be used to analyse the security status and security trends of the station. In recent years, the security situation prediction has been widely studied. There are some typical prediction algorithms, such as time series forecasting method, neural network forecasting method, gray model and Support Vector Machine (SVM). Fan [1] proposed the prediction of the rail train security situation based on SVM and succeeded to forecast the situation that the train passes curve. A security situation prediction algorithm based on radial basis function was proposed by Ren [2]. The algorithm can be used to predict the value of network

X. Yao · Y. Zhang (✉)
School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China
e-mail: 34445721@qq.com

F. Wang
School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_37

security situation by analysing nonlinear relationship between the values of security situation. Xuan et al. [3] proposed the method of network security situation prediction based on the combination of quantitative and qualitative cloud, and realized the prediction of future development of the situation.

In this paper, Empirical Mode Decomposition (EMD) and Particle Swarm Optimization (PSO) are applied to the SVM model to establish the security situation predictable model. The security situation value of Chen Jiaci station in Guang Zhou metro is selected as the experimental data in 365 days to validate the proposed method. The EMD-PSO-SVM model is expected to be more accurate in security situation prediction.

37.2 The Basic Theory

37.2.1 Empirical Mode Decomposition (EMD)

Huang et al. [4] proposed EMD in 1998. Which is a generally nonlinear, non-stationary data processing method EMD method can depict the non-stationary signal trends, cycles and fluctuation effectively. So, it has been widely cited in the field of mechanical vibration signal detection and fault analysis. EMD is used to decompose the original security situations data into several IMFs and IMFs meet the following conditions:

- The number of extreme points should be equal to the number of zero crossings, or differ at the most by one;
- For each data points, upper and lower envelopes should be symmetrical.

The basic steps of the EMD are shown as follows. First, find all extreme points of the signal $x(t)$ which include maximum and minimum values, and then generate its upper and lower envelopes with cubic spline curve fitting. Second, calculate the mean envelope $m(t)$ according to upper and lower envelopes point-by-point. Third, get the difference between $x(t)$ and $m(t)$ and the first part $c_1(t)$ can be acquired. Repeat above steps until the remainder $r_n(t)$ meets some stopping criterions. The original security situation series can be expressed as the sum of IMFs and a remainder $r_n(t)$:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{37.1}$$

37.2.2 Support Vector Machine (SVM)

SVM [5] is a machine learning method based on statistical theory. The basic idea is transforming the input space into a high-dimensional space by using the nonlinear change which is defined by inner product function, and then a nonlinear relationship is found between the input and output in the high-dimensional space.

For training data (x_i, y_i), $i = 1, 2, \ldots, L$, x_i is the input values, y_i is the output values, L is the number of training data. SVM looks for a nonlinear mapping f from

input space to the output space, and maps data x into a high-dimensional feature space, then gets the nonlinear regression result in the original space by doing linear regression in the high-dimensional feature space. The equation is as follows:

$$f(x) = \omega\phi(x) + b \tag{37.2}$$

In Eq. (37.2), ω is the weight vector, b is threshold, and $\Phi(x)$ is the nonlinear mapping from input space into high-dimensional feature space.

The slack variables ξ_i and ξ_i^* are introduced as Eq. (37.3), in order to determine the coefficient ω and b,

$$\min R = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{37.3}$$

$$\text{s.t.}\begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i, \\ f(x_i) - y_i \le \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, n. \end{cases} \tag{37.4}$$

The first half of the Eq. (37.3) is the minimizing structural risk, the second half is the minimizing sample error, the constant C is the compromise of structure risk and sample error, ε is used to define the linear insensitive loss.

It is very difficult to solve the Eq. (37.3) directly, because of the high dimension feature space and the non-differentiable objective function. However, it can be solved by SVM [6] in which the dot product kernel function $K(x_i, x_j)$ and Wolfe dual tips are introduced. In this paper, RBF kernel function is used, such as Eq. (37.5).

$$K(x_i, x_j) = \exp(-|x_i - x_j|^2/\sigma^2) \tag{37.5}$$

37.2.3 Particle Swarm Optimization (PSO)

Dr. Kennedy [7] was inspired by bird predation and proposed PSO algorithm. The PSO algorithm is described as follows: a community is composed by n particles in the D-dimensional search space, each particle i is represented by a D-dimensional vector and its state is represented by three vectors: $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$ indicates the current position of the search space; $V_i = (v_{i1}, v_{i2}, \ldots, v_{id})$ indicates the flight speed of the ith particle; $P_i = (p_{i1}, p_{i2}, \ldots, p_{id})$ indicates the best position of the ith particle after searching. In addition, $P_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$ indicates the entire PSO best position after searching. The optimal position p_i is recorded when searching the D-dimensional solution space. Each iteration, the position and speed of particles should be changed constantly to find the optimal position, and we use a fitness function to measure the advantages of the particles [8].

The following equations are used to update particle state in PSO algorithm.

$$v_{id}^{k+1} = w * v_{id}^{k} + c_1 r_1 (p_{id}^{k} - x_{id}^{k}) + c_2 r_2 (p_{g1}^{k} - x_{id}^{k}) \tag{37.6}$$

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \tag{37.7}$$

In the equation: i = 1, 2, …, N; d = 1, 2, …, D; ω is the inertia coefficient; Learning factors c_1 and c_2 are non-negative constants; r_1 and r_2 are random number between (0, 1); $v_{id} \in [-v_{\max}, v_{\max}]$.

37.3 EMD-PSO-SVM Prediction Model

37.3.1 EMD-PSO-SVM Prediction Model

The structure of EMD-PSO-SVM security situation prediction model is showed in Fig. 37.1.

First, EMD is used to decompose the original security situation series into several independent IMFs according to different frequencies. Second, each IMF is predicted by using SVM-PSO model in which PSO is used to optimize the parameters of SVM. Finally, the final prediction is obtained by accumulating each IMF's forecasting result [9].

37.3.2 PSO-SVM Prediction Model

The original security situation series are decomposed into several IMFs by EMD method. PSO-SVM is used to predict each IMF and get prediction of each part. The structure of PSO-SVM security situation prediction model is showed in Fig. 37.2.

The purpose of normalize IMF data is to get the model training data set and test data sets. SVM model is established with the PSO algorithm to optimize the

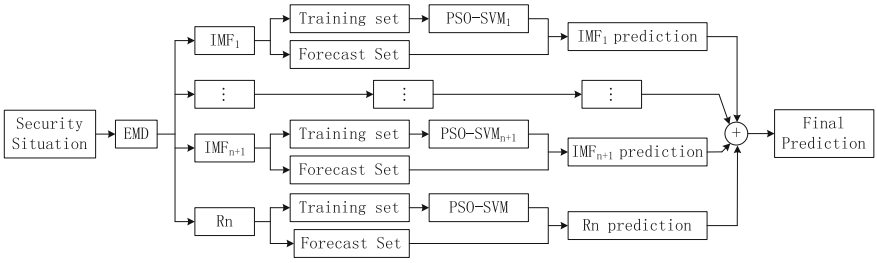


Fig. 37.1 EMD-PSO-SVM prediction model

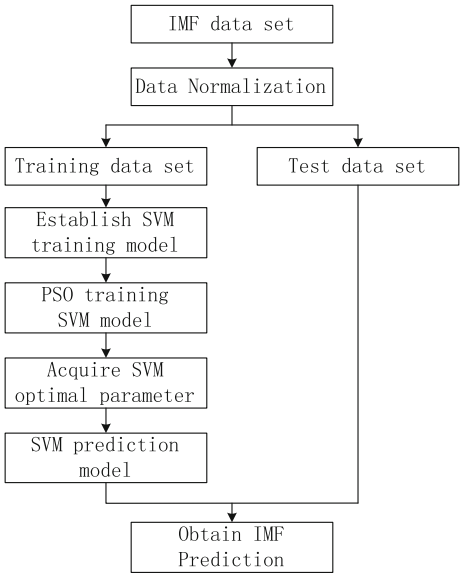


Fig. 37.2 PSO-SVM prediction model

parameters of SVM by using training data set. According to the SVM prediction model, IMF prediction is obtained by using the test data set.

37.4 Experimental Validation

37.4.1 Evaluation Criteria

Mean square error (MSE), mean absolute error (MAE) and the correlation coefficient R are selected as evaluation criteria in order to evaluate the predictive performance of the model. In this paper, MSE can reflect the accuracy of the measurement data well. MAE can better reflect the actual situation of the prediction error. R represents the curve fitting situation between security situation value predicted by the model and the actual security situation value. The value R is closer to 1, prediction accuracy is higher.

37.4.2 Model Prediction

The security situation value of Chen Jiaci station in Guang Zhou metro in 365 days are selected as the experimental data which are used to be constructed the experimental sample sets whose capacity is 358 groups. The distribution of security

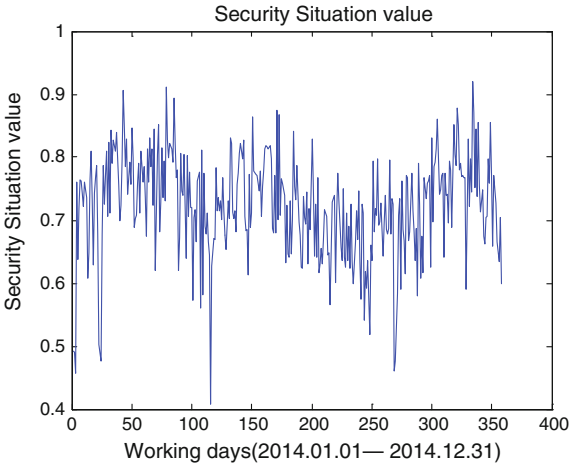


Fig. 37.3 The distribution of security situation value

situation value is showed in Fig. 37.3. The security situation is normalized and then used the EMD method to decompose it, shown in Fig. 37.4.

After EMD method acting on the security situation, six components and a margin sequence are obtained in Fig. 37.4. For each part, the PSO-SVM model is used to train and predict the data, and gets the predicted value of each part in the end. The absolute error of training and test sets are shown in Tables 37.1 and 37.2.

The prediction of each part is accumulated to get the final prediction. The comparison between the original data and the final prediction data is shown in

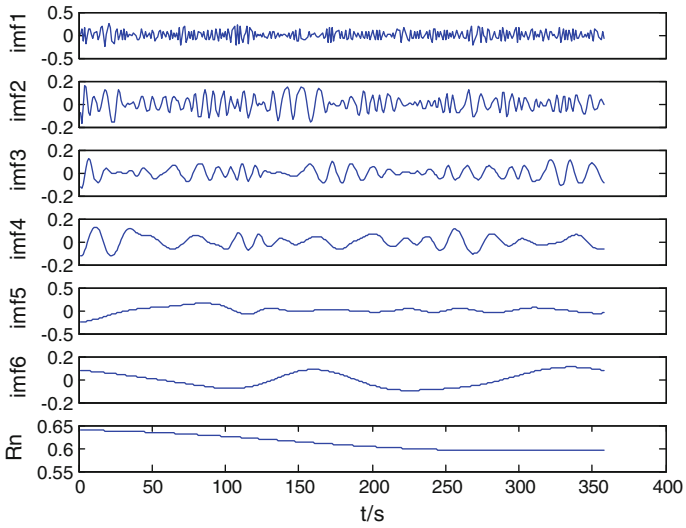


Fig. 37.4 The security situation is decomposed by the EMD method *t*

Table 37.1 The absolute error of training sets

Error	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	Rn
The max error	0.2908	0.1119	0.0373	0.1065	0.2615	0.0935	0.0178
The min error	−0.2544	−0.0977	−0.0128	−0.1205	−0.1502	−0.1198	−0.0243
Average absolute error	0.0513	0.0169	0.0090	0.0393	0.0515	0.0365	0.0089

Table 37.2 The absolute error of test sets

Error	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	Rn
The max error	0.2073	0.1518	0.1141	0.0898	0.0998	0.0564	0.0235
The min error	−0.2267	−0.1112	−0.1091	−0.1120	−0.0999	−0.1351	−0.0210
Average absolute error	0.0655	0.0503	0.0483	0.0377	0.0371	0.0585	0.0102

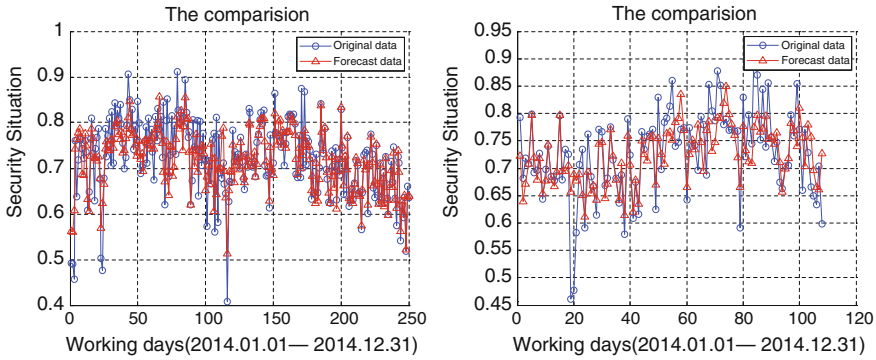


Fig. 37.5 The comparison between the original data and the final prediction data

Fig. 37.5. The correlation curve between the security situation predicted by model and the real data is shown in Fig. 37.6. The R of the training sets is 0.8218 and the test sets are 0.7914. The comparison of results between training and test sets based on EMD-PSO-SVM model is shown in Table 37.3. The EMD-PSO-SVM model is expected to be more accurate and feasible in security situation prediction.

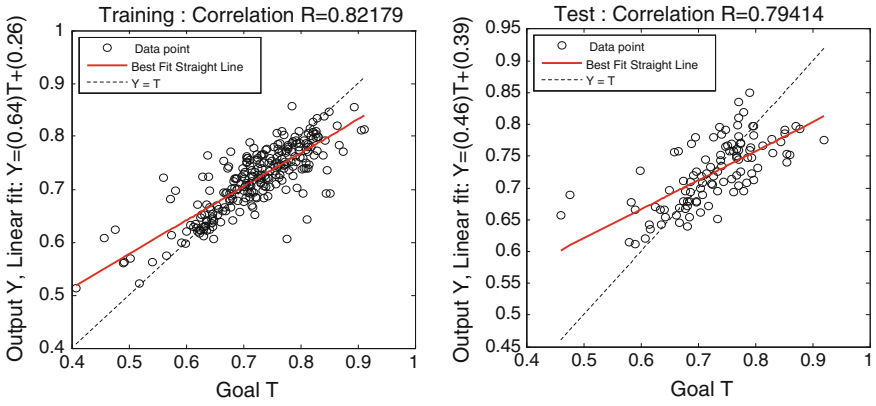


Fig. 37.6 The correlation curve between the security situation predicted by model and the real data

Table 37.3 The comparison of results between training and test sets

Simulation type	MAE	MSE	R
Training data	0.0329	0.0022	0.8218
Test data	0.0414	0.0032	0.7941

37.5 Conclusion

An EMD-PSO-SVM model is established which is used to predict the security situation in this paper. First, EMD is used to decompose the original security situation series into several independent IMFs according to different frequencies. Second, PSO-SVM is used to forecast each IMF in which PSO is used to optimize the parameters of SVM. Finally, the sum of each IMF's forecasting result is the final prediction. The result shows that the final prediction value and the actual value are consistent. Therefore, the EMD-PSO-SVM model is more accurate and feasible in security situation prediction.

Acknowledgments This study was the Fundamental Research Funds for the Central Universities (AE89454), and the Science and Technology Program of Guangzhou (201508010010). The author gratefully acknowledges the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially.

References

1. Fan H (2011) Rail security situation awareness research. Beijing Jiaotong University graduate thesis, Beijing (in Chinese)
2. Ren W (2007) Intelligent research of network security situation assessment. MS thesis, Shanghai Jiaotong University, Shanghai (in Chinese)

3. Xuan L, Hao S, Zhang Q (2010) Prediction of network security situation based on cloud. In: Proceedings of international conference of China communication and technology (ICCCT2010) (in Chinese)
4. Huang NE, Shen Z, Long SR (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. The Royal Society A: mathematical, physical and engineering sciences, pp 903–995
5. Zhang X (2000) Introduction to statistical learning theory and support vector machines. Acta Automatica Sinica 26(01):32–42 (in Chinese)
6. Bernhard S, Burges CJC (1999) Advances in kernel methods: support vector learning. The MIT Press, London
7. Kennedy J (2010) Particle swarm optimization. Encyclopedia of machine learning. Springer, US, New York, pp 760–766
8. Xu H, Wang X, Liao Y et al (2010) A new approach for optimizing the model of RBF-SVM based on PSO. Control Decis 25(3):367–370 (in Chinese)
9. Wang W, Zhao H, Liang Z, Tao M (2012) Hybrid intelligent prediction method based on EMD and SVM and its application. Comput Eng Appl 48:225–227 (in Chinese)

Chapter 38
A Study on the Forecast Method of Urban Rail Transit

Yiyu Hou, Honghui Dong and Limin Jia

Abstract Simple and quick short-term passenger flow forecast is the basis of improving the efficiency of urban rail transit operation. This paper discusses the advantages and disadvantages of BP neural network in the short-term passenger flow forecast of urban rail transit and puts forward to optimize the BP neural network forecast model by genetic algorithm method. After analyzing the actual data of time and space characteristics of passenger flow in urban rail transit, this paper determines the forecast object of the section passenger flow. Then, this paper selects the BP neural network as a forecast model and makes optimization or adjustment focus on the rate of convergence, local optimum, and overfitting for BP neural network. Making short-term section passenger flow forecast by using the passenger flow data and comparing the forecast results with the actual results, the model can be verified feasibility. Through this method, the optimized BP neural network can forecast passenger flow after 30 min accurately and provide the basis for transit operation.

Keywords Urban rail transit · Short-term passenger flow forecast · Time and space characteristics · BP neural network · Genetic algorithm

38.1 Introduction

Short-term passenger flow forecast is an important evaluation index for operating condition of rail transit system. In order to find high precision nonlinear forecasting methods that fit for their own characteristics of rail transit system. Researchers at home and aboard put forward some methods of short-term passenger flow forecast

Y. Hou · H. Dong (✉) · L. Jia
Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, No. 3 ShangYuan Road, Beijing, HaiDian District, China
e-mail: hhdong@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_38

such as time series and neural network, and obtained many significant achievements and development, especially neural network model.

Aiming at training methods of neural network, M. Clelland and Rumelhart put forward BP neural network algorithm. After comparing various short-term forecasting models, Karlaftis, Goliasand, and Vlahogianni pointed out that neural network has the best potential for development [1]. Tsung-Hsien and Trai put forward parallel integration neural network model (PENN) and multi temporal order neural network model (MTUNN), and apply it into short-term passenger flow forecast of railway [2]. Ping Chen amends nearest neighbor-clustering algorithm, and apply it into short-term traffic flow forecast to improve forecast accuracy and optimize matching results. Tao Ding and Huicheng Zhou improve reasonable degree of center amount and position of radial basis function by means of cross iterative fuzzy clustering algorithm.

The purpose of this paper is to find short-term passenger flow forecast that accord with rail transit passenger flow characteristic and the difficulties is whether the optimized BP neural network model has the feasibility. According to the above contents, this paper analyzes the traffic flow characteristics of rail transit, using the optimized BP neural network model which adapts to the characteristics well for short-term passenger flow forecast. Then this paper analyzes the feasibility and optimization effect of the model.

38.2 Passenger Flow Characteristic of Urban Rail Transit

In the research, it is obvious that urban rail transit has obvious time and space distribution characteristics. The collection of historical traffic data is the basis of the forecast. Understanding the time and space characteristics of passenger flow thoroughly or not can determine the accuracy of passenger flow forecast.

38.2.1 Analysis of Passenger Flow Time Characteristic

38.2.1.1 Per Hour of Passenger Distribution Characteristic

Through the change of daily passenger flow within the operating days, Per hour of station passenger flow distribution can be divided five types as follows: ① Single peak type; ② Double peak type; ③ Full peak type; ④ Sudden peak type; ⑤ No peak type.

The stations on the road network gradually produce differences, because of its different structure location in urban planning and road network. The line 4 runs through the north and south of the Beijing urban rail network, which covers all the station types, has great significance for study.

38.2.1.2 Per Week Passenger Distribution Characteristic

In the investigation of relevant literature [3], as for the overall road network, Weekend Passenger flow is only 76.5 % of the working days, and Sunday slightly lower than Saturday, it shows that the proportion of commuter passenger flow is higher. In addition, the common passenger flow level of the weekend is not less than the working day, but the passenger flow distribution equilibrium, no obvious morning or evening peak. Weekday peak passenger flow accounted for more than 40 %, peak hour factor for an average of 12 %. In the course of the study, according to the source of road network pressure, the forecast and analysis of the passenger flow of the working days should be emphasized.

38.2.2 Analysis of Passenger Flow Space Characteristic

Because of the scale of construction and the form and direction of the route layout, it is easily caused unbalanced distribution of passenger flow. Different types of urban rail transit routes are mainly divided into four space distribution forms as follows: ① Full line equality; ② Reduction of both ends; ③ Middle sudden increase; ④ Gradually reduce. The form of passenger flow distribution on the line 4 has the characteristics of ② and ③ at same time.

As shown in Table 38.1, the maximum passenger flow section at different time is different. Considering the passenger flow throughout the day, there are two main passenger flow sections. The numerical and the section of the maximum section passenger flow is the important standard in urban rail operation and management. During the evening peak, the maximum passenger flow section is the Zoo-National Library, the two stations next to the section are transfer station and even transfer hub. In this paper, the content will be studied for this section.

Table 38.1 2013 working days each time maximum section passenger flow of the line 4

Time slot	Section	Passenger flow	The proportion (%)
13:00–14:00	Caishikou → Xuanwumen	11,734	5.71
14:00–15:00	Caishikou → Xuanwumen	11,351	5.77
15:00–16:00	Caishikou → Xuanwumen	10,208	5.39
16:00–17:00	Caishikou → Xuanwumen	11,733	5.29
17:00–18:00	Zoo → National Library	16,296	5.15
18:00–19:00	Zoo → National Library	16,221	5.25
19:00–20:00	Zoo → National Library	8100	4.92

38.3 BP Neural Network Forecast Model

In this paper, the AFC data is extracted from per 5 min, and 30 min is the forecast interval for short-term passenger flow forecast, then the BP neural network is constructed.

38.3.1 Overview BP Neural Network

The topology of the BP neural network is the input layer, one layer or the multi-layer of hidden layer and the output layer. In the three layers structure, the input signal will according to the hierarchical structure of the forward transfer, error correction is the reverse communication. Through these two directions processes the network learning and training steps were completed [4, 5].

38.3.2 Constructing BP Neural Network Model

38.3.2.1 Selection of Transfer Function

Due to the superiority of Tan-sigmoid transfer function in the nonlinear and speed of convergence, we choose it as transfer function between the input layer and the hidden layer. In addition, the Purelin function can enlarge or decrease input value and maintain input value, so it is chosen as the transfer function between the hidden layer and the output layer.

38.3.2.2 Selection of Training Function

The neural network for urban rail passenger traffic forecast just belongs to small or medium network. In the process of short-term passenger flow forecast, timely and effective judgment is very important, so it is necessary to run faster rate and reduce memory consumption, at the same time in the choice of training function must also take into account the time varying and nonlinear. To meet the above requirements, the Trainscg training function the best choice for this paper.

38.3.2.3 Setting Parameters

The BP network of the single hidden layer can successfully approximate any continuous function within the interval. And the less hidden layers can improve the network forecast speed, so the hidden layer number can be determined as only one for the short-term passenger flow forecast of the line 4.

In this experiment, there are 5 input dimensions and 1 output dimension, according to the empirical formula, the node number should be between 5 and 13. After the observation of the 1500 training cases, the number of the hidden layer nodes can be determined at 10.

After using the Trainscg training function, the following parameters need to be set [6]:

1. The maximum training times (epochs): Because of the low demand for forecasting accuracy, and need the fast speed of convergence, it is set to 1000.
2. The training accuracy (goal): Considering that this paper can accept slightly higher forecast error, according to the metro traffic volume, it is set to 0.002.
3. Learning rate (LR): To ensure the convergence of the network is stable, the smaller learning rate is often selected. At the same time, the learning rate should be slightly more than the training accuracy, so it is set to 0.01.

38.4 Design Optimization and Forecasting Example

Based on the built model, we will further optimize the network structure and simulated analyze the actual date, and we make simulation analysis based on actual data.

38.4.1 Data Analysis

This paper extract Zoo-National Library, upper reaches of Xizhimen-Zoo section flow and leaving and entering flow of Zoo station in AFC data during 15 min. To ensure specific network input dimension, the correlation coefficient between the section flow data and other data still need to be calculated [7, 8].

The resulting value is mapped to a fixed band with maximum and minimum value method, and input data is in [−1, 1] interval, thus normalization of samples is completed. By using SPSS this paper get the correlation coefficients between many sets of data.

As shown in Table 38.2, there is a strong correlation between the six groups of data. Therefore, BP neural network model for forecasting passenger flow has five input dimensions in this experiment. There are: three connected section flow data in

Table 38.2 Analysis results of correlation between each set of data

Correlation coefficient	Forecast period	Previous period	Previous section	Latter section	Entering flow	Leaving flow'
Forecast period	1.000	0.964	0.877	0.919	0.790	0.944

the previous period and in and out section data of beginning station. Forecasting period section flow data is the only output dimension.

38.4.2 Model Optimization Based on Genetic Algorithm

In this article, we need to solve two problems, slow convergence and partial optimal solution, to get better results. So, we consider using the global searching feature of the genetic algorithm to optimize the network (Fig. 38.1).

Through using the GAOT toolbox main function we assign the GA optimized weights and threshold to previous BP network and forecast the passenger flow. The speed of program convergence gets faster practically.

As shown in Table 38.3, BP neural network get better in performance through genetic algorithm optimized. But it is not obvious because the structure of BP neural network passenger flow forecasting model is simple. The improvement of performance is mainly reflected in the cut of training step, and there is still some room to improve the stability of network structure.

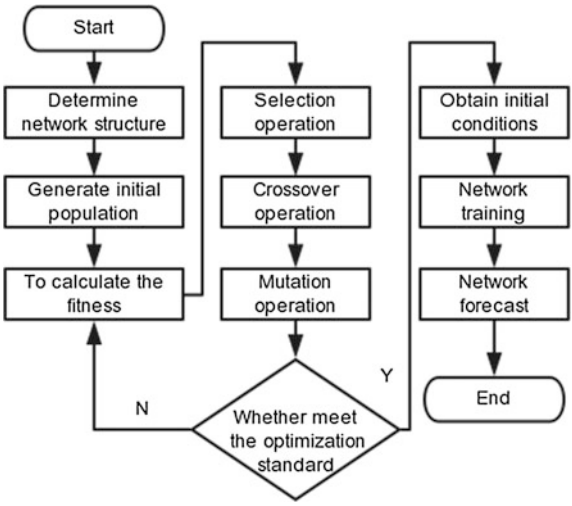


Fig. 38.1 GA-BP neural network process chart

Table 38.3 Comparison of the results before and after optimization

GABP step	BP step	GABP accuracy (%)	BP accuracy (%)
382	679	7	9
245	724	8	10
353	752	6	8
371	496	6	8

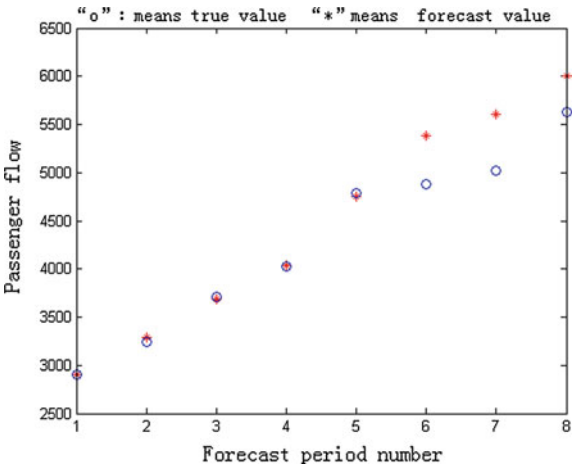


Fig. 38.2 Value and forecast value contrast chart

38.4.3 Simulation Results

Figure 38.2 is the results of the training simulation model.

In Fig. 38.2, the first four points using the training data as the input and output value, so the data fitting degree is extremely high. Although the predicted value of the next four points increases with the real value, there are obvious errors about 5 % when the accuracy is acceptable, and it is the fitting phenomenon of the network. In the course of the short time passenger flow forecast of the urban rail transit, it needs more passenger flow data to mitigate the phenomenon of overfitting. Further experiments are also needed to improve the selection of parameters in the network structure.

38.5 Conclusion

First, this paper analyzes the time-space characteristics of passenger flow in urban rail. And then, it chooses the training and transfer function base on the characteristics, set the network structure parameters, and write simple BP neural network program. In addition, aiming at the inherent defects of BP neural network, it optimize the network structure of defects, and improve the accuracy and training speed of forecast model through the combination of BP neural network and genetic algorithm. In the end, the passenger flow data and BP neural network is used to predict the passenger flow of the future and the feasibility of the model is also verified by simulation.

There are still many problems and limitations in this paper. In the process of further research, I am looking for the optimal fit between the learning ability and generalization ability help to solve the problem like overfitting. At the same time,

this paper hope to take more passenger flow data as input dimension, such as passenger data of the same period of the previous week, these can greatly improve the accuracy of forecast, and improve the network structure.

Acknowledgments This work was supported by Beijing New-star Plan of Science and Technology (Z1211106002512027).

References

1. Yansuo S, Ruihua X (2011) Estimation of rail transit passenger route using automated fare collected data. Traffic Transp 12:85–90 (in Chinese)
2. Tsai T, Lee C, Wei C (2009) Neural network based temporal feature models for short-term railway passenger demand forecasting. Expert Syst Appl 36(2):3728–36
3. Jing W, Jianfeng L et al (2013) Temporal and spatial passenger flow distribution characteristics at rail transit stations in Beijing. Urban Traffic 11(6):18–27 (in Chinese)
4. Weisheng D (2013) The research of short—time passenger flow forecasting based on improved BP neural network in urban rail transit. Beijing Jiaotong University, Beijing (in Chinese)
5. Minxin T (2014) The research of short term passenger flow prediction based on neural network in city rail transit. Comput Knowl Technol Acad Exch (2):809–812 (in Chinese)
6. Guo J (2005) Study on the algorithm of BP neural network. Huazhong University of Science and Technology, Wuhan (in Chinese)
7. Xingjie S, Changmei H (2011) Calculation and application of section passenger flow of nanjing metro AFC system. Commun Audio Video 1:31–36 (in Chinese)
8. Yanshuo S, Ruihua X (2011) Estimation of rail transit passenger route using automated fare collected data. Traffic Transp 27(B12):85–90 (in Chinese)

Chapter 39
Reinvestment Strategy-Based Project Portfolio Selection and Scheduling with Time-Dependent Budget Limit Considering Time Value of Capital

Baolong Wang and Yuantao Song

Abstract The selection and schedule of right project portfolio is an important part of successful project management. Considering annual budget and time value of capital, a mathematical optimization model is proposed to address the problem of simultaneously selecting and scheduling a suitable project portfolio among available projects based on reinvestment strategy. For practical situations, annual budget is assumed time-dependent and time value of capital is also integrated into the model. Through reinvestment strategy, the profit generated by completed projects can be reinvested to finance other projects to maximize total profit of the portfolio. Additional constraint of interrelationship among projects, such as partial exclusiveness is also incorporated into the model to represent closely specified problems. Finally, an experiment study is presented to validate the model capability for dealing with optimal selection and schedule of project portfolio under considerations.

Keywords Project interrelationship · Project portfolio selection and scheduling · Reinvestment strategy · Time-dependent budget · Time value of capital

39.1 Introduction

With the implementation of project management in most organizations, more operations or business activities, such as the R&D of new products, quality improvement, technological innovation, etc., are considered as a complete project to manage. Single-project management plays a crucial role in the business operations of the organizations in the past decades. However, considering the economic globalization and the increasing competition, many enterprises should grasp the

B. Wang · Y. Song (✉)
School of Engineering Science, University of Chinese Academy of Sciences (UCAS), No. 19A Yuquan Road, 100049 Beijing, China
e-mail: songyuantao@ucas.ac.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_39

dynamic opportunity and make quick response to maintain sustainable advantage and healthy growth by simultaneously managing multiple projects, which makes related project management to become more complex. The national project of large-scale construction in China is often not a single project but a program or portfolio. For instance, the municipal facilities for the 2008 Beijing Olympic Games contains a number of venues, roads, Olympic Village and others; the South–North Water Diversion of China is divided into the main project, supporting projects and pipeline projects, with each category including varieties of subprojects. There are usually a large amount of available projects or investment opportunities for decision makers in the early stage of project management, but financial and physical limits may not support all alternatives simultaneously. Effectively selecting and reasonably scheduling project portfolio, aligned with strategic objectives has always been the focus of project management [1]. Given that only 26% of information technology projects are completed on time within the budget [2], this will greatly affect the achievement of the long-term strategy, which makes portfolio management theory and its practice an active research field in project management.

39.2 Literature Review

Project portfolio is an array of projects, gathered together for organizations' strategy or other business objectives, which share same resource in a certain time period, among which interrelationship may exist. Project portfolio selection and scheduling problem (PPSSP) has been widely studied and applied in many fields, such as project management, financial management and investment [3]. The PPSSP can be described as follows. It is a process about selecting an optimal project portfolio, aligned with the strategy, by analyzing and evaluating each project and then scheduling them to make them implemented and completed within the given time horizon. These projects may differ in the volume of required resource, duration, annual profit after completion and the interrelationship. With adequate resource, any project can start at any moment. However, when the budget is limited and just available for implementing several projects, the return generated by completed projects can be reinvested to finance more projects to start. So, the order of project implementation does matter, and the selection and schedule process interact with each other.

Reference [4] first introduced the concept of portfolio and applied it in the field of financial investment to assess the earnings and risk of investment assets, laying the foundation of modern portfolio. Then, further researches about the portfolio were conducted by later scholars. McFarlan integrated the concept of portfolio with selection and management of projects, where better choice may be made by separately considering the individual and overall risk of projects, and thereby established the decision mechanism of project portfolio [5]. So far, the theory of project portfolio has achieved a leap of progress. Based on basic project portfolio,

numerous extensions or variants have been developed, such as stochastic, deterministic, and static or dynamic environments [3].

Many procedures have also been used to evaluate and select projects, such as analytical hierarchy process and multi-attribute decision making methods [6]. As PPSSP becomes more complex, like the introduction of renewable resource with limited capacity and multiple objectives, mathematical programming models become another interest. For example, a multi-objective model based on Lean and Six Sigma methodologies was developed for selecting projects [7]. However, the lack of resource usually leads to the impossibility of simultaneous commerce for all projects, so the integration of selection and schedule better treats these problems. Reference [8] introduced a comprehensive framework which incorporated the efficient selection, optimum schedule and the project interdependence. A mixed integer programming model for selecting and scheduling projects was proposed by [9]. Based on selected projects and limited resource, the corresponding schedule problem is usually called resource constrained project scheduling problems (RCPSP), which have attracted considerable attention for decades. As a generalization of job shop, the introduction of resource capacity makes RCPSP a non-polynomial hard (NP-hard) problem in strong sense [10]. The traditional planning methods, like CPM and PERT, could not well deal with the RCPSP. And then, classic RCPSP of scheduling a set of activities of a project to minimize the completion time has been extended to more general problems with multiple projects, preemptive activities and nonrenewable resource [10, 11]. Preference [12] provided a scheme to classify RCPSP by using a unifying notation. Since all ongoing projects share same resource, the allocation of limited resource plays a crucial role in resolving the RCPSP. Techniques from mathematical programming techniques to flexible heuristic procedures, such as zero one programming, priority rule-based heuristics and metaheuristics have also been summarized to solve the RCPSP [2, 13].

However, considering that capital is a type of special resource and has special characteristics, like the feature of cumulation and reinvestment, the involved complexity under considerations is great. To our knowledge, few literature concerns the PPSSP with these features of capital. Although the cumulative resource, namely the resource is consumed at the beginning and reproduced after its completion, was introduced into the RCPSP [14], it is not involved with project selection. The same resource was also used to solve a batch production scheduling problem in process industries [15]. Reference [16] combined cumulative resource with multi-mode RCPSP and considered a project in the automotive industry, but project selection is not still mentioned. Moreover, the reinvestment of capital is not considered in these papers. Reference [17] used reinvestment strategy to invest the profit of completed projects in other projects, but it assumed all initial budgets to be available at first and did not consider annual budget reaching by multiple periods. However, this assumption might be too strict in some cases, like changing availability of budget for each year. In addition, the time value of capital is not well treated with. In practice, the initial capital is not in place at a time, i.e., a certain amount of capital is available each period. Meantime, the time value of capital will influence the

investment cost, which implies that the investment cost will change if a project does not start immediately. Even though the reinvestment strategy is generally used in the field of finance investment, like equity security and fund, it still widely exists in other situations, like investment decision. So, it is advisable and necessary to consider reinvestment strategy when solving the PPSSP under considerations.

In this paper, we analyze PPSSP subjected to annual budget, which is time-dependent, to maximize overall profit of the portfolio. Besides, the reinvestment strategy is included in the model, so the execution and completion of projects in prior period will affect the selection and schedule of projects in later period. In addition, the time value of capital is also considered. To represent closely the practice, the partial exclusivity among projects is also integrated, where the relationship of partial exclusivity means that at most one of each pair with partial exclusivity could be selected because of extremely similar technology or function. Moreover, the deadline is also set to force the portfolio finish with the planned time horizon.

39.3 Problem Formulation and Model

Suppose that an organization, like a corporate, needs to select a suitable project portfolio from a pool of available projects to realize the strategy. A series of budget is provided during the specified time intervals, such as the five-year plan of the government in China. Limited annual budget does not guarantee simultaneous implementation of all alternatives, i.e., even though some projects have been chosen, it may be unable to execute them simultaneously. So the PPSSP under considerations becomes a concern confronted by practitioners and scholars. Besides, a duration of return is assigned separately with each project upon its completion, during which annual profit may be equal or not. The initial capital is from the budget of first year and the balance of budget in previous periods, following budget provision and the return of completed projects make up the capital pool to finance other projects to execute.

For simplicity, we assume that the investment required by a project is a one-off put; the project is non-preemptive, i.e., the project continues until it is completed once commencing; the selected candidate will start at the beginning of each period; annual revenue is generated a year after the completion of the project.

Based on the above, a mathematical optimization model is established to optimize the total profit of PPSSP under considerations.

The number of available projects is assumed to be N ($i = 1, 2, \ldots, N$). The symbol of T_1 is the planned time horizon of completing selected projects. T_2 represents the time periods of budget occurring. M_j is used to denote the budget of period j, occurring at the moment $j-1$ ($j = 1, 2, \ldots, T_2$). H_i is used to represent the duration of return for project I and a are the discount rate. Besides, P_{ik} is applied to denote annual profit generated by project i during period k after completion if the project i immediately commences. If project i starts immediately, the corresponding

investment cost is Q_i. The duration of project i is D_i. We use symbol S_t to represent the set of launched projects before or on period t, i.e., $S_t = \{ i \mid 0 < Y_i \leq t \}$, and symbol F_t to represent the set of completed projects before the end of period t, i.e., $F_t = \{ i \mid Y_i - 1 + D_i \leq t \}$.

Let the decision variables be X_i and Y_i. X_i is a binary variable for project selection, $X_i = 1$ if the project i is included in the portfolio, otherwise $X_i = 0$; Y_i is an integer variable and is used to represent the start time of project i. If project i is selected, that is, $X_i = 1$, then $Y_i \geq 1$; otherwise, $Y_i = 0$. If project i is selected and will be launched at period t, it means that the project will commence at moment $t-1$.

Besides, we classify the interrelationship between projects as partial exclusivity (E). Accordingly, E denotes the set of project pair with partial exclusivity, from which at most one of each pair can be included in the portfolio.

An integer programming model for PPSSP under considerations is formulated as follows:

$$\text{Maximize } Z = \sum_{i=1}^{N} \left(Xi(\sum_{k=D_i+1}^{D_i+Hi} P_{ik} e^{-ak} - Q_i) \right) \tag{39.1}$$

Subject to

$$\sum_{i \in St} (X_i Q_i e^{a(Y_i-1)}) \leq \sum_{j=1}^{\min\{T_2,t\}} M_j + \sum_{i \in F_t - 1} \sum_{k=Y_i+D_i}^{t} X_i P_{ik} e^{a(Y_i-1)} \; t = 1, 2, \ldots, T_1 \tag{39.2}$$

$$X_i + X_j \leq 1 \quad (i,j) \in E \tag{39.3}$$

$$Y_i - 1 + D_i \leq T_1 \quad i = 1, 2, \ldots, N \tag{39.4}$$

$$Y_i = 1, 2, \ldots, T_1 \quad i = 1, 2, \ldots, N \tag{39.5}$$

$$X_i = 0, 1 \; i = 1, 2, \ldots, N \tag{39.6}$$

The objective function is illustrated as Eq. (39.1) and denotes the total profit of the portfolio. Equation (39.2) represents that during each period the cumulative investment required does not exceed the sum of available resource, including cumulative budget and total return generated by completed projects considering the time value of capital. Equation (39.3) presents the partial exclusiveness between each pair of projects from the set of E, i.e., at most one of project pair from the set of E can be included in the mixture. Equation (39.4) forces all selected projects to be completed within planned time horizon. Equations (39.5) and (39.6) shows decision variable Y_i is an integer and project i, if selected, must commerce within T_1, and X_i is a binary variable for project selection.

About Eq. (39.1), a simple proof is given: If project i commences at period Y_i, i.e., at the moment of Y_i-1, the investment required will amount to $Q_i e^{a(Y_i-1)}$.

Annual profit will be $P_{ij}e^{a(Y_i-1)}$ after its completion. By discounting the total profit to initial moment, the overall benefit $\sum_{j=Y_i+D_i}^{Y_i+D_i+H_i-1} P_{ij}e^{a(Y_i-1)}e^{-aj} - Q_i e^{a(Y_i-1)}e^{-a(Y_i-1)}$ will be simplified in the form of $\sum_{j=D_i+1}^{D_i+H_i} P_{ij}e^{-aj} - Q_i$, and the ultimate objective function is shown by (39.1).

Although the proposed problem belongs to the NP problems, it can still be successfully and optimally solved by standard optimization technology for some instances of interest for application. As shown in the model, the decision variables are Boolean and Integer; and the problem is not too complex. For instance, the number of total variables in the model equals 2 N and all of them are Integer, whereas the number of total constraints is not too much. Numerous commercial software for integer programming technique can deal with the problem under considerations. So, MATLAB and Lingo can be applied to solve the model, obtain global optimal solution and ensure its precision.

39.4 Experiment Analysis

In the section, the selection and scheduling of construction projects is taken an example to illustrate the application of the proposed model. Similarly, the R&D of new products or the transformation of production systems in manufacturing industry can also be used as another case of this paper.

Annual profit is assumed to be same during the profit duration, namely, $P_{ik} = P_i$, to simply the operation. The sample information of projects, including project cost, duration, annual profit and profit duration is listed in Table 39.1. In addition, let the

Table 39.1 Sample information

Project name i	Project cost($) Q_i	Project duration(y) D_i	Annual profit($) P_i	Profit duration(y) H_i
P1	12 M[a]	3	6 M	4
P2	1 M	4	5 M	5
P3	7 M	2	3 M	6
P4	4 M	1	2 M	4
P5	2 M	1	1 M	4
P6	5 M	2	2 M	6
P7	4 M	2	1 M	8
P8	8 M	2	3 M	7
P9	1 M	3	5 M	4
P10	3 M	1	1 M	6

[a]The symbol of M represents a million dollars

Table 39.2 Annual budget limit (*unit* Million Dollars)

Year	First	Second	Third	Fourth
Available budget	8 M[a]	9 M	8 M	9 M

[a]The symbol of M represents a million dollars

discount rate, *a,* be 10 %. And a deadline, $T_1 = 5$, is set to force all selected projects to be completed within the planned time horizon.

The capacity of annually available budget is given in Table 39.2, accordingly, $T_2 = 4$. Another constraint is the interrelationship among projects, where P5 and P6 are partially exclusive.

This section analyzes separately the scenarios with and without reinvestment strategy. We describe Scenario I as following: the optimization of overall profit of the portfolio using reinvestment strategy. Scenario II is defined as the same to Scenario I but without reinvestment strategy. We can analyze in detail the difference in ultimate solution by comparing the optimal results of Scenario I and Scenario II, like ultimate project portfolio, total profit and the change of cumulative profit.

Here, MATLAB and Lingo is used to solve the proposed model optimally. The optimal mixture and schedule of project portfolio for Scenario I is shown in Fig. 39.1, where P3, P4, P6, P8, P9, and P10 is included in the portfolio. Correspondingly, P6 commences at period $t = 1$; P4 at $t = 2$; P9 at $t = 3$; P3 and P8 at $t = 4$; P10 at $t = 5$ separately. For Scenario II, the optimal solution is displayed in Fig. 39.2, where P3, P4, P6, P8 and P10 are selected, and P4 commences at $t = 1$;

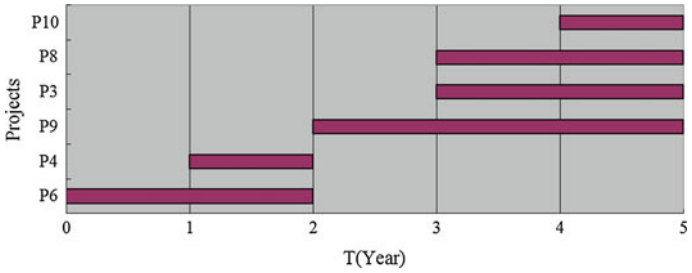


Fig. 39.1 Final selection and schedule for Scenario I

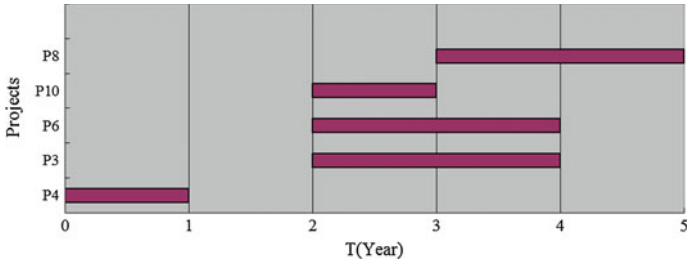


Fig. 39.2 Final selection and schedule for Scenario II

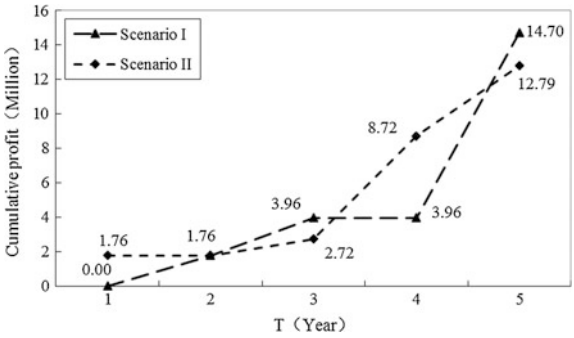


Fig. 39.3 Cumulative profit for both scenarios

P3, P6 and P10 at $t = 3$; and P8 at $t = 4$. More details, like the completion time can also be obtained. From the difference between Fig. 39.1 and Fig. 39.2, we can conclude that reinvestment strategy not only greatly influences the selection of the portfolio, but it also affects the schedule of the portfolio. For instance, P10 is selected in Scenario I and excluded in Scenario II. According to the results in Fig. 39.3, the maximum profit is \$14,697,000 for Scenario I and \$12,789,000 for Scenario II. The use of reinvestment strategy makes the total profit significantly increase by 15 %. The increased profit can also be regarded as the profit of executing additional project P10 through reinvestment strategy. Based on the same budget, reinvestment strategy finances project managers to execute more projects to optimize the total profit. So, it is reasonable and profitable to apply reinvestment strategy in project selection and schedule. Meantime, the efficiency of capital can also be enhanced accordingly.

39.5 Conclusion

This paper establishes an optimization model of simultaneously selecting and scheduling project portfolio through reinvestment strategy to solve combinatorial optimization problems in project management. Within the proposed formulation, annual budget is time-dependent, and time value of capital and interrelationship between projects are considered to represent closely realistic problems. As a kind of special resource, the capital is consumed at the beginning and reproduced after completion, so it can be reinvested to finance more projects while satisfying various restrain conditions. The optimization model is verified by considering two different scenarios involving ten projects. The optimal results of selection and schedule guarantee the optimization of objective. Two scenarios, with and without reinvestment strategy, are analyzed comparatively to demonstrate that reinvestment strategy can alleviate the shortage of limited budget, finance more projects and therefore increase the overall profit of project portfolio, enriching the theory and facilitating the practice of project management.

References

1. Archer NP, Ghasemzadeh F (1999) An integrated framework for project portfolio selection. Int J Project Manage 17(4):207–216
2. Chen J, Askin RG (2009) Project selection, scheduling and resource allocation with time dependent returns. Eur J Oper Res 193(1):23–34
3. Jafarzadeh M, Tareghian HR, Rahbarnia F et al (2015) Optimal selection of project portfolios using reinvestment strategy within a flexible time horizon. Eur J Oper Res 243(2):658–664
4. Markowitz HM (1952) Portfolio selection. J Finance 7(1):77–91
5. Mcfarlan FW (1981) Portfolio approach to information-systems. Harvard Bus Rev 59(5):142–150
6. Heidenberger K, Stummer C (1999) Research and development project selection and resource allocation: a review of quantitative modelling approaches. Int J Manage Rev 1(2):197–224
7. Hu G, Wang L, Fetch S et al (2008) A multi-objective model for project portfolio selection to implement lean and six sigma concepts. Int J Prod Res 46(23):6611–6625
8. Carazo AF, Gómez T, Molina J et al (2010) Solving a comprehensive model for multiobjective project portfolio selection. Comput Oper Res 37(4):630–639
9. Medaglia AL, Hueth D, Mendieta JC et al (2008) A multiobjective model for the selection and timing of public enterprise projects. Socio-Econ Plan Sci 42(1):31–45
10. Hartmann S, Briskorn D (2010) A survey of variants and extensions of the resource-constrained project scheduling problem. Eur J Oper Res 207(1):1–14
11. Wu C, Wang X, Lin J (2014) Optimizations in project scheduling: a state-of-art survey. In: Wang X (ed) Optimization and control methods in industrial engineering and construction. Springer, Netherlands
12. Brucker P, Drexl A, Möhring R et al (1999) Resource-constrained project scheduling: notation, classification, models, and methods. Eur J Oper Res 112(1):3–41
13. Naderi B (2013) The project portfolio selection and scheduling problem: mathematical model and algorithms. J Optim Ind Eng 6(13):65–72
14. Neumann K, Schwindt C (2003) Project scheduling with inventory constraints. Math Methods Oper Res 56(3):513–533
15. Neumann K, Schwindt C, Trautmann N (2005) Scheduling of continuous and discontinuous material flows with intermediate storage restrictions. Eur J Oper Res 165(2):495–509
16. Bartels JH, Zimmermann J (2009) Scheduling tests in automotive R&D projects. Eur J Oper Res 193(3):805–819
17. Belenky AS (2012) A Boolean programming problem of choosing an optimal portfolio of projects and optimal schedules for them by reinvesting within the portfolio the profit from project implementation. Appl Math Lett 25(10):1279–1284

Chapter 40
Modeling of Jigsaw Game Software with a Map Using UML

Naiwen Wang, Wu Xie, Zhipeng Zhong, Huimin Zhang and Maoxuan Liang

Abstract In view of the problem that current jigsaw games have no functions of upgrades or accumulation scores, the UML (Unified Modeling Language) technology is adopted to analyze to design the mobile game software via spitting jigsaw puzzle pieces of a map. The results of software development illustrate that it is very easy to upgrade and accumulate scores. The gaps between design and coding are smaller using UML than before. Also, the software development period is shorter. Not only the material resources for entity toys can be saved greatly via this software, but also it is significant for children users to study the knowledge of map and traffic.

Keywords Jigsaw game · UML · Modeling · Map · Traffic

40.1 Introduction

Jigsaw puzzle game is a children activity that all piece parts of a picture (i.e., a map) are combined into a complete figure. By playing the game with a map, not only the intellectual for young children are exercised imperceptibly in the entertainment, but also children become strong with self-confidence little by little. This game plays an important role in learning related geography knowledge of maps and traffics from childhood. There are many researches about jigsaw puzzle game [1–8], yet the problem also exists since children like playing on new electronic devices.

The business modeling processes of irregular jigsaw puzzle are not described much in these literatures [1–9]. The jigsaw procedures are difficult to express using such traditional methods as block diagram, E-R (Entity-Relationship) diagram, etc.

N. Wang · W. Xie (✉) · Z. Zhong · H. Zhang · M. Liang
School of Computer Science and Engineering, Guilin University of Electronic Technology, No. 1 Jinji Road, Guilin, Guangxi, People's Republic of China
e-mail: xiewu588@guet.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_40

Especially in the traditional physical puzzle games, the childhood potentials are not easy to be stimulated further for lacking the designs of grades and scores. Children's thinking and activities are not easily described without object-oriented methods, keeping distances from software programming. The specific functions of children users and the system administrators are not easy to explain. All kinds of activities of the puzzle games are not clear. These have resulted in a problem that the software workflows of jigsaw puzzle games are too complicated to develop for no upgrades or accumulation scores.

Therefore, in order to solve this problem, the traditional methods are combined with the theory of modern-oriented object software engineering to model the system business processes, and the software is developed with the Android technology. The remainder sections of this paper are organized as follows.

In Sect. 2, the module function demands are analyzed statically using the use case diagrams in UML (Unified Modeling Language). The software system is modeled dynamically with the UML methods of sequence diagram, collaboration diagram and state diagram. The third section provides the results of the code framework class diagrams and the software implementations. In the fourth section, the advantages of UML technology to model the map puzzle game software are discussed. Finally, the brief conclusions of this paper are followed in Sect. 5.

40.2 Modeling of Game Software Using UML

In order to describe the specific functions of jigsaw puzzle game systems by the children users of different ages, the grades can be set according to the difficulty and time for the game software. Here, the users mainly include such young child as low-age children, teenagers or youth, etc. The grades for the children of different ages that can be surpassed are not identical. Thus, from the perspectives of different children, their functional requirements changes with ages for the software of puzzle game. It is necessary to utilize the UML technology to analyze the requirements of this software, and the use case diagram is shown in Fig. 40.1.

The objects of the jigsaw puzzle software include user layer, section layer, grade layer and game layer, etc. Children users enter the game interface, and then choose

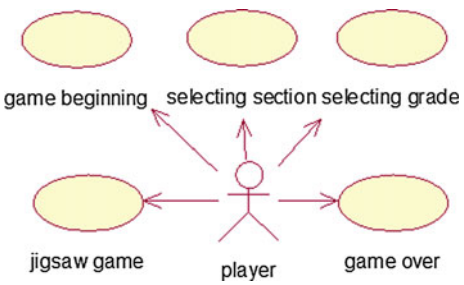


Fig. 40.1 The use case diagram of this game software as a whole

the section and grade before starting playing map puzzle games. At last, they have to exit the playing interface of this software to end the game when the time is up.

To clearly describe the time sequences during the processes of message sending and receiving among these software objects and show their dynamic interactions, the UML methods with the object-oriented technology are used to model these message sequences, and the corresponding sequence diagram is designed as Fig. 40.2 according to the events from the children players.

To clearly determine the message communication relationships among different children user objects where the information of the events needed to be illustrated, the collaboration diagram is conceived as Fig. 40.3. The logical business through sending and receiving messages among these objects are denoted with some Arabic numerals.

During playing the jigsaw puzzle game with a map, children always have to face the changing software interface, i.e., game states. The states include legal user interface, welcome interface, count-clock interface, starting game, playing game, pass, next grade, etc. Therefore, the state diagram is constructed to describe the relations of these states as Fig. 40.4.

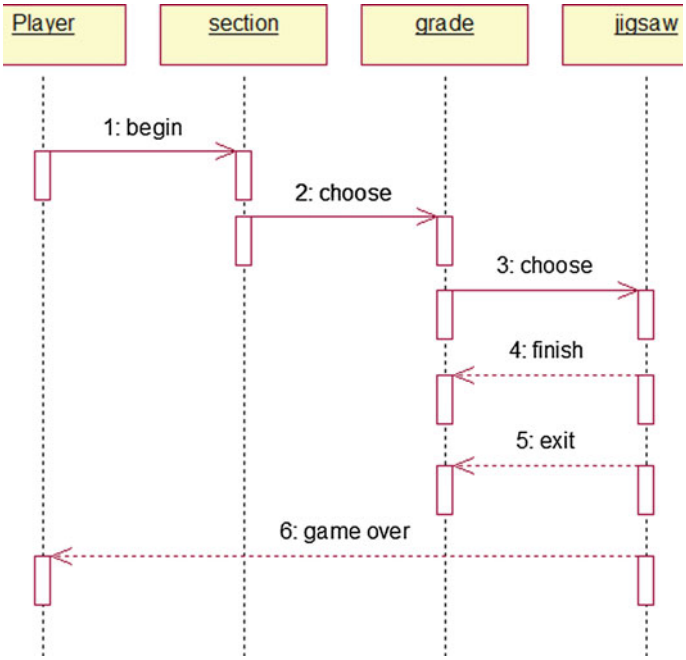


Fig. 40.2 The sequence diagram of this game software

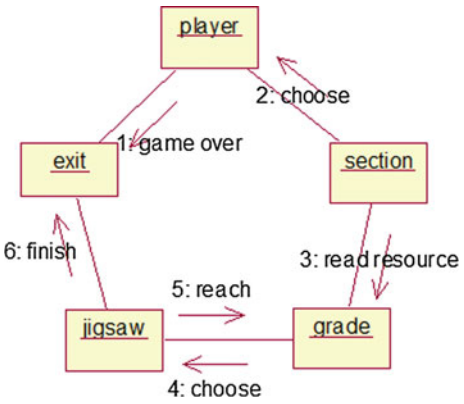


Fig. 40.3 The collaboration diagram of this game software

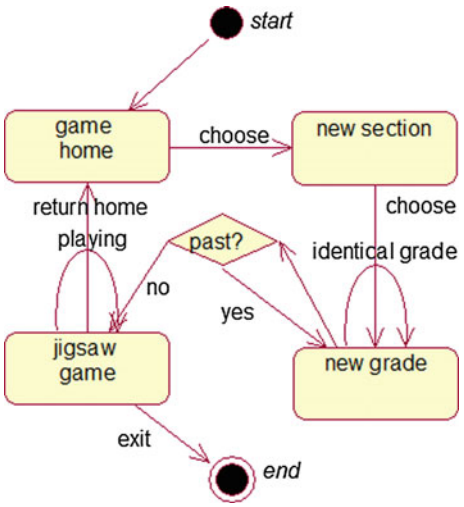


Fig. 40.4 The state diagram of this game software

40.3 Results

By using the technology of IBM Rational Rose, the requirement businesses of the software system are modeled by the software engineering theory with the use case diagram, sequence diagram, collaboration diagram and state diagram. According to these model diagrams, the class diagram for coding is obtained and shown as Fig. 40.5.

Using the Android technology with the tools of Java SE Development Kit (JDK) and Eclipse (Plug-in Development Environment, PDE), the software is implemented via Java programming based on the businesses of the use case diagram, sequence diagram, collaboration diagram and state diagram in UML.

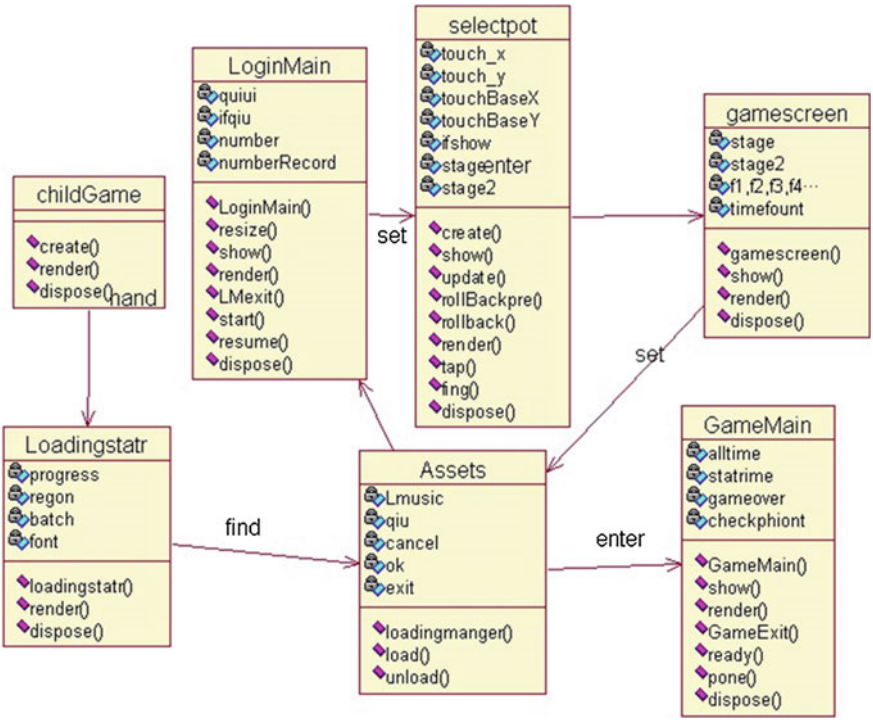


Fig. 40.5 The class diagram of this game software

Fig. 40.6 One of the main interfaces of this game software with a map

Based on the class diagram from the modeling processes of this jigsaw game software, one of the results of the mobile game with simple and friendly interactions is shown in Fig. 40.6.

40.4 Discussion

From the results of the class diagram and jigsaw game software with the Chinese map, the UML modeling processes have more advantages than before.

1. The function requirement businesses of this software are clearer for children users to deal with the puzzle games of the Chinese map using UML than via traditional software engineering approaches. The relevant complete concepts of an entity semantic game are not established by children users, which results in the complex puzzle businesses. However, the mobile puzzle shapes and the change statuses for the game are simplified via the UML models, and the businesses of playing puzzle games are close to the actual processes of hand-made jigsaws. It can be seen from the results above that clearer software models have been constructed by using the UML technology to produce the main system frameworks for the game software programming than ever.
2. The modeling results of the puzzle game software with maps are visual via UML, which is easier to understand for the system designer to exchange with children users. Children's understanding abilities are not strong, and it is difficult for the system designer to explain the software businesses of the puzzle game. Yet the UML modeling methods and processes of this software are intuitive via the use case diagram, class diagram, sequence diagram, collaboration diagram, and state diagram. Thus, children users can understand more easily than using the methods of text documents. It is easier for the designer to exchange with children users through the modeling processes than before.

40.5 Conclusion

In this paper, the jigsaw game software with maps has been modeled using the methods of UML, and the corresponding class diagram for coding of this system has been obtained via the IBM Rational Rose. By the Android tools, this game software has been developed with complete functions. The results show that children users are able to understand more clearly the functional requirement business of puzzles games via UML than before. Also, the system designers are more convenient to exchange with children users via the visual modeling results of the game software than such traditional approaches as text documents. The UML modeling of the jigsaw game software with maps for multiple operating systems needs to be further researched.

Acknowledgment This work is supported by the national college students' innovative training program project (201210595006), the teaching reform project (JGB201408, YJG201403, JGS201124) and the network teaching resources construction project (ZJW14004) of Guilin University of Electronic Technology, Guangxi Education Science "11th five-year plan" project (2010C059).

References

1. Syungog A (2010) Q-STAR game with a new number ordered cube puzzle game. J Knowl Inf Technol Syst 5(2):133–139
2. Peter K, Samuel Rota B, Michael D, Marcello P, Horst B (2011) Semantic image labeling as a label puzzle game. In: Proceedings of the british machine vision conference 2011(BMVC 2011), Dundee, Scotland, pp 111.1–111.12
3. Shi-Jim Y, Shih-Yuan C, Cheng-Wei C, Tsan-Cheng S (2011) Elimination search for puzzle games: an application for Hashi solver. In: IEEE international conference on fuzzy systems (FUZZ 2011), Taipei, pp 185–189
4. Jho Cheung W (2012) Polyomino tiling method for puzzle video game. J Korean Soc Comput Game 25(1):135–142
5. Yong-Seok C et al (2012) Deadlock detection in puzzle game environment. J KIISE Comput Practices Lett 18(6):450–457
6. Jing Z (2012) The design and implement of puzzle kind educational game software. Macao, pp 185–193
7. Bastian B, Nguyen-Tuong D, Heiner M, Alois K (2013) Solving the 15-puzzle game using local value-iteration. Aalborg, Denmark, pp 45–54
8. Martyn A, Jack C (2012) A genetic algorithm for the zen puzzle garden game. Nat Comput 11 (3):353–359
9. Xiang C, Michael M, Ravin B (2008) Flashlight jigsaw: an exploratory study of an ad-hoc multi-player game on public displays. ACM, San Diego, pp 77–86

Chapter 41
Optimization Research of Bus Stops that Are Based on Berths

Kang yongzheng, Chang aixin and Zhao zhen

Abstract For the purpose of optimizing the bus stations, this article finds the methods and steps of designing the bus stop berth number in the process of analyzing the factors that affects traffic capacity. Meanwhile, analyzing the two berths bus stops queuing system particularly by using M/M/s Model of queuing theory, which provide basic theory of application example for calculating the maximum number lines of bus stops. Making data investigation of Shijiazhuang Baolongcang shopping plaza bus stations and analyzing the characteristics of the data, calculating the berths and lines by using M/M/s Model of queuing theory, raise improvement suggestions and optimum proposal, comprehensive analysis results that when the berth number is 3, the bus stations service ability of Shijiazhuang City is the best. This article introduces the process of calculating the bus stations berth number, also presents the Particle Swarm Optimization Algorithm methods applied in bus stations berth numbers by theoretical analysis and practical applications, which provides particular theory and technical support for optimizing the bus stations of Shijiazhuang City.

Keywords Optimization of bus stops · Berths · Queuing system

41.1 Introduction

Bus stops are the indispensable parts in public transportation, which plays vital role in realizing Bus Priority Strategy. Whether the bus stops are designed reasonable not only affects buses' transportation speed and efficiency, but also has influence on public vehicles' transportation capacity in the same route section.

There are many factors influencing bus stops, such as design form of these stops and number of berths and so on. At present, there are serious mismatches on the berths and lines of stops around new train station in Shijiazhuang, under which situation they fail to meet the transportation capacity need in rush hours. These also

K. yongzheng (✉) · C. aixin · Z. zhen
Shijiazhuang Tiedao University, Shijiazhuang 050043, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_41

cause long queues along these bus stops and some buses even stop their vehicles outside these stops. Some passengers have to take long walk to the station, which brings them lots of inconvenience.

This article starts its study from berths of bus stops, taking Shijiazhuang Baolongcang shopping mall bus stops as example, doing statistics survey and analysis and calculating its transportation capacity by adopting related models, and put forward relevant improvement measures and optimization plans.

41.2 Influential Factors of Transportation Capacity

The named transportation capacity of bus stops actually refers to the maximum number of buses that can hold in a certain bus stop in the same unit time and under same road condition. Many factors can affect transportation capacity of a bus stop, for instance, ways of selling tickets, access ways, traffic signal control of crossings, berths, types of stops, types of buses (lengths, width, and numbers), etc.

Generally speaking, people will queue up to access to a bus station, and it is rarely to see buses overtake queues and pull into a station. Therefore, bus station is a typical queuing system of M/M/s.

41.2.1 How to Define Berths of a Bus Stop

In order to ensure sufficient berths to meet the needs of these buses, stopping time and arrival routines of these buses in rush hours should be investigated and analyzed, i.e., when $p < 1$, which means $k > \lambda/\mu$, bus stops can work stably. Therefore, the initial value of berth of a bus stop should be set as the least positive integer:

Step 1 Set the berth initial value of this bus stop, i.e., k_0.

Step 2 After calculation, if berth of one bus stop is $k = k_0$, then queuing system of a bus stop has the average number of buses. Assume probability of the average number of buses which is larger than that of berths in the queue system of this bus stop is P:

Step 3 If Ls is less than or equal to effective berths in this bus stop, and simultaneously $p > \alpha$, that means the berths need to be set. And if Ls is larger than the berths, then there should add one berth and step 2 should be recalculated and obtain values of Ls and P.

41.2.2 Two-Berth Stop Queue System Analysis

Hereby, we will provide the two-berth data analysis because most of Shijiazhuang bus stops have two berths. When there are two berths in a bus stop, i.e., $K = 2$, i.e., $\frac{\lambda}{\mu} = 2\rho$, then the probability of no bus in this stop is:

$$p(0) = \left[1 + \frac{\lambda}{\mu} + \frac{1}{2(1-\rho)}\left(\frac{\lambda}{\mu}\right)^2\right]^{-1} = \frac{1-\rho}{1+\rho} \tag{41.1}$$

Probability of number of buses less than two berths in this stop is:

$$p(\leq 2) = p(0) + p(1) + p(2) = 1 - \frac{2\rho^3}{1+\rho} \tag{41.2}$$

Probability of number of buses more than two berths in this stop is:

$$p(> 2) = 1 - P(\leq 2) = \frac{2\rho^3}{1+\rho} \tag{41.3}$$

Average bus number in this bus stop is:

$$Ls = \rho + \frac{\rho^2(k+1)}{2k(1-\rho)} = \rho + \frac{3\rho^2}{4(1-\rho)} \tag{41.4}$$

41.3 Data Analysis in Bus Stop

Do bus stop optimization in Baohongcang shopping mall. Road segment Zhonghua South Street is the main city road features 2228 m long with three-carriage ways and a 45 m wide red line. It is two-way six-lane road with 5 m wide non-vehicle lane. Sidewalks are 5 m wide and g/c is 0.75. Baolongcang shopping mall stop is linear bus stop with two berths and 11 bus lines. See below, Table 41.1 for data investigation.

As there are two berths in this bus stop, the queue system can be regarded as single-way multiple queue system, and it is subject to first come first served principle.

41.3.1 Transportation Capacity Calculation of Stops

Most of buses in Shijiazhuang have two doors, and front doors are for on while rear doors for off. Data shows it takes 10 s from buses starts to they completely leave the

Table 41.1 Data investigation in Baolongcang shopping mall stop, Zhonghua South Street

No of bus lines	Arrived buses in a certain time	Total passengers on buses	Total passengers off buses	Total on-and-off-bus time
No. 1 ring	29	184	97	511
No. 2 ring	27	118	172	429
8	59	531	767	1893
13	37	265	375	923
20	32	117	69	311
22	38	275	254	817
55	41	490	510	1523
59	46	320	249	1011
148	55	726	378	1381
319	36	226	354	938
Quick13	31	190	339	739

stop, which means t_{c1} is 10 s. As the bus station is linear stops, it can be drawn from the above data that the clearance time of this stop is t_{c1}, i.e., 10 s.

According to Table 41.1, it can conclude that for each passenger on average they spend $t_{\text{io}} = 23.25$ s getting on–and-off buses. Because in rush hours there will be more passengers, time on opening and close doors t_{oc} is 3 s. Then, the total time that a bus spends in a stop is $t_d = t_{\text{io}} + t_{\text{oc}} = 23.25 + 3 = 26.25$ s, from which it can be concluded that the actual transportation capacity in this station is:

$$C_B = \frac{3600 \times 0.833 \times 0.75}{0.75 \times 31.6 \times 10} \times 1.85 \approx 123(\text{veh/h})$$

When transferred to service rate of this stop, it is: $C_B/60 = 123/60 = 2.05$ (veh/min).

41.3.2 The Maximum Calculated Bus Lines of Bus Stops

According to Table 41.2, it can come to conclusion that among the bus lines in Baolongcang shopping mall in Zhonghua South Street, average arrival rate of these buses is $\overline{\lambda} \approx 0.163$ (veh/min). If $\alpha = 0.05$, the average stopping time of this bus stop is 23.25 + 3 + 10 = 36.25 s. And the service rate of this berth is $\mu = (1/36.25) \times 60 \approx 1.655$ (veh/min), $\rho = (11 \times 0.163)/(2 \times 1.655) \approx 0.542$.

And probability that number of stopping buses exceeds that of berth is:

$$\rho(>2) = \frac{2\rho^3}{1+\rho} = 0.207,$$
$$m \le \frac{\mu}{\lambda}\left(\sqrt[3]{2\alpha + 2\alpha\sqrt{1+2\alpha/27}} + \sqrt[3]{2\alpha - 2\alpha\sqrt{1+2\alpha/27}}\right) = \frac{1.655}{0.163} \times 0.528 \approx 5.4$$

Assume m = 6, based on the above statistics, it can be drawn in the system, probability of number of buses larger than that of berth is described as P (>2), the data is 0.207, and if larger, it will be 0.05, which means at least six bus lines can operate stably. And on the other hand, service capacity of this bus stop does not definitely mean it meets the stopping needs of the 11 bus lines.

41.4 Optimization Plan of Bus Stops

Let us assume that there are K berths in one stop and μ is the service rate of this berth while λ is the average arrival rate of buses in this stop. To ensure the stable operation of this stop, first $P < 1$ is needed to be guaranteed.

Therefore, the initial value of these berths can be set as the minimal positive integral that is greater than λ/μ. From the above value it can be calculated the berths of the bus stop system, which can be described as the average arrival rate of these stops by the following:

$$\lambda = \sum_{i=1}^{m} \lambda_i \approx 0.121 + 0.112 + 0.246 + 0.154 + 0.133 + 0.158 + 0.170 + 0.192 + 0.229 + 0.150 + 0.129 = 1.793$$

(veh/min) average stopping time in one stop: 23.25 + 3+10 = 36.25 s service rate of this berth: $\mu = \frac{1}{36.25} \times 60 \approx 1.655$ (veh/min), $\frac{\lambda}{\mu} = \frac{1.793}{1.655} \approx 1083$.

Thus, when k_0 = 2, the average buses number in the bus queue system can be stated as: Ls equals 1.021 while less than 1.85, it means Ls is less than or equal the effective berth number, which conforms to the rules; When $P \ge 0.207$ while larger than α, then there needs to add one berth, i.e., K = 3. And when K = 3, Ls = 0.496 < 2.45, P(>3) = 0.039. If value α equals 0.05 or 0.1, then the effective berth number in this stop is bigger than the average buses number. Meanwhile, in the system it is permissible that buses number is larger than the berth's probability P(>3). In consequence, it can be confirmed that berth in this stop can be set as 3. Assume one berth is added, i.e., 3, the transportation capacity in this stop is:

$$C_B = \frac{3600 \times 0.833 \times 0.75}{0.75 \times 26.25 \times 10} \times 2.45 \approx 186$$

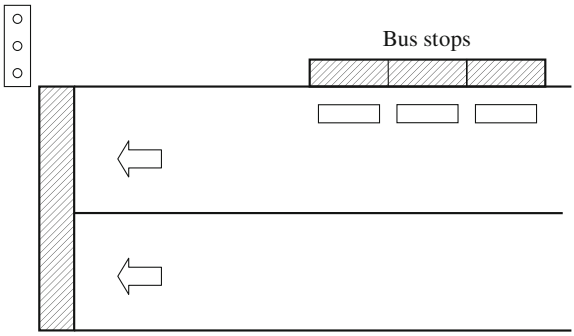


Fig. 41.1 Optimization chart of bus stop

Table 41.3 Optimization of Southern China street before the delay time cars

	Delay (s)	Stopd (s)	Stops	#Veh	Pers (s)	#Pers
Number	1	1	1	1	1	1
3600	96.1	30.6	1.97	63	96.1	63
All	96.1	30.6	1.97	63	96.1	63

Table 41.4 Optimization of Southern China street before the delay time buses

	Delay s)	Stopd (s)	Stops	#Veh	Pers (s)	#Pers
Number	1	1	1	1	1	1
3600	90.9	54.7	2.85	20	122.9	371
All	90.9	54.7	2.85	20	122.9	371

Table 41.5 Optimization of Southern China street after the delay time cars

	Delay (s)	Stopd (s)	Stops	#Veh	Pers (s)	#Pers
Number	2	2	2	2	2	2
3600	65.6	11.2	1.49	39	65.6	39
All	65.6	11.2	1.49	39	65.6	39

When transferred into service rate of this stop, it is $C_B/60 = 186/60 = 3.1$ (veh/min). When $\bar{\lambda} = 0.163$(veh/min), $\alpha = 0.05$ and $P(>3) = 0.039$, from which it can be concluded that upgrading Baolongcang shopping mall into a three-berth stop will greatly improve its transportation capacity as well as cut down its queue probability in this stop. Shown as below Fig. 41.1.

Do simulation on the proposed solution by applying Vissim and compare the delayed time before and after optimization. See Tables 41.3, 41.4, 41.5 and 41.6.

Among which, Stopd is the average stopping time of each bus (s); Stops refers to average stopping times of each bus; #Veh means the passed vehicles; Pers is the average delayed time when there no no-bus passenger; #Pers means number of passed passengers, all the above do not include on-bus time of each passengers in this stop.

Table 41.6 Optimization of Southern China street after the delay time buses

	Delay (s)	Stopd (s)	Stops	#Veh	Pers (s)	#Pers
Number	2	2	2	2	2	2
3600	69.4	34.2	2.52	23	115.1	413
All	69.4	34.2	2.52	23	115.1	413

From the above table can be seen after the optimization of the vehicle delay is reduced, especially for large passenger cars to reduce the number of passengers increased by.

41.5 Conclusion

This article analyzes transportation analysis of bus stops. Starts from berths, it then defines bus lines, and do data calculation by applying knowledge of queuing theory.

Taking Shijiazhuang Baolongchang shopping mall as examples, this article does survey first, then calculates its berths and bus lines of these stops by adopting models. It also does optimization of these stops in an attempt to change their berths and analyze their transportation capacity as well as probability of queue condition before and after this optimization. These are all followed by a best optimization plan. It can meet the stopping needs of this bus stops when add the berths to three, which lows down the queue probability, which also provides reference and technical support for optimization of bus stops in Shijiazhuang.

Acknowledgment This work was supported by Social Science Fund Project of Hebei Province (HB13GL008).

Chapter 42
Passenger Flow Trend Analysis and Forecasting for Large High-Speed Railway Stations During Holidays: A Case Study for Beijing West Railway Station

Jinjin Tang, Leishan Zhou, Yong Zhao and Jingjing Shao

Abstract With the rapid development of China's social economy and the improvement of people's living standards, people's travel demand grows rapidly. As high-speed rail provides fast and comfortable services, the number of high-speed railway passengers is increasing exponentially. During the holidays, huge numbers of railway passengers always bring great pressure and challenge to the transport companies, particularly at large railway stations. Accurately forecasting passengers' travel demand during holidays is of practical significance to efficiently develop train schedules, formulate effective emergency plans and early warning schemes and guarantee the operating safety. With historical passenger flow data of Beijing West Railway Station during the Spring Festival, we analyse and forecast the future passenger flow using a trend analysis method. The result provides suggestions on organization optimization and improvement for Beijing West Railway Station, which can be used by other railway passenger stations as references.

Keywords High-speed rail · Passenger flow forecasting · Trend analysis method · Beijing west railway station

42.1 Introduction

As a resource-saving and environment-friendly transport mode, railway is the major and fundamental infrastructure of a country and plays an irreplaceable role on speeding up the industrialization and urbanization process, optimizing transportation structure, reducing social logistics costs, facilitating people's safe travel. After years of construction, China railway network makes great achievement. By the end

J. Tang (✉) · L. Zhou · J. Shao
Beijing Jiaotong University, Beijing, China
e-mail: jjtang@bjtu.edu.cn

Y. Zhao
Beijing West Railway Station, Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_42

of 2014, China's railway mileage exceeded 110,000 km with a total mileage of high-speed railway over 15,000 km. According to the "long-term railway network plan (2008 adjusted)", by 2020, the national railway mileage will be 120,000 km or more with the mileage of passenger dedicated lines more than 16,000 km. As a result, railway transportation plays an increasingly important role in China's comprehensive transportation system.

As the basic production units for railway passenger transport, railway stations/terminals are the locations where trains stop in order to load or unload freight or passengers. Terminals are responsible for passenger and cargo transport of long-distance, short-distance and inner-city; they are also the major locations where connections/transfers may be available to crossing rail lines or other transport modes, effective convergence between inner-city and external transport, different mode transference. One of the most important tasks for railway passenger transport is to offer safe, rapid and accurate services to satisfy travellers' demand. Under large passenger volume scenarios, where passenger components are complex, evacuation space is limited and the flow lines intersect with each other, the merits and demerits of passenger flow organization will seriously affect the quality and efficiency of railway passenger transport.

Beijing West Railway Station, affiliated with Beijing Railway Bureau, China, is one of Asian largest passenger railway terminals. With the sustained and rapid economic development and the improvement of people's living standards, the daily passenger traffic in Beijing West Railway Station increases dramatically during holiday periods, especially during the Spring Festival. Accurately forecasting passenger demand during holiday period is the foundation to provide reasonable transportation services with limited railway resources. In this research, using the historical passenger flow data of Beijing West Railway Station during the Spring Festival, we forecast the passenger flow trend in order to provide useful information that can be used to improve the station's transport organization quality and efficiency during the Spring Festival. This paper is organized as follows. Section 42.2 is about review of research, Sect. 42.3 is about passenger flow analysis and forecasting of Beijing West Railway Station during the Spring Festival and the last section is about conclusion.

42.2 Review of Research

Scholars have made great efforts on passenger flow organization during holidays. Since 1960s, researchers in Germany, France, Japan and other developed countries have begun to take advantage of abundant capital and advanced design concepts to transform railway terminals, to enhance passenger travel conditions, and to improve transport organizations efficiency and quality [1–5]. Subprason [5] focused on optimizing passenger flow lines and calculated optimal number of passengers occupying space under different flow line conditions to provide references to passenger waiting-room area construction. Some scholars [6–10] focused on large

passenger flow formed during holidays, such as "National Day", "Spring Transportation" and "Summer Transportation" in China. They analysed the forming causes of large railway passenger flow, summarized the passenger transport characteristics during the holidays, and discussed passenger transport organization methods during holidays. There are some approaches that are commonly used, to name a few, passenger survey analysis, ticketing organization and management optimization, ordering passengers take down, strengthening the local control as well as strengthening equipment and facilities management. Sun [7] analysed various factors affecting passenger flow such as time, space, flow properties, price and transport capacity, and proposed the forecast model based on complex system with bringing modal concept Xu et al. [8] proposed a new railway passenger flow prediction method based on spatio-temporal data mining, combined the statistical theory and neural network algorithm to improve the prediction accuracy. Sun et al. [9] studied factors affecting intercity passenger flow, using the correlation analysis system to establish a new railway intercity passenger forecast model. Xia [10] analysed the main factors influencing the fluctuation of passenger flow on holidays, proposed the forecast model based on the flow fluctuation, and used the actual ticket data for the validation. Previous studies provide references for the study on passenger flow of Beijing West Railway Station.

42.3 Passenger Flow Trend Analysis and Forecasting of Beijing West Railway Station During the Spring Festival

Accurately analyzing and forecasting passenger flow trends during the Spring Festival is of critical importance on reasonably allocating transport resources of Beijing West Railway Station and easing insufficient capacity of both stations and sections.

Most commonly used methods of railway passenger flow forecasting include: passenger flow forecasting based on grey theory, short-term forecasting based on radial neural network theory, four-stage method based on passenger flow nature, gravity model forecasting, trend extrapolation and analogy method, regression analysis method. The existing research results indicate that it is better to choose the function fitting method to forecast data if there are sufficient sample data. With actual passenger flow data of Beijing West Railway Station from 2004 to 2015, shown in Fig. 42.1, it is possible to adopt function fitting method to forecast future passenger flow.

Automatically fitting the passenger number of Beijing West Railway Station during the Spring Festival through Microsoft EXCEL software, we get the fitting functions shown in Figs. 42.2, 42.3, 42.4, 42.5, 42.6, 42.7 and 42.8 (corresponding to linear function fitting, quadratic function fitting, cubic function fitting, four equation fitting, logarithmic function fitting, exponential function fitting and

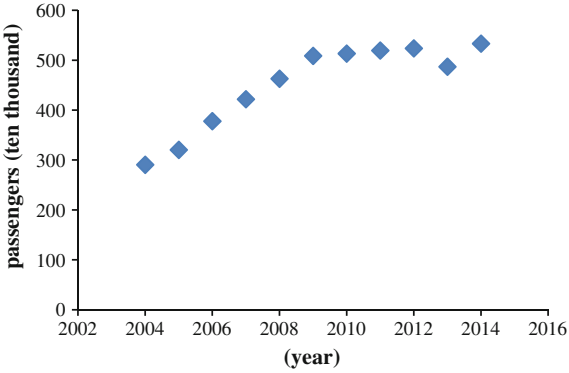


Fig. 42.1 Number of travellers at Beijing West Railway Station during the spring festival

piecewise linear function fitting, respectively). It should be noted that R^2, also known as the determination coefficient reflects how well a statistical model fits actual data. If R^2 is higher, the fitting degree is better. The value of R^2 is between 0 and 1 and when it is equal to 1 or close to 1, the reliability of trend line is higher. This paper selects the fitting function in terms of the value of R^2.

The function corresponding to the linear fitting is

$$y = 23.302x - 46363 \tag{42.1}$$

where $R^2 = 0.8028$, the figure corresponding to the linear fitting function is shown in Fig. 42.2.

The function corresponding to the quadratic fitting is

$$y = -3.755\,x^2 + 15111\,x - 2\mathrm{E}+07 \tag{42.2}$$

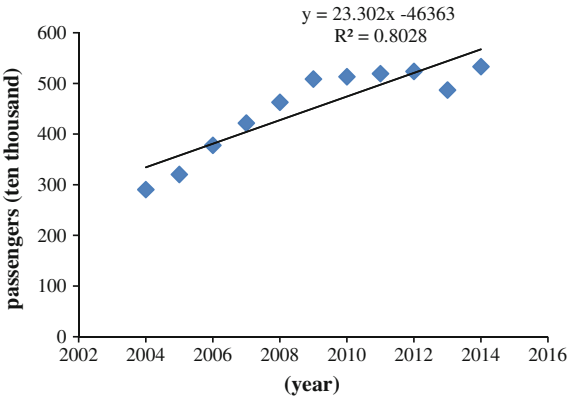


Fig. 42.2 Linear fitting

where E represents scientific notation, and E in other formulas below has the same meaning, for example 2E + 07 indicating 2×10^7. $R^2 = 0.9654$. The figure corresponding to the quadratic fitting function is shown in Fig. 42.3.

The function corresponding to the cubic fitting is

$$y = 0.0235x^3 - 145.65x^2 + 300176x - 2\text{E} + 08 \tag{42.3}$$

where $R^2 = 0.9655$, the figure corresponding to the cubic fitting function is shown in Fig. 42.4.

The function corresponding to the four equation fitting is

$$y = 0.205x^4 - 1647x^3 + 5\text{E} + 06x^2 - 7\text{E} + 09x + 3\text{E} + 12 \tag{42.4}$$

where $R^2 = 0.9887$, the figure corresponding to the four equation fitting function is shown in Fig. 42.5.

The function corresponding to the logarithmic fitting is

$$y = 46828 \ln(x) - 355694 \tag{42.5}$$

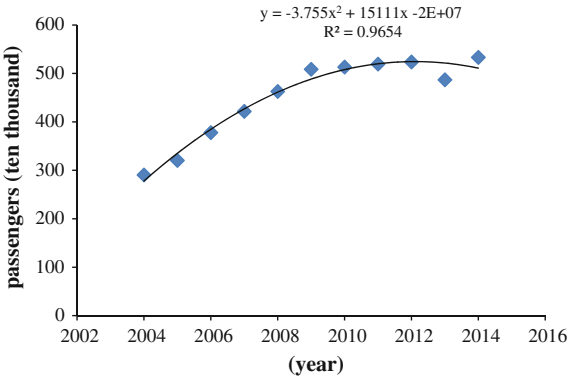


Fig. 42.3 Quadratic fitting

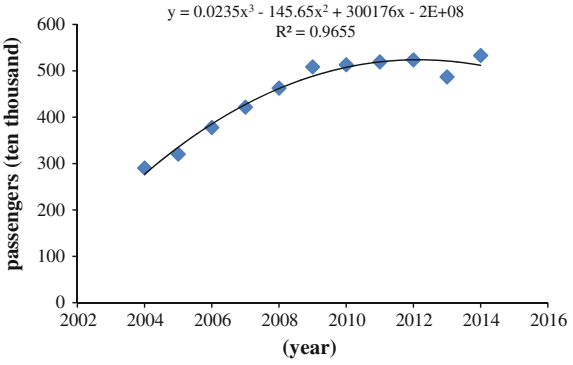


Fig. 42.4 Cubic fitting

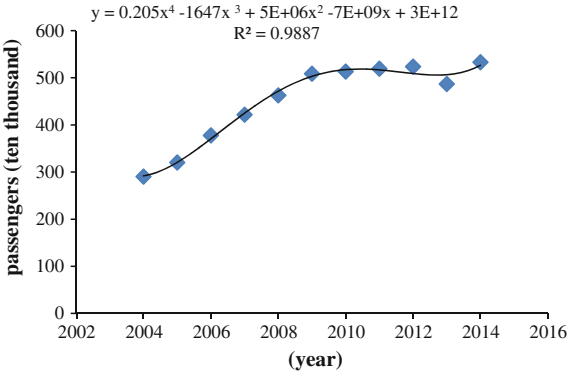


Fig. 42.5 Four equation fitting

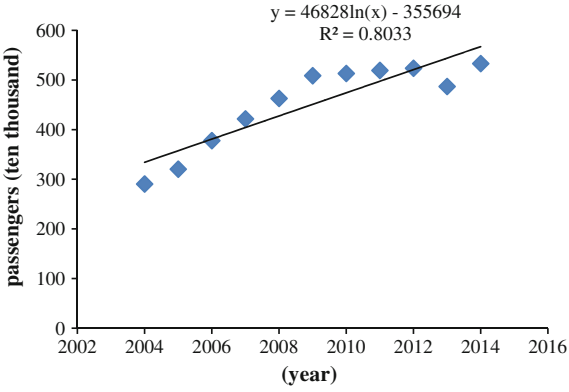


Fig. 42.6 Logarithmic function fitting

where $R^2 = 0.8033$, the figure corresponding to the logarithmic fitting function is shown in Fig. 42.6.

The function corresponding to the exponential fitting is

$$y = 2\mathrm{E} - 47\mathrm{e}^{0.0565x} \tag{42.6}$$

where $R^2 = 0.779$, the figure corresponding to the exponential fitting function is shown in Fig. 42.7.

The function corresponding to the piecewise linear fitting is

$$y = \begin{cases} 40.357x - 80583, & \text{if } x \leq 2009 \\ 4.8x - 9214.5, & \text{if } x > 2009 \end{cases} \tag{42.7}$$

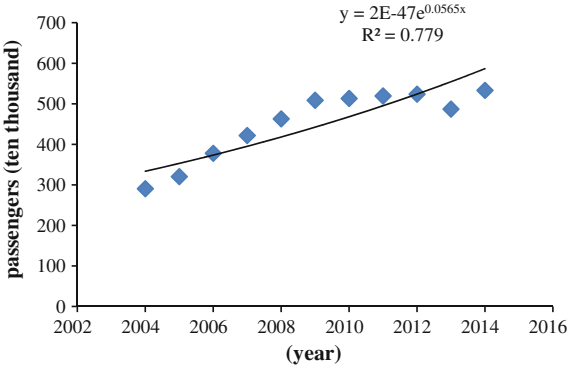


Fig. 42.7 Exponential fitting

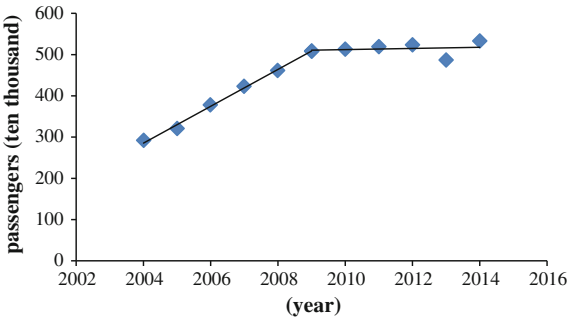


Fig. 42.8 The linear fittingness of passengers

where $R^2 = 0.9957$, the figure corresponding to the piecewise linear fitting function is shown in Fig. 42.8.

From the above fitting functions, the maximum value of R^2 is 0.9957 which means the piecewise linear fitting best. According to the above piecewise function (42.7), the predicted passenger flow from 2010 to 2019 shown in Table 42.1. In Table 42.1, the calculation formula of error rate is shown in Formula (42.8).

$$\text{error rate} = \frac{\text{predicted data} - \text{actual data}}{\text{actual data}} \times 100\,\% \tag{42.8}$$

The error rates of most predicted data are within the permissible range from 2010 to 2015 except for the data in 2013, when Beijing West Railway Station was under platform construction. So, predicted data from 2016 to 2019 by this method have a higher reliability.

Table 42.1 Forecasting result from 2010 to 2019 by the piecewise function method

Year	Estimated data (ten thousand)	Actual data (ten thousand)	Error rate (%)
2010	513.86	513.30	0.11
2011	518.7	519.40	0.13
2012	523.54	522.30	0.23
2013	528.38	487.00	7.83
2014	533.22	533.20	0.01
2015	538.9	538.20	0.1
2016	542.9	–	–
2017	547.74	–	–
2018	552.58	–	–
2019	557.42	–	–

42.4 Conclusion

The passenger demand forecasting during holidays aims to provide essential information for train timetabling, platform planning, emergency countermeasures and the railway safety controls. To this end, it is necessary to forecast passenger flow reasonably and accurately. It is very important to select high prediction accuracy methods in terms of various factors that affect railway passenger traffic (such as time, economic conditions, weather, etc.). Using the historical passenger flow data of Beijing West Railway Station during the Spring Festival we analyse and forecast the future passenger flow by adopting the fitting method with piecewise linear function (42.7), whose R^2 is equal to 0.9957. And predicted data in 2014 and 2015 by this method are closest to actual data. Thus, the results by the piecewise method are chosen to be the fundamental basis for the operating plan of Beijing West Railway Station. Moreover, the forecasting method of Beijing West Railway Station can be used by other railway passenger stations in China as a reference.

Acknowledgements The material in this paper is based on research supported by "The Fundamental Research Funds for the Central Universities" under grant No. 2015JBM057 "Research on assessment and improvement train schedule for high-speed railway based on multi-resource constraints".

References

1. Bussieck MR, Winter T, Zimmermann UT (1997) Discrete optimization in public rail transport. Math Program 79:415–444
2. Dorfman M, Medanic J (2004) Scheduling trains on a railway network using a discrete event model of railway traffic. Transp Res Part B 38:81–98

3. Gatersleben MR, van der Weij SW (1999) Analysis and simulation of passenger flow in an airport terminal. In: Proceeding in Winter Simulation Conference. Phoenix, AZ, USA
4. Takakuwa S, Oyama T (2003) Simulation analysis of international departure passenger flows in an airport terminal. In: Proceedings in winter simulation conference. New Orleans, LA, USA
5. Subprason K, Seneviratne P, Kilpala H (2002) Cost-based space estimation in passenger terminals. J Transp Eng 128(2):13
6. Wang Hua (2004) Character of passenger transportation under fastigium and counter measurements. Railway Transport Econ 26(10):38–39 (in Chinese)
7. Quanxin S (2006) Train organizational theory with application for railway passenger lines. Beijing Jiaotong University, Beijing (in Chinese)
8. Wei X, Huang H, Qin Y (2004) Study of railway passenger flow forecasting method based on spatio-temporal data mining. J Beijing Jiaotong Univ 5:1–4 (in Chinese)
9. Sun W, Liu G (2005) Research on the model of forecasting passenger flow of the intercity train. J Beijing Jiaotong Univ 29(3) (in Chinese)
10. Xia Q (2011) Research on fluctuation of holiday railway passenger flow and application on forecast of passenger flow. Beijing Jiaotong University, Beijing (in Chinese)

Chapter 43
Research on the Topological Structure Description of Urban Rail Transit Network

Xuan Li

Abstract The topological structure of urban rail transit network directly determines the degree of difficulty of passenger transfer in the transit system. Specific methods are proposed to establish a dual-topology map for an urban rail transit network. Then, the matrix is used for describing the connection relationships between transit lines within a network. Accordingly, a case study is carried out for the rail network in Beijing. Based on the matrix description, some issues are found out in transit accessibility and transfer convenience of Beijing's rail transit network. Finally, several suggestions are provided for the planning of urban rail transit networks that would assure convenience for transfer passengers.

Keywords Urban rail transit network · Topological structure · Matrix · Accessibility

43.1 Introduction

The topological structure of an urban rail transit network (URTN) is constrained by several conditions, including the city's geography, the land use layout, and population distribution. Meanwhile, the topological structure depends on the function and status of URTN in an urban public transport system. Each city has its own unique conditions, which lead, to a certain extent, to different rail transit network structures in different cities [1, 2]. It is a common point of view in systems science that structure determines function [3]. Different topological structures found in URTNs determine various transit accessibility, resulting in different convenience levels for passengers. Therefore, when studying transit accessibility, it is first necessary to extract topological structures and construct topology maps.

X. Li (✉)
School of Maritime and Transportation, Ningbo University, No. 818, Fenghua Road, Jiangbei District, Ningbo 315211, China
e-mail: lixuan@nbu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_43

Different features stand for different network structures, and different structures cause different systematic functions [4, 5]. The topological structure of URTN is directly related to whether passengers can successfully reach destinations by the rail network, and whether it is necessary to transfer within the network and the transfer times [6]. Studies into the topological structure will show useful applicable value in terms of the design, evaluation and reliability analysis of the URTN. However, there has been little research in the area. In this study, I have proposed some methods to establish a dual-topology map for an URTN. Then I used the matrix for describing the connection relationships between transit lines within a network, which can be used as the quantitative analysis tool of assessing, selection, and optimization of a transit network.

43.2 Construction of Topology Map of URTN

The construction of a transit network topology map is meant to abstract the actual elements into vertices and edges of undirected graphs. Depending on the various methods used to abstract, there are two categories: the primal approach and the dual approach. With the primal approach, nodes (intersections, stations, etc.) in the transit network are abstracted as graph vertices, while connecting lines between nodes (roads, intervals, etc.) are abstracted as graph edges [7]. With the dual approach, vertices represent linear facilities in the transit network (roads, intervals, etc.), and edges represent the connecting relationships between these facilities [8].

In the existing studies of topological structure of urban public transport network, the primal approach is widely applied to construct the topology map as well as crafting mappings from stations to vertices and intervals to edges. For example, the L-space method and P-space method, which are the two methods for describing public transport networks, summarized by Gao et al., both use the primal approach [9]. However, when dealing with the complex public transport network, where there are a number of stations, and the dimensions of the corresponding vertices and edges will greatly increase. At the same time, primal maps reflect neither the attribution relationships between stations and lines nor the connection relationships between those lines. Differences between connection relationships in various public transport networks are the causes of different ISAs. To pave the way for the subsequent studies of ISA, I adopted the dual approach to generate a topology map of the URTN.

A dual-topology map of the URTN is denoted as an undirected graph $G(V, E)$, where: $V(G) = \{v_1,\ v_2,\ \ldots, v_n\}$ is a non-empty finite vertex set, v_i is vertex and V is vertex set. Also, $E(G) = \{e_1,\ e_2,\ \ldots, e_m\}$ is a finite edge set, e_i is edge and E is edge set. I abstracted rail lines as vertices regardless of directions. $V(G) = \{v_1,\ v_2,\ \ldots, v_n\}$ represented the lines set. n was the number of lines. Direct connection relationships between lines were represented by edges. If line i and line

j intersect, there is an edge between vertex v_i and v_j; namely they are adjacent. On the contrary, when v_i and v_j, are not adjacent, there is no edge in between them.

For example, Fig. 43.1 is a sketch map of a URTN. It includes 5 lines (L1, L2, L3, L4, and L5); a corresponding dual topology map is shown in Fig. 43.2.

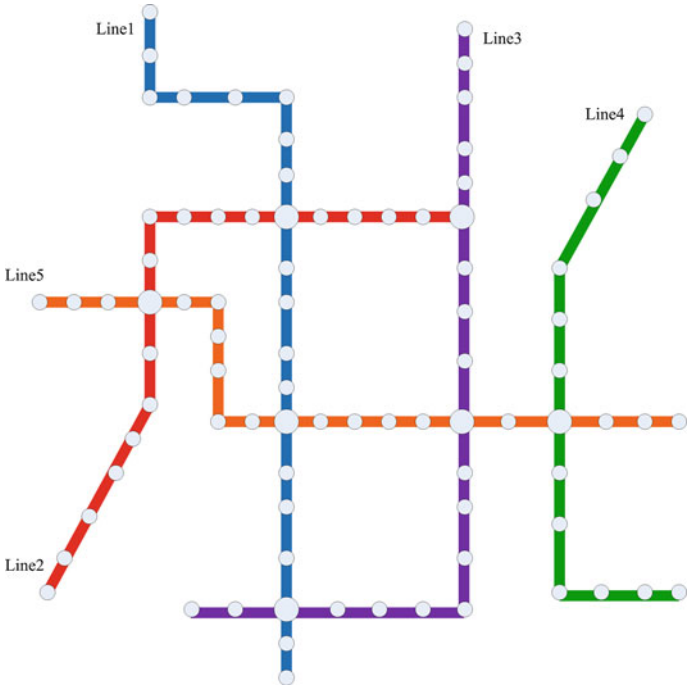


Fig. 43.1 A sketch map of A URTN

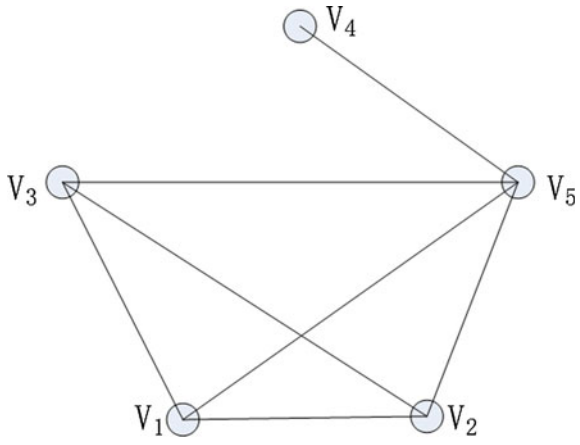


Fig. 43.2 Dual-topology map the URTN in Fig. 43.1

43.3 Matrix Description of URTN

Based on the dual topology map of the URTN, I applied the concepts such as an adjacency matrix to describe the connection relationships between lines in the URTN and established a direct transfer matrix, accessibility matrix, and a minimum transfer time matrix.

43.3.1 Direct Transfer Matrix

For the undirected graph $G(V,E)$, $V(G) = \{v_1,\ v_2,\ \ldots, v_n\}$, $A\left(a_{ij}\right)_{n\times n}$ is denoted as the adjacency matrix of $G(V,E)$, where

$$a_{ij} = \begin{cases} 1, & v_i v_j \in E(G) \\ 0, & v_i v_j \notin E(G) \end{cases} \tag{43.1}$$

The adjacent matrix of the undirected graph is a symmetric matrix with elements of either 0 or 1, and the diagonal elements are 0. After abstracting the URTN into the dual topology map G, the corresponding adjacency matrix $A\left(a_{ij}\right)_{n\times n}$ can be denoted as a direct transfer matrix of the URTN. $a_{ij} = 1$ ($a_{ji} = 1$) represents that line i and line j intersect directly. Passengers interchanging between these two lines need only one transfer; that is, for passengers these two lines are accessible to each other through transfer once. Conversely, $a_{ij} = 0$ ($a_{ji} = 0$) represents that line i and line j do not intersect. For passengers these two lines are not accessible to each other through transfer once.

The direct transfer matrix of the URTN shown in Fig. 43.1 is $A = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$.

Apart from the diagonal elements, only elements on row 4 and column 4 are 0, indicating that each two lines in L1, L2, L3, and L5 intersect directly, and for passengers each two lines in them are accessible to each other through transfer once. L4 intersects with L5 but not L1, L2, or L3. For passengers, L4 and L5 are accessible to each other through transfer once, but not L1, L2 or L3.

43.3.2 Accessibility Matrix

A direct transfer matrix of the URTN reflects whether passengers can interchange between any two lines through transfer once. In addition, we also concerned

whether passengers can interchange between any two lines through transfer several times, so I introduced an accessibility matrix to represent the accessibility relationships between lines.

For an undirected graph $G(V,E)$, $V(G) = \{v_1,\ v_2,\ \ldots, v_n\}$, a matrix $P\left(p_{ij}\right)_{n\times n}$ was defined, where

$$p_{ij} = \begin{cases} 1, & \text{there are paths between } v_i \text{ and } v_j, \\ 0, & \text{otherwise.} \end{cases} \tag{43.2}$$

Thus, the accessibility matrix for graph G is denoted as matrix P, which is also a symmetric matrix.

The accessibility matrix shows whether a path exists between any two nodes. This can be determined by calculating the adjacency matrix of graph G. Since this is a Boolean matrix with elements of either 1 or 0, matrix $A, A^2, \ldots,\ A^n$ (A^n refers to n-th power of the adjacency matrix A) can be changed to Boolean matrixes, denoted as $A(1), A(2), \ldots,\ A(n)$, then: $P = A(1) \vee A(2) \vee \ldots \vee A(n)$, where $\vee$ represents Boolean sum.

In this study, $P\left(p_{ij}\right)_{n\times n}$ corresponding to the topology graph G is called directly as accessibility matrix of the URTN, and $p_{ij} = 1$ (p_{ji}=1) represents that passengers can interchange between line i and line j via transfer. Contrarily, $p_{ij} = 0$ ($p_{ji} = 0$) represents that passengers cannot interchange between line i and line j via transfer.

The accessibility matrix of the URTN shown in Fig. 43.1 is $P = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$, exhibiting that passengers can interchange between any two lines within the network via transfer.

43.3.3 The Minimum Transfer Time Matrix

If the undirected graphs of the URTN are connected graphs, passengers are transferable between any two lines and elements of its accessibility matrix are all 1. Therefore, I am concerned more with how many times of transfer are needed for passengers to interchange between any two lines.

The adjacency matrix has the following features: set $A\left(a_{ij}\right)_{n\times n}$ as the adjacency matrix of $G(V,\ E)$, the number of paths with distance of k from vertex v_i to v_j is $a_{ij}^{(k)}$ on row i and column j in matrix A^k; A^k refers to k-th power of the adjacency matrix A. A^k is also a symmetric matrix. It can be seen that after calculating the k-th power of the direct transfer matrix $A\left(a_{ij}\right)_{n\times n}$, the obtained elements $a_{ij}^{(k)}(a_{ji}^{(k)})$ of matrix A^k represent the number of accessible paths between line i and line j through transfer

k times. Accessible paths are unrepeatable paths from line i to line j through transfer k times ($k \leq n-1$).

Through an analysis of matrix A^k, the following can be concluded:

1. If $a_{ij}^{(k)} = 0,\ k = 1, 2, \ldots, m\ ;\ m < n$, then passengers cannot interchange between line i and j through transfer m times;
2. If $a_{ij}^{(n-1)} = 0$, then passengers cannot interchange between line i and line j via transfer;
3. If $a_{ij}^{(k)} > 0$ and $a_{ij}^{(m)} = 0,\ k < n\ ;\ m = 1, 2, \ldots, k-1$, then passengers can interchange between line i and j through transfer at least m times.

I then established matrix $R\left(r_{ij}\right)_{n\times n}$ to record the minimum transfer times passengers can interchange between any two different lines. This is called the minimum transfer time matrix of the URTN, where

$$r_{ij} = \begin{cases} k, & i \neq j \quad \text{and} \quad a_{ij}^{(k)} > 0 \quad \text{and} \quad a_{ij}^{(m)} = 0; \quad k < n; \\ & m = 1, 2, \ldots, k-1; \\ \infty, & i \neq j \quad \text{and} \quad a_{ij}^{(n-1)} = 0; \\ 0, & i = j. \end{cases} \tag{43.3}$$

In the equation, $a_{ij}^{(k)}$ is the element on row i and column j of the A^k matrix. R is also a symmetric matrix. When $i \neq j$ and $a_{ij}^{(n-1)}=0$, passengers cannot interchange between line i and line j via transfer, so I defined $r_{ij} = \infty$. When $i = j$, there is no need to transfer, so $r_{ij} = 0$.

The minimum transfer time matrix of the URTN shown in Fig. 43.1 is $R = \begin{bmatrix} 0 & 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 2 & 1 \\ 2 & 2 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$, suggesting that passengers interchanging from L4 to L1, L2, or L3 need transfer at least two times.

43.4 Case Study

As shown in Fig. 43.3, the URTN in Beijing (2014) consists of 16 lines, which are Line 1, Line 2, Line 3, Line 4-Daxing Line (because Line 4 and the Daxing Line connect with each other and there is no need to transfer, they were considered to be the same line in this study), Line 5, Line 6, Line 8, Line 9, Line 10, Line 13, Line 14, Line 15, the Batong Line, the Changping Line, the Yizhuang Line, the Fangshan Line, and the Airport Line.

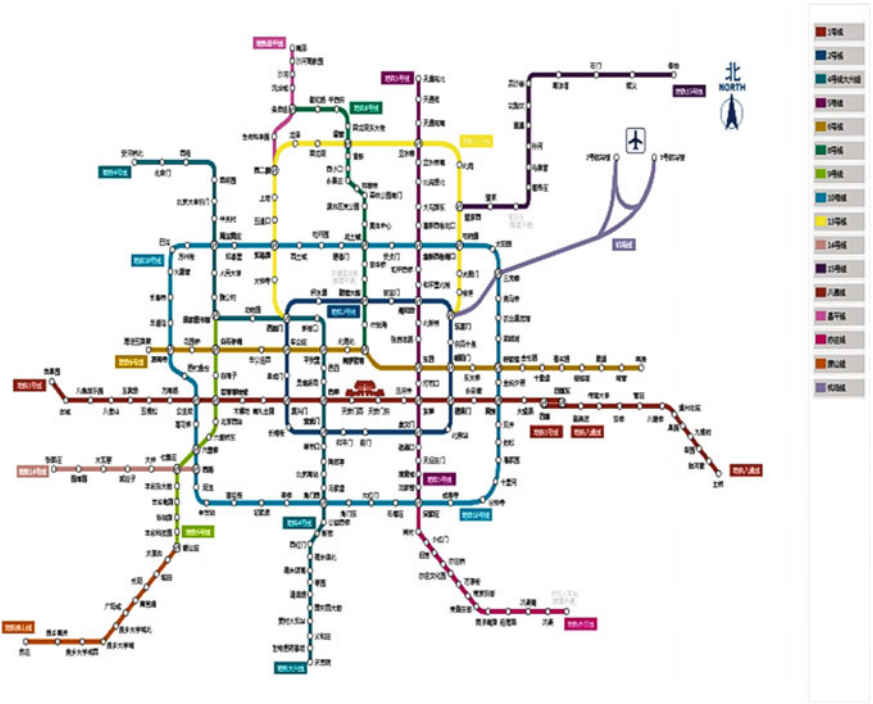


Fig. 43.3 A sketch map of Beijing's URTN (2014)

Set the corresponding vertices in an undirected graph as v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, and v_{16}. Thus, the direct transfer matrix, accessibility matrix and minimum transfer time matrix are, respectively:

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$P = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

$$R = \begin{bmatrix} 0 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & 2 & 2 & 3 & 1 & 3 & 2 & 2 & 2 \\ 1 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 1 & 3 & 2 & 2 & 2 & 2 & 3 & 1 \\ 1 & 1 & 0 & 2 & 1 & 2 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 1 & 2 & 0 & 1 & 2 & 2 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 3 & 2 \\ 2 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 2 & 2 & 3 & 3 & 3 & 2 & 2 & 2 \\ 2 & 1 & 2 & 2 & 1 & 0 & 2 & 1 & 1 & 2 & 2 & 3 & 1 & 2 & 3 & 2 \\ 1 & 2 & 1 & 2 & 1 & 2 & 0 & 1 & 2 & 1 & 3 & 2 & 3 & 2 & 1 & 2 \\ 1 & 2 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 2 & 2 & 2 & 1 & 2 & 1 \\ 2 & 1 & 1 & 1 & 2 & 1 & 2 & 1 & 0 & 2 & 1 & 3 & 1 & 2 & 3 & 1 \\ 2 & 3 & 2 & 2 & 2 & 2 & 1 & 1 & 2 & 0 & 3 & 3 & 3 & 2 & 2 & 2 \\ 3 & 2 & 2 & 2 & 3 & 2 & 3 & 2 & 1 & 3 & 0 & 4 & 2 & 3 & 4 & 2 \\ 1 & 2 & 2 & 2 & 3 & 3 & 2 & 2 & 3 & 3 & 4 & 0 & 4 & 3 & 3 & 3 \\ 3 & 2 & 2 & 2 & 3 & 1 & 3 & 2 & 1 & 3 & 2 & 4 & 0 & 3 & 4 & 3 \\ 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 & 2 & 2 & 3 & 3 & 3 & 0 & 3 & 2 \\ 2 & 3 & 2 & 3 & 2 & 3 & 1 & 2 & 3 & 2 & 4 & 3 & 4 & 3 & 0 & 3 \\ 2 & 1 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 2 & 2 & 3 & 3 & 2 & 3 & 0 \end{bmatrix}.$$

Using the above matrix, I can compute and get some significant indicators. For example, the number of maximum transfer times of Beijing's URTN is 4, which means that in Beijing's URTN there are some cases that passengers required to transfer four times to reach their destination. The combinations of line-pair, between which passengers interchanging need to transfer four times, include Line 15 and the Batong Line, Line 15 and the Fangshan Line, the Batong Line and the Changping Line, and the Fangshan Line and the Changping Line. Meanwhile, the direct transfer coefficient of Beijing's URTN is 30 %, which means the ratio of the number of direct intersecting line-pair combinations to the number of all line-pair combinations in Beijing, indicating a rather lower probability in Beijing's URTN to

transfer only once in order to reach the destination. Thus, because of the poor connection relationships between rail lines, the transit accessibility is relatively weaker in Beijing's URTN and the travel convenience in Beijing's URTN determined by the topological structure is unsatisfactory.

Through a detailed analysis, I have identified the chief cause of the poor accessibility of Beijing' URTN. That is, the number of the suburban lines (a suburban line in this study refers to a line that intersects with only one other line) in Beijing' URTN is redundant, including the Batong Line, the Changping Line, the Fangshan Line, and Line 15. In Beijing's URTN, the line-pair combinations requiring four-time transfers all exist within the suburban lines. It is evident that passengers interchanging between suburban lines or between one suburban line and other lines generally require a greater number of transfer times. Hence, suburban lines greatly reduced the overall convenience level of Beijing's URTN. If the four suburban lines were excluded, the maximum transfer times of Beijing's URTN would decrease to two.

43.5 Conclusion

The planning of the URTN generally involves multiple factors and targets; consequently, it is a complex, multi-constrained, and multi-objective optimization problem. In this study, the proposed description method of the network topological structure provided quantitative analysis tools for the evaluation, selection, and optimization of network planning schemes.

In order to improve the transit accessibility of the URTNs, we should pay attention to the following problems during the planning stage:

1. During network structure design, isolated lines should be avoided to strengthen network connectivity.
2. During network structure design, optimization and selection, the connectivity between rail lines should be considered, and suburban lines within the network should be minimized.
3. Based on network connectivity, rail lines should be constructed in a reasonable order. As shown in Fig. 43.1, if the construction order is Line 1 → Line 2 → Line 3 → Line 4 → Line 5, Line 4 remains isolated and does not intersect with any other line before Line 5 is put into operation. If the construction order of lines 4 and 5 were reversed, the network disconnectivity issue could be avoided.
4. Rail-line construction is time-consuming. The formation of the transit network is a long process. If there is an isolated line at a certain stage in the network formation, we should establish regular bus lines between the isolated line and another nearby line to facilitate passenger travel.

Acknowledgment This work is financially supported by the Zhejiang Provincial Natural Science Foundation (No. LQ13G010010) and the National Natural Science Foundation of China (No. 51408323).

References

1. Luo XQ et al (2010) Theory and methods of layout and planning of urban rail transit network. Chang'an University, Xi'an (in Chinese)
2. Gu BN et al (2000) Structural analysis of urban rail transit network. J China Railway Soc 5 (22):25–29 (in Chinese)
3. Geurs KT, van Wee B (2004) Accessibility evaluation of land-use and transport strategies: review and research directions. J Transp Geogr 12(2):127–140
4. Wu JJ et al (2008) Studies of topology complexity of urban traffic network. Beijing Jiaotong University, Beijing (in Chinese)
5. Chen Yong-Zhou, Li Nan, He Da-Ren (2007) A study on some urban bus transport networks. Phys A 376:747–754
6. Hu JH et al (2012) City accessibility research in subway mode. Planners 1(28):29–33
7. Porta Sergio, Crucitti Paolo, Latora Vito (2006) The network analysis of urban streets: a primal approach. Environ Plan 33(5):705–725
8. Porta Sergio, Crucitti Paolo, Latora Vito (2006) The network analysis of urban streets: A dual approach. Phys A 369:853–866
9. Gao ZY et al (2006) Complexity of urban traffic network. Shanghai Science and Technology Education Press, Shanghai (in Chinese)

Chapter 44
The Rail End Face Contour Extraction Based on Mathematical Morphology and Adaptive Threshold

Yan Dou, Meihuan Chen, Yuqian Zheng, Hongzhu Liu and Po Dai

Abstract To solve the problem of extracting rail end face contour in high quality, this paper proposes a rail end face contour extraction algorithm based on mathematical morphology and adaptive threshold. First, use median filter on the rail end face image captured by the industrial camera, and then design a "※" shaped kernel for image denoising based on mathematical morphology; finally, in contour extraction section, an improved Otsu algorithm combined with a histogram of gradient magnitude automatically determine the threshold. By double threshold detection, extract a complete clear contour of the rail end face. Experimental results show that the algorithm can not only suppress the smaller noise, but also have a good inhibition on noisy blocks. It is a strong self-adaptive algorithm, can get a great rail end face contour.

Keywords Rail end face · Contour extraction · Mathematical morphology · Adaptive threshold

44.1 Introduction

Rail is a key equipment of railway for train operation, while the turnout rail play an more important role in the railway lines, which is always used in stations with more cross lines as the lines connected devices. So the manufacturing precision of turnout rail is count for much, in order to guarantee quality, measuring the size is inevitable. In this paper, we propose an algorithm to exact the end face contour to measure the turnout rail end face size well.

Y. Dou (✉) · M. Chen · Y. Zheng
College of Information Science and Engineering, Yanshan University, Qinhuangdao, China
e-mail: douyan@ysu.edu.cn

H. Liu · P. Dai
China Railway Shanhaiguan Bridge Group Co., Ltd., Qinhuangdao, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_44

There are two methods for measuring the face size; one is contact measurement [1], which uses some contact instruments, such as retractable steel rule, vernier calipers, to get the size of rail face. This method is relatively original, low efficiency, and big-error because of the influence of measuring tools and personnel.

The another method is non-contact measurement method based on machine vision, which gains the image of rail face, then exacts the outline, accordingly acquires the face size. This way is more applied with far-distance, non-contact, high-precision, and fast-measurement. In the process, extracting the surface profile of the rail is the most important part, and the results will directly affect the measurement results.

Two important steps in contour extraction are image denoising and edge detection. Because of the influence of environment, illumination and other factors, there are some unnecessary noise when the image is acquired, and simple image smoothing cannot meet the ideal requirements, so this paper propose a new method to eliminate noise block based on mathematical morphology. The key using mathematical morphology denoising is to determine structural elements, which will have a direct impact on the input image processing results.

There are many methods of edge detection [2]: Sobel, Laplace, Log, Robert, Kirsch, Canny, etc. At present, the most representative one is Canny operator [3], which is a robust algorithm with dual threshold when the light changes and show a better advantage than other algorithms facing edge of the refinement. But double thresholds of the Canny algorithm are usually determined by many experiences, artificially set two fixed threshold. When image is changed, it is necessary artificially again to find a new threshold, so it is poor adaptability and low efficiency. Ref. [4] obtained a double threshold based on the class within the variance minimization principle of the gradient magnitude histogram. Ref. [5] detected the edge using the four threshold edge detection method. These methods only focus on the determination of the optimal threshold, but the effect of edge detection is not much improved. So, this paper detects the contour using adaptive threshold, the threshold range is determined based on the gradient magnitude histogram, and then the maximum of the variance of the class is derived from the adaptive threshold value.

This paper has two contributions for detecting contour of the rail end face, the one is to propose a new method to eliminate noise block based on mathematical morphology, defining a "※" shaped kernel. The second is to give a new definition of adaptive threshold, in order to make the contour detection effect better.

44.2 The Rail End Face Contour Detection Algorithm

In this part, we will introduce our algorithm in detail. The main steps of the proposed algorithm are shown in Fig. 44.1. The flow chart indicates the procedure of the method minutely: firstly obtain the image; and then wipe off the small noise with median filter; dislodge the larger-size noise using the morphological denoising

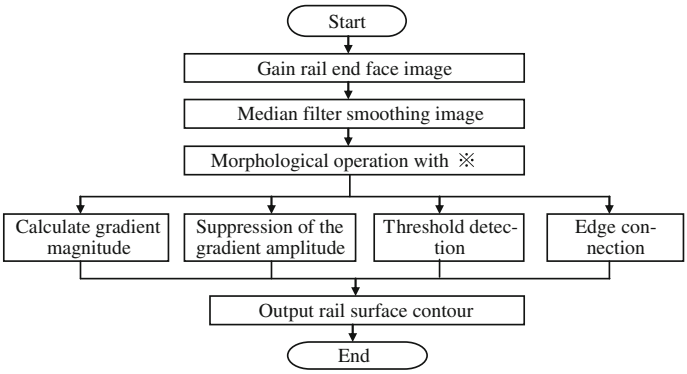


Fig. 44.1 Main steps of the proposed algorithm

operation based on "※" shaped kernel; extract the contour with improved adaptive threshold, which includes four steps.

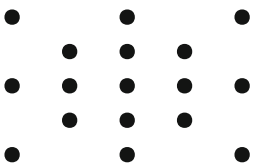
44.2.1 Median Filter Smoothing Image

Due to the interference of external factors, the image will have the noise, and smoothing image could wipe out the small noise. In this paper, we adopt the median filter [6, 7] which sorts the gray value of all the pixels (S totally) in the $N * N$ neighborhood, and takes the median value instead of the gray value of the target pixel. The result is shown in Fig. 44.3b.

44.2.2 Morphological Denoising Based on "※" Shaped Structural Elements

Using image smoothing can eliminate most of small noise, but due to the lighting angle and the environmental factors, image surface is not uniform and produce a wide range of noise. For larger area of noise, this paper introduces morphological [8] removal.

The key content of morphological operation is choosing appropriate structure elements, which is used to implement the input image morphology transformation of a certain size of the background image. At present, commonly used structural elements include cross, oval, and rectangle. There are four basic operations of mathematical morphology: corrosion, inflation, opening operation, and closing operation.



1	0	1	0	1
0	1	1	1	0
1	1	1	1	1
0	1	1	1	0
1	0	1	0	1

Fig. 44.2 "※" structure element and template representation

Fig. 44.3 The process of the algorithm, from *left* to *right* **a** input image with light source in the daytime, **b** median filter **c**, binary image, **d** morphological denoising with"※", **e** end face contour

With the experimental environment and the shape of the rail end face, through the experimental contrast, we design the "※" shaped structural element shown in Fig. 44.2. This structure element can be split into two 5 × 5 cross nuclear, one of which is the result of another 45° rotating, the "※" increased by 45°, 135°, 225°, 315° four directions, from the four directions of cross structural elements expand to eight direction. Compared with 3 × 3 rectangular structure element, the size of "※" is expansion and neighborhood pixels from 8 to 16, local information of the target pixel is more rich. Compared with the same size of 5 × 5 rectangular structure element of 24 pixels, the calculation of the "※" structure element is small and the efficiency is high. Through the above analysis, "※" structure element has a great advantage than the existing ones. The result after morphological denoising is shown in Fig. 44.3d. We will show the superiority of the "※" using experience in Sect. 44.3.2.

44.2.3 Contour Extraction Based on Adaptive Threshold

44.2.3.1 Non-maximum Suppression of the Gradient Amplitude

Calculating the magnitude and direction of the gradient using finite difference based on the first order partial derivative of the neighborhood. Then, traverse the gradient amplitude figure and along the gradient direction find two adjacent pixels as to each pixel gradient intensity value of the non-zero point, which is judged to be the candidate edge points, if greater than the adjacent pixel gradient amplitude along the gradient direction, otherwise it is not the edge points and set intensity value as zero.

The process of preserving the local maximum value of the gradient magnitude is non-maximum suppression, and the final goal is to generate a refinement of the edge.

44.2.3.2 Improved Otsu Algorithm to Determine the Threshold

The images after the maximum inhibition are set two thresholds for double threshold detection. The traditional methods determine the thresholds by artificial multiple trials to determine, this paper puts forward a method using improved Otsu algorithm. The Otsu method [9] can automatically determine the threshold and obtain better segmentation results, whose principle is simple and easy to understand. But the Otsu method's calculation amount is larger, and it is proved that when the size of targets and background in the image is very small, obtain the threshold tend to have larger error. So, this paper improves the Otsu method.

First, get the mean value U of gradient amplitude of the whole image according to the statistical histogram.

Second, take the average U as the boundary, find the left and right peaks Cl, Cr, and then calculate the average value Ca using the two peaks.

Third, find the minimum tr between the two peaks Cl Cr.

Finally, use Otsu seeking the best threshold decision method to find the class variance maximum between Ca and tr, which is set as the high threshold Hth, and the 0.4 * Hth is set as the low threshold Lth. The formulas are as following:

$$H\text{th} = \omega_0\omega_1(\mu_0 - \mu_1)^2 \tag{44.1}$$

$$L\text{th} = 0.4 * H\text{th} \tag{44.2}$$

where, ω_0 is the percentage of the foreground to the whole image segmented by threshold T; ω_1 is the percentage of the background to the whole image segmented by threshold T, and the two meet $\omega_0 + \omega_1 = 1$. μ_0 is the mean value of the foreground pixels, and μ_1 is the mean value of the background pixels.

Use the two thresholds to detect the non-maximum suppression image, if the gradient magnitude of the pixel point is greater than the high threshold Hth, the pixel is the edge point; if it is smaller than Lth, the pixel is not the edge point; when is less than high threshold Hth and lower than Lth, the pixel will be detected the connectivity with the edge point, if so, mark as an edge point.

44.2.3.3 Edge Connection

Link edge point based on the results obtained from Hth. First as to endpoint detection, if meet the endpoint which only has a neighboring point, then find the weak edge points from low threshold detection results in order to make up the edge of the gap [10]. So, the contour is complete and the result is shown in Fig. 44.3e.

44.3 Experiments

We evaluate the performance of method in win7 system, processor is Intel(R) Core (TM)2 Duo CPU T5800 @2.00 GHz 2.00 GHz, RAM is 2.00 GB. The codes are completed using Visual Studio 2010 and Opencv2.0. The experimental images are all from industrial camera.

44.3.1 Obtain the Contour Results

We compare our method with existing two popular algorithms, Canny and Otsu. The experiment maps are obtained, respectively, from without light source, with light source in the night and with light source in the daytime.

- Qualitative analysis

The comparison results are shown in Fig. 44.4. From the results, we easily conclude ours are best among the three methods. The proposed method is equipped with strong adaptability and denoising, implement simply and get better rail end face contour map.

- Quantitative analysis

We get the groundtruth of the rail end face segmented by several persons accurately. Calculate the T based on the formula: $T = (N_{\text{dst}} - N_{\text{src}})/N_{\text{src}} * 100\,\%$.

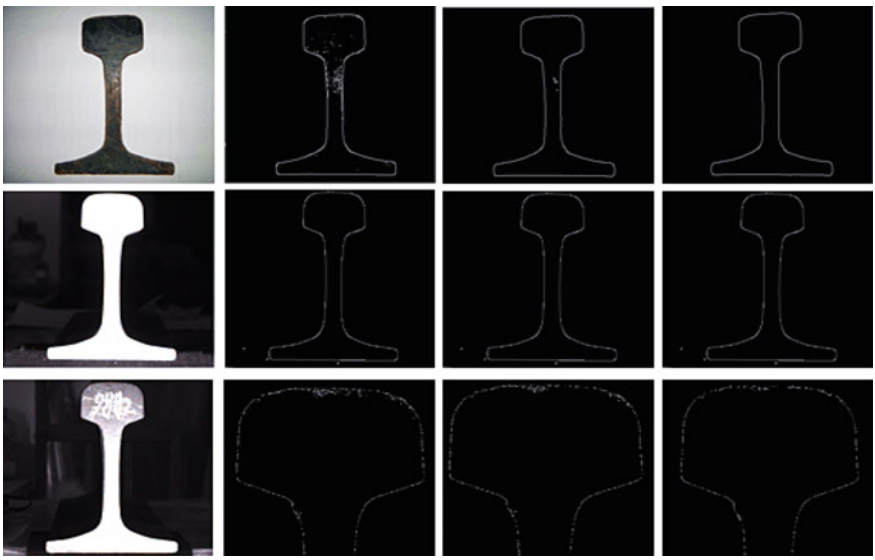


Fig. 44.4 The comparison of existing methods with ours on three images, from *left* to *right* **a** input image, **b** Canny, **c** Otsu, **d** Ours

Table 44.1 The results of different methods on three images

Images	Canny (%)	Otsu (%)	Ours (%)
a	14.29	4.124	3.960
b	4.884	0.535	0.535
c	32.443	4.518	4.215

Where, N_{dst} and N_{src} indicate the number of the edge point from our algorithm and groundtruth. Numerator represents the number of non-edge points in the edge map obtained by the algorithm, the less the number of non-edge points, the better the algorithm. So when T is smaller, the effect is better. The results from three methods on the different images are shown in Table 44.1.

44.3.2 Detail Experience

In this part, we will discuss the advantage of "※" shaped structural elements from the experience aspect.

We compare it with other existing structural elements, such as rectangle of 3 * 3 and cross of 5 * 5 to wipe off the noise in the binary image with light source in the daytime. The results are shown in Fig. 44.5 from the maps, we easily know the "※" shaped structural elements we proposed is effective.

In order to compare the inhibition degree of different structural elements more clearly and intuitively, we use the ratio of signal-to-noise (*snr*) to reflect the effect. If *snr* is larger, then the structural element has the lower inhibition ability to the

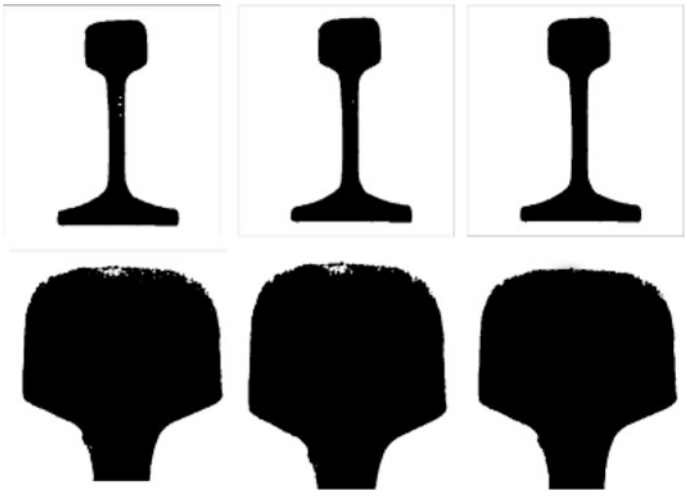


Fig. 44.5 The morphological denoising results of different structural elements, from *left* to *right* **a** rectangle of 3 * 3, **b** cross of 5 * 5, **c** "※"

Table 44.2 The *snr* of different structural elements

Structural element	Cross of 3 * 3	Rectangle of 3 * 3	Cross of 5 * 5	Rectangle of 5 * 5	Oval of 5 * 5	"※"
snr	39.181397	32.731180	14.372123	13.695593	13.695593	13.689715

noise, on the contrary, *snr* is small, structural element has the high inhibition ability to the noise and the denoising results are better. The results are shown in Table 44.2.

44.4 Conclusions

We presented a method of the rail end face contour detection based on mathematical morphology and adaptive threshold. Use median filter to smooth the source image and then remove the noise of the large blocks with the proposed "※" shaped structural elements; at last, detect the contour according to the improved Otsu theory. Compared with other methods, ours produces more perfect contour. But morphological denosing enlarges the number of the edge pixels, increased workload calculation to extract the contours. In the next work, we will optimize the algorithm and make it better application to practice.

Acknowledgment This research is supported by Hebei Province Science and Technology Support Program, China (no. 13211801D); partly supported by Yanshan University Dr. Fund projects (no. B540).

References

1. Zhu HT, Cao XC, Wang ZY (2013) Data processing of contact-type rail profile measurement. Railway Stand Des 7(26–28):32 (in Chinese)
2. Gao WJ (2012) Research of rail contour detection system based on machine vision. Beijing Jiaotong University, Beijing (in Chinese)
3. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 6(8):679–698
4. Li M, Yan JH, Zhu YH et al (2008) An application of improved Otsu in Machine Vision. Jilin Univ (Eng Sci) 38(4):913–918 (in Chinese)
5. Li H, Hu FM, Chu XQ (2008) An improved Canny edge detection algorithm. Microcomput Inf 24(12–3):298–299 (in Chinese)
6. Lin HM, Willson AN (1988) Median filters with adaptive length. IEEE Trans Circ Syst 35 (6):675–690
7. Qi L, Zhang W, Chen Q et al (2013) Research of EMCCD image filtering method based on noise detection and fuzzy adaptive median filter. Int Conf Opt Instrum Technol Optoelectron Imaging Process Technol 9045:2714–2739

8. Geng S (2012) Image denoising based on mathematical morphology. Shandong Normal University, Jinan, pp 34–35 (in Chinese)
9. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybernet 9(1):62–66
10. Coleman SA, Scotney BW, Suganthan S (2011) Multi-scale edge detection on range and intensity images. Pattern Recogn 44(4):821–838

Chapter 45
A Queuing Simulation Research of Fare Collection Equipments in Xi'an Subway Longshouyuan Station

Guang Li, Meng Xi and Lu Ni

Abstract Automatic fare collection system is one of the important urban rail transit infrastructures. It is closely related with passenger experience. It is important to determine the appropriate number of fare collection equipments, because too many equipment will bring waste and too little equipment will bring congestion. Always, it is needed to determine the appropriate number of fare collection equipment through simulation method. This paper developed queuing simulation software for station fare collection equipment with R language, and used it to simulate the situation of Longshouyuan station of Xi'an subway. The simulation result gives suggestion for the appropriate number of fare collection equipment in Longshouyuan station.

Keywords Urban rail transit · Fare collection equipment · Queuing simulation

45.1 Introduction

Over the past decade, the urban mass transit has been developed very fast in the China Mainland. In nearly twenty cities, urban mass transit is operating. And more cities decided to build their urban mass transit system. Every year, hundreds of kilometers of subway lines have been built [1].

In the modern urban rail transit system, the AFC (Automatic Fare Collection) system [2, 3] is the essential infrastructure. The AFC system is based on computer, communications, networking and automatic control technology. It can automate complete operations in rail transportation system, like ticketing, billing, charging, statistics, management, and so on. The AFC system can be divided to two parts, one is the computer system, and the other is the AFC station equipment. In the urban rail transit systems in China Mainland city, like Xi'an or Shanghai, the AFC station equipments mainly include automatic ticket vending machines and the automatic ticket gate machines.

G. Li (✉) · M. Xi · L. Ni
School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China
e-mail: hit6006@126.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_45

The number of the AFC station equipments should be studied carefully. If setting too many AFC station equipments, the equipment utilization will be too low, and will waste investment and increased maintenance costs. But if setting too few, it will produce passenger congestion. It will increase the time spent on buying ticket of passengers, giving passengers a bad experience and bring great security risk. So, it is very important to decide how many the AFC station equipments should be set for every station.

The reasonable number of the AFC station equipments is closely related with the station passenger traffic. It is a queuing system and always need be studied by using computer simulation [4, 5].

A computer simulation is a simulation, run on a single computer, or a network of computers, to reproduce behavior of a system. The simulation uses an abstract model to simulate the system. The computer simulation is a useful tool for analysis of complex systems. It has wide applications in environmental ecology [6], physics [7], electronic engineering [8] and other fields.

For the queuing system, the computer simulation always uses the Matlab platform or specialized simulation software like Arena. The queuing can be used for aviation [9], traffic [10], logistics [11] and other fields. This paper designed and implemented a queuing simulation software for the AFC station equipment using R platform [12, 13], and applied it for Xi'an subway Longshouyuan station. According to the simulation result, we give our recommendations for the setting of the AFC station equipment in Xi'an subway Longshouyuan station.

The rest of our paper is organized as follows. Section 45.2 describes the AFC station equipments. Section 45.3 describes the queuing simulation software designed and implemented by this paper. Section 45.4 describes the simulation result when using this software for Xi'an subway Longshouyuan station. Section 45.5 is consists of the conclusions.

45.2 Analysis of the AFC Station Equipment

In China Mainland, the AFC station equipment mainly includes two kinds of device. One is the automatic ticket vending machine, and the other is the automatic ticket gate machine. The automatic ticket gate machine can be divided into inbound ticket gate and the outbound ticket gate. And in China Mainland the security checking apparatus is widely used in the urban mass transit system.

Figure 45.1 shows the inbound process of the Xi'an subway. Almost in all the cities in China Mainland, the subway system has the similar inbound process as Xi'an subway. First, passengers carrying luggage have to accept security checking, and the other passengers generally do not need it. Then, the passengers having stored value cards can accept ticket checking directly and inbound. And the passengers without stored value cards have to buy ticket on the ticket vending machines and then accept ticket checking.

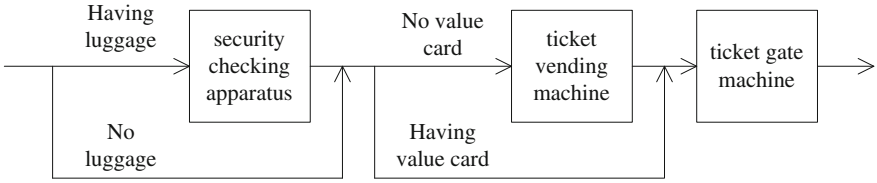


Fig. 45.1 Inbound process

In the inbound process, there are three possible queues, which, respectively, are at the security checking apparatus, ticket vending machines, and inbound ticket gate machines.

The outbound process is much simpler than the inbound process. The passengers get off the train and accept ticket checking directly. There is only one possible queue, which is at the outbound ticket gate machine.

The outbound passenger traffic is typical burst passenger traffic. Only when the train stops at the station platform, many passengers will get off and use the ticket gate machine. Most of the time, there is no outbound passenger. So, at the outbound ticket gate, the congestion and idle are alternating.

The inbound passenger traffic changes more relaxed. In weekdays, in early morning, only few passengers and there is always no queues at the AFC station equipment. Then, more people want to use the subway. And it reaches a peak when people go to work. At this moment, we will have longest queues. After the peak, the passengers will turn fewer, and the queues will turn shorter and then disappear. When people go home from work, another peak will appear, and the queues appeared again at the AFC station equipment.

45.3 Design and Development of the Simulation Software

This paper develops the queuing simulation software on R platform. R is a programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners. R is an implementation of the S programming language and is freely available under the GNU General Public License. Compared with MATLAB, which is widely used for queue simulation, R has not bad function and is free.

A queuing system mainly includes three parts: input process, queuing discipline and service law of the help desk. The input process is the process of that the customer arrive queuing system. In this paper, the input process is gotten by field observations. The queuing discipline is the ways and rules of the customer waiting for the service when they arriving queuing system. This paper uses the waiting system. That means if the customer arrives queuing system, he will wait until the help desk is free. Then, he goes to the free desk and receives services. This paper

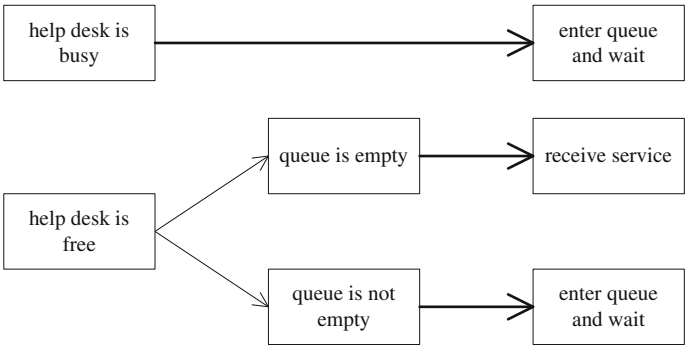


Fig. 45.2 The code of conduct for a newly arrived passenger

uses the FCFS (first come first served) criterion. The service law of the help desk is also gotten by field observations in this paper. Single desk and multi desk both can be simulated in this paper.

In the case of the single desk, there is only one queue. As Fig. 45.2 shows, for a newly arrived passenger, he should go to the help desk to receive service if the help desk is free and the queue is empty. In other cases, a newly arrived passenger will enter the queue and waiting. For the passengers in the queue, if he is the head of the queue and the help desk is free, he should go to the help desk to receive service and be out of the queue. In other cases, he just waits. For the help desk, it will check the queue each time finishing once service. If the queue is empty, it will wait. Otherwise, it will provide services to the head of the queue. In the whole simulation process, the pure waiting system is used. The passengers will wait until they receive service, and will not go away if they wait too long.

In the case of the multi desk, there is a plurality of queues. The basic rules for this case are the same as the single desk case. It should be noted that, for a newly arrived passenger, if he cannot receive service at once, he will choose the shortest queue to enter and wait, and he will not change to another queue. If a passenger has more than one help desk that can be used at once, he will select one of them randomly.

45.4 Data and Simulation Results

The measured data is come from the Longshouyuan station of Xi'an subway. The observation time is the morning peak period of April 17, 2014. In the morning peak, the stations will have the largest passenger traffic and the longest queue at the AFC equipment. The outbound passenger traffic is typical burst passenger traffic. And in Longshouyuan station, the outbound ticket gate is often free. The queues only

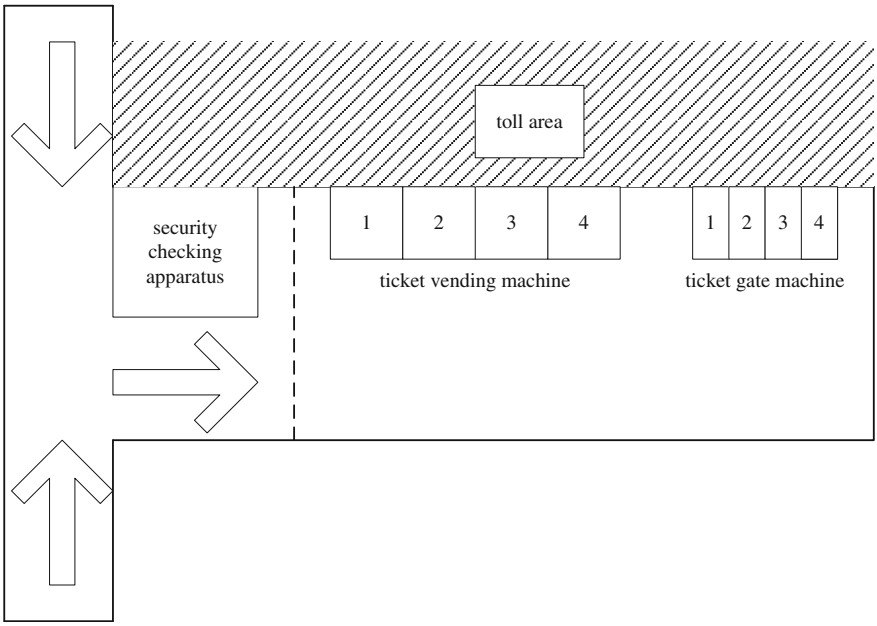


Fig. 45.3 The setting of the AFC station equipments of the A–B entrances in the Longshouyuan station

appear when the train stops on the station. So, we only observe the inbound passenger traffic and only simulate the inbound AFC station equipment.

Figure 45.3 shows the plan view of the A–B entrances in the Longshouyuan station of Xi'an subway. There are one security checking apparatus, four automatic ticket vending machines and four inbound ticket gate machines.

Field observations found the inbound ticket gate machines have enough passing capacity. Even in the morning peak, queues appear scarcely at them. It also can be found that a serious shortage of the security checking apparatus. In the morning peak, there is a queue about thirty persons at the security checking apparatus. But in the Longshouyuan station, the passage is barrow. And there is no space for another security checking apparatus. So, the research focuses on the automatic ticket vending machines.

This paper first observes the passenger traffic through the dotted line in Fig. 45.3. The observation time is from the 8:00 am to the 8:36 am. And data is recorded once every three minutes. We also observe if the passenger using the stored value cards. Figure 45.4 shows the observation result. In Fig. 45.4, the i-th set of data corresponding to the time from the 8:(3i-3) am to the 8:3i am. Because the passengers arrive very intensive in the morning peak, we can assume in every data set, the passengers are uniform arrives.

We also observe the spending time for every service of the automatic ticket vending machines. Figure 45.5 shows the observation result. It shows that most of

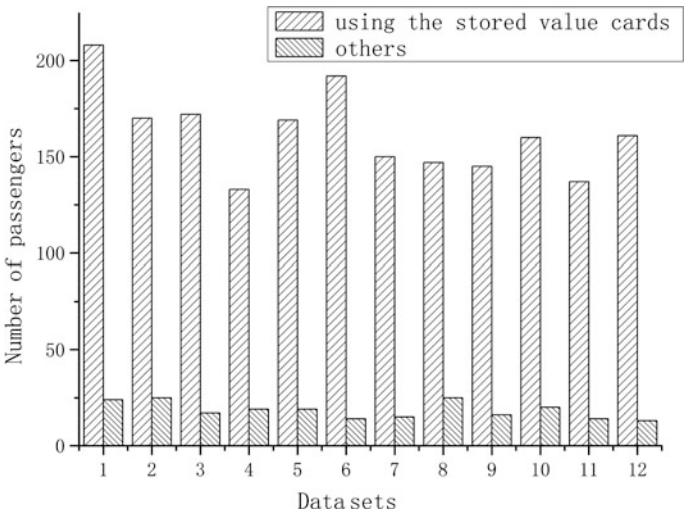


Fig. 45.4 The number of the inbound passengers

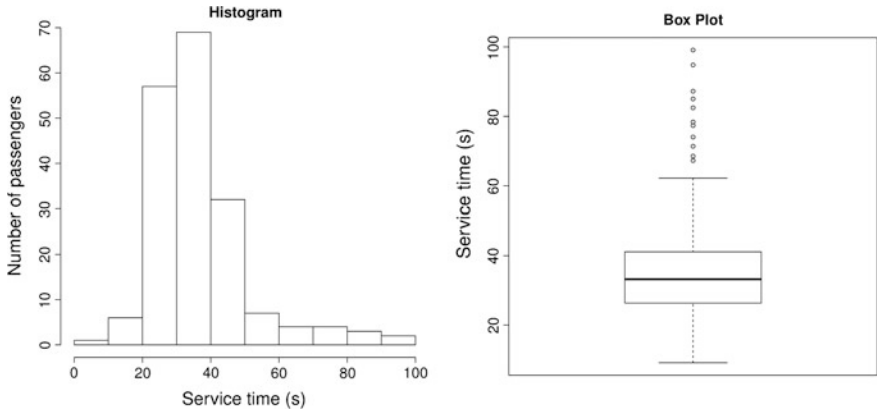


Fig. 45.5 The serving time of the automatic ticket vending machine

the passengers can finish the operating in one minute, but part of the passengers need much more time.

The simulation result is shown in Table 45.1. If holding four automatic ticket vending machines, about 80 % of the passengers need to wait and the average waiting time is 1.52 min. If using 5 automatic ticket vending machines, the situation will be greatly improved. Only 40 % of the passengers need to wait and the average waiting time is 0.16 min. If using six automatic ticket vending machines, the average waiting time further decline to 0.04 min.

Table 45.1 The simulation result

The number of ticket vending machine	The rate of passengers needs to wait (%)	The average wait time (min)
4	81.3	1.52
5	40.4	0.16
6	21.5	0.04

Considering it is the morning peak, this paper thinks it is more appropriate to add 1 automatic ticket vending machine. That means set five automatic ticket vending machines. Waiting 1.52 min is too long but waiting 0.16 min is not serious in the morning peak. And it is no need to set six automatic ticket vending machines to reduce waiting time to 0.04 min.

In summary, by simulation of the Longshouyuan station of Xi'an subway, this paper suggests to add one automatic ticket vending machine for the A–B entrances in this station.

45.5 Conclusion

The AFC station equipment is critical Infrastructure in the modern urban rail transit system. The number of the AFC station equipments should be studied carefully. If setting too many AFC station equipments, investment is wasted. If setting too few, it will produce passenger congestion and bring great security risk. The reasonable number of the AFC station equipments always need be studied by using computer queuing simulation.

This paper design and develop queuing simulation software for the AFC station equipments on the R platform. And this paper applies this software for the Longshouyuan station of Xi'an subway and gives suggestions for the setting number of the AFC station equipments.

Acknowledgments This work is supported by the National Nature Science Foundation of China (No. 51407012) and the Chang'an University's Fundamental Research Funds for the Central Universities (0009-2014G6114024).

References

1. Min G (2011) Status of urban rail transit development and prospects. Chin Railways 2011 (10):53–56 (in Chinese)
2. Zhu W, Hao H, Huang Z (2014) Calibrating rail transit assignment models with genetic algorithm and automated fare collection data. Comput-Aided Civil Infrastruct Eng 29 (7):518–530

3. Jiang Y, Luo N (2013) Urban rail transit station automatic fare collection configuration research and simulation. Proc ICTE 2013:1142–1149
4. Law A (2014) Simulation modeling and analysis, 5th edn. McGraw-Hill Education, New York
5. Ross SM (2012) Simulation, 5th edn. Elsevier Academic Press, Waltham
6. Yun SM, Kang CS, Kim J, Kim HS (2015) Evaluation of soil flushing of complex contaminated soil: an experimental and modeling simulation study. J Hazard Mater 287:429–437
7. Densmore JD, Park H, Wollaber AB, Rauenzahn RM, Knoll DA (2015) Monte Carlo simulation methods in moment-based scale-bridging algorithms for thermal radiative-transfer problems. J Comput Phys 284:40–58
8. Marlier C, Videt A, Idir N (2015) NIF-based frequency-domain modeling method of three-wire shielded energy cables for emc simulation. IEEE Trans Electromagn Compat 57 (1):145–155
9. Zhang J, Minghua H, Zhenya W (2013) Airspace capacity evaluation of terminal area based on flight queuing simulation. Aeronaut Comput Tech 43(4):9–13 (in Chinese)
10. Zhang W, Li D, Luo W (2013) Simulation for speed spatial distribution and queuing characteristics on traffic accident section of freeway. J Syst Simul 1:158–163 (in Chinese)
11. Ma X, Qin X, Cheng C (2012) Simulation optimization based on production of queuing theory of logistics system. Comput Eng Appl 48(20):244–248 (in Chinese)
12. Matloff N (2011) The art of R programming: a tour of statistical software design. No Starch Press, San Francisco
13. Kabacoff R (2015) R in action: data analysis and graphics with R, 2nd edn. Manning Publications, Greenwich

Chapter 46
The Vehicle Distance Measurement System Based on Binocular Stereo Vision

Yongjie Li, Peijiang Chen and Mei Zhang

Abstract Vehicle distance measurement is the basis of intelligent navigation and secondary navigation which plays an important role in traffic safety. Based on the machine vision technology, a vehicle distance measurement system is designed and realized. Two digital images of the tested vehicle are obtained by binocular camera. On the basis of camera calibration, the technologies of feature point detection and matching are used. According to the principle of parallax, the three-dimensional geometric information of the vehicle's feature point is reconstructed and the distance between vehicles is obtained. The system can provide a basis for visual navigation and vehicle monitoring, improve road safety, and reduce traffic accidents which has certain practical value.

Keywords Binocular stereo vision · Vehicle distance measurement · Camera calibration · Corner detection · Image matching

46.1 Introduction

With the rapid increase of car ownership and driving speed, the traffic accident rate is getting higher and the issue of traffic safety is more prominent. The vehicle distance measurement technique can ensure traffic safety and improve the intelligence level of vehicles. In critical situations, through an alarm or automatically pre-set operation,

Y. Li · P. Chen (✉)
School of Automobile, Linyi University, Shuangling Road, Linyi 276000, Shandong, China
e-mail: chenpeijiang@163.com

M. Zhang
Library, Linyi University, Shuangling Road, Linyi 276000, Shandong, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_46

such as emergency braking, some accidents which are caused by the drivers' fatigue or negligence can be avoided based on the measured vehicle distance [1, 2].

Today, the commonly used vehicle distance measurement methods mainly include ultrasonic ranging, millimeter-wave radar ranging and laser ranging, etc. These methods have their advantages and disadvantages.

The technology of vision measurement is to use image sensors to detect the three-dimensional coordinates of certain space object and then get the size, shape, and motion of the tested object [3]. It originated from the computer vision technology in the 1950s. After decades of research and development, various theories and techniques become more mature and perfect. Now, many techniques about machine vision are really applied to people's daily life, such as production automation, vehicle identification [4], and so on.

Based on the technologies of binocular stereo vision technology and image processing, a vehicle detecting and distance measurement system is designed. The principle of binocular stereo vision is analyzed and the corner detection and matching methods are used to compute the actual vehicle distance, which can provide essential information to vehicle safe driving.

46.2 Design of Vehicle Distance Measurement System

The binocular stereo vision system can measure the distance by the simulation of human eyes. Two images of the same object are captured by using two cameras, and the coordinates of certain points can be obtained by corner extraction and matching, and then the distance between cameras and target can be obtained.

46.2.1 Principle of Binocular Stereo Vision Measurement

In binocular stereo vision, using two cameras simultaneously at different locations to shoot two pictures of the same target vehicle [5]. According to the coordinates of same points in their two image coordinate systems, based on feature points detection and matching, the coordinates of certain points in the world coordinate system can be obtained. The principle of three-dimensional measurement is given, as shown in Fig. 46.1.

Assuming that the left camera coordinate system $oxyz$ is made as the world coordinate system, the left image coordinate system is $X_lO_lY_l$, and the effective focal length is f_l; the right camera coordinate system is $o_rx_ry_rz_r$, the corresponding image coordinate system is $X_rO_rY_r$, and the effective focal length is f_r. The point

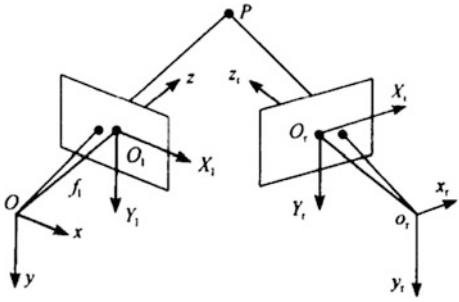


Fig. 46.1 The spatial point reconstruction of binocular stereo vision

P is target, according to the camera perspective transformation model, the expression can be obtained as in

$$s_l \begin{bmatrix} X_l \\ Y_l \\ 1 \end{bmatrix} = \begin{bmatrix} f_l & 0 & 0 \\ 0 & f_l & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad s_r \begin{bmatrix} X_r \\ Y_r \\ 1 \end{bmatrix} = \begin{bmatrix} f_r & 0 & 0 \\ 0 & r_l & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} \tag{46.1}$$

The mutual positional relationship between the world coordinate system and the right camera coordinate system can be obtained as in

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{46.2}$$

$$M = [R|T]$$

where, $R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$, $T = \begin{bmatrix} t_x \\ t_y \\ t_y \end{bmatrix}$, they are the rotation matrix and translation transform vector between the world coordinate system $oxyz$ and the right camera coordinate system $o_r x_r y_r z_r$ respectively.

For the spatial point in the world coordinate system, the correspondence of a certain point in the two image plane can be obtained as in

$$\rho_r \begin{bmatrix} X_r \\ Y_r \\ 1 \end{bmatrix} = \begin{bmatrix} f_r r_1 & f_r r_2 & f_r r_3 & f_r t_x \\ f_r r_4 & f_r r_5 & f_r r_6 & f_r t_y \\ r_7 & r_8 & r_9 & t_z \end{bmatrix} \begin{bmatrix} z\frac{Y_l}{f_l} \\ z\frac{Y_l}{f_l} \\ z \\ 1 \end{bmatrix} \tag{46.3}$$

Thus, the three-dimensional coordinate can be obtained as in

$$\begin{cases} x = z\dfrac{X_l}{f_l}, \quad y = z\dfrac{Y_l}{f_l} \\ z = \dfrac{f_l(f_r t_x - X_r t_z)}{X_r(r_7 X_l + r_8 Y_l + r_9 f_l) - f_r(r_1 X_l + r_2 Y_l + r_3 f_l)} \\ \quad = \dfrac{f_l\left(f_r t_y - Y_r t_z\right)}{Y_r(r_7 X_l + r_8 Y_l + r_9 f_l) - f_r(r_4 X_l + r_5 Y_l + r_6 f_l)} \end{cases} \tag{46.4}$$

Therefore, if the focal lengths f_l,f_r and the coordinates of left image and right image of the spatial point are known, as long as rotation matrix R and translation transform vector T are obtained, the three-dimensional coordinates of the object point can be obtained.

46.2.2 *Vehicle Distance Measurement Based on Binocular Stereo Vision*

In the vehicle distance measurement system, it is based on machine vision technology; the two images of the tested vehicle are captured and processed. Using the binocular vision technology, the distance between two vehicles can be computed.

The realization process of vehicle distance measurement is given, as shown in Fig. 46.2.

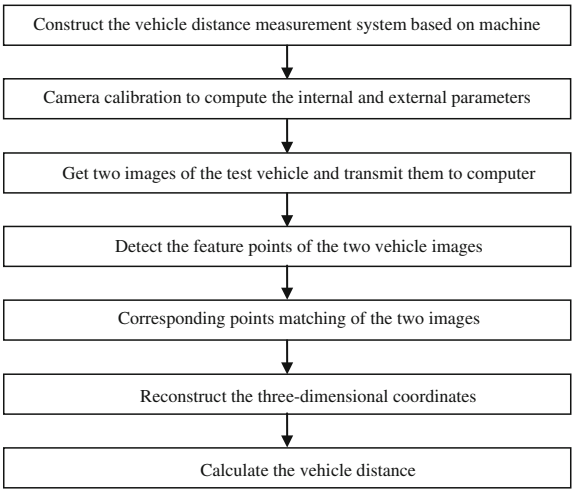


Fig. 46.2 The process of vehicle distance measurement based on binocular stereo vision

46.3 Realization of Vehicle Distance Measurement System

46.3.1 Camera Calibration

The camera calibration means to build the relationship between image pixel position and actual scene location of certain feature points. Its approach is to analyze the model parameters of cameras by image coordinates and world coordinates of some known feature points based on camera model [6, 7]. The model parameters of camera are divided into two parts: internal parameters and external parameters. The camera imaging model has two types: linear model and nonlinear model. Wherein, the linear camera model is based on the principle of pinhole imaging, and the nonlinear camera model adds the distortion amendment on the basis of linear model which improves the camera calibration accuracy effectively.

Tasi gives a two-step calibration method based on radial constraint. The first step is to use the least square method to solve an over-determined linear equation, so the external parameters can be obtained. The second step is to solve the internal parameters. If the camera is not distorted, they can be solved by an over-determined linear equation, but if there are radial distortion in the camera, these parameters can be obtained by combining the nonlinear optimization method. The method has moderate calculating amount and high precision. In this project, in order to compute the internal and external parameters, the two cameras are calibrated by the two-step method.

46.3.2 Corner Detection of Vehicle Images

In the process of computer image processing, the corners can reflect the essential characteristics of the original image [8]. There is no exact definition of corner. In general, the corner means a point where the brightness changes strongly. As the important feature of original image information, the corner can effectively reduce the data amount and the corner detection is particularly important in image processing and machine vision.

Harris algorithm extracts corners based on gray scale information of image pixel. The method has excellent detection result and it is used widely. It uses autocorrelation matrix and computes the threshold value to decide a corner. In this project, the Harris method is adopted to detect the image corners of the tested vehicle.

46.3.3 Stereo Matching of Vehicle Images

The stereo matching is the key technology to machine vision application system. It is to get the correct corresponding feature points in the two images of the same

target [9, 10]. On the basis of corner detection and image matching, the vision system can quickly get the coordinates of the object in the world coordinate system.

In order to improve the accuracy rate of the image matching, certain constraints are used. For the detected corners by the Harris algorithm, the gray scale similarity is used to construct the initial matching relation. The matching method based on gray scale distribution is to use correlation function to evaluate the similarity of the two images and to determine the corresponding feature points. The parallax theory is used to remove some false matched corners. According to the camera calibration, the fundamental matrix is found to eliminate some false matching. Then, the final corresponding corners in the two images of the tested vehicle can be obtained.

46.3.4 Measurement of Vehicle Distance

In the vehicle distance measurement system, based on binocular stereo vision, two images of the same target vehicle are captured simultaneously at different locations by two cameras. Based on the camera calibration technology, Harris corner detection algorithm, stereo matching technology, and the three-dimensional reconstruction theory, the vehicle distance can be computed.

In the actual test, the two cameras are set with a moderate angle, the left and right images of the tested vehicle are captured, and the corners are detected and matched, as shown in Fig. 46.3.

According to the matched corners and the results of camera calibration, the vehicle distances are measured, as shown in Table 46.1.

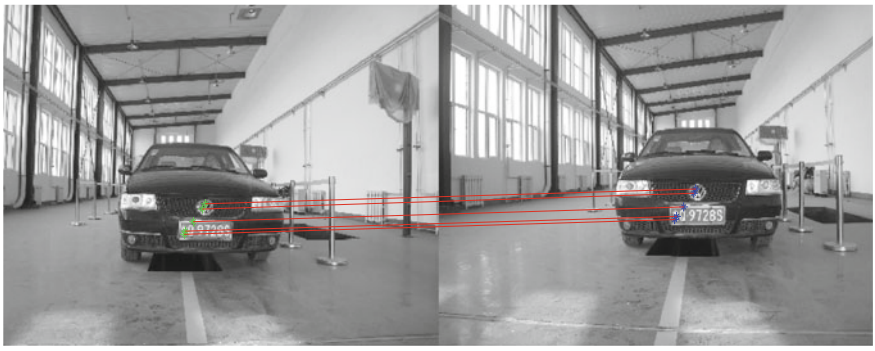


Fig. 46.3 The matched corners of the tested vehicle images

Table 46.1 The measurement results of vehicle distance

No.	Coordinates in left camera image	Coordinates in right camera image	World coordinates	Vehicle distance/m	Actual distance/m	Relative error (%)
1	(1194.4763, 1165.0427)	(1525.8455, 1079.4907)	(−180.31609, −202.23938, −2434.7925)	2.4498	2.454.9	−0.21
2	(1091.0814, 1290.8289)	(1424.7415, 1206.8273)	(−4.4261856, −352.28821, −2430.7969)	2.4562	2.454.9	0.05
3	(1206.764, 1138.7637)	(1538.672, 1052.4248)	(−217.42348, −186.19931, −2450.8438)	2.4675	2.454.9	0.51
4	(1135.3666, 1240.5409)	(1467.2797, 1156.5664)	(−74.991882, −286.10071, −2421.0759)	2.4391	2.454.9	−0.64
5	(1081.2616, 1309.0554)	(1415.4536, 1225.1615)	(21.507919, −367.56281, −2434.4548)	2.4621	2.454.9	0.29

46.4 Conclusions

The binocular stereo vision system has the special function of simulating human visual system. It can get outside information effectively and quickly, and it is an online, non-contact information exchange platform. Vehicle distance measurement is important in intelligent vehicle and safe driving. The vehicle distance measurement system based on machine vision is designed and realized. By the corner detection and stereo matching for the two images of the tested vehicle, based on the camera calibration, the three-dimensional information of the target vehicle is reconstructed. The vehicle distance measurement model is established based on principle of parallax. The system can automatically measure the distance between two vehicles to provide driving information for drivers and ensure driving safety.

Acknowledgment This study is sponsored by the National Undergraduate Training Programs for Innovation and Entrepreneurship (No. 201410452031) and the Science and Technology Development Planning Project of Shandong Province (No. 2013YD05004), People's Republic of China.

References

1. Zhang F, Zhou J, Xia W (2011) Research of vehicle security spacing measurements based on binocular stereo vision. CAAI Trans Intell Syst 6(1):79–84
2. Wu Y, Li YJ, Liu JX, Wei LF, Yin XQ (2010) Measurement of safety space between two vehicles based on machine vision. Mach Eng Autom 3:67–68 (in Chinese)
3. Fang Q (2009) The apparent quality inspection system design based on machine vision artificial marble deck. Nanjing University of Aeronautics and Astronautics, Nanjing (in Chinese)
4. Chen P, Min Y (2014) Automobile longitudinal axis detection method based on image segmentation and preliminary results. Jordan J Mech Industr Eng 8(5):297–303
5. Yu HY, Zhang WG (2012) Vehicle distance measurement and its error analysis based on monocular vision sensor. Transducer Microsyst Technol 31(9):11–13 (in Chinese)
6. Lin J, Huang C, Liu B, Jiang K (2011) Binocular stereo vision camera calibration and precision analysis. Huaqiao University 32(4):364–367 (in Chinese)
7. Liu J, Yun S, Zhang Q, Liu D (2011) Research on binocular stereo vision camera calibration technology. Comput Eng Appl 33(6):237–239 (in Chinese)
8. Zhang Y, Yu J, Sun J (2011) Matching algorithm based on harris's corner. Comput Mod 11(2):132–136 (in Chinese)
9. Wang X, Wang Z, Wu F (2009) Harris correlation and feature matching. Pattern Recog Artif Intell 22(4):504–513 (in Chinese)
10. Bai M, Zhuang Y, Wang W (2008) Progress in binocular stereo matching algorithms. Control Decis 23(7):721–729 (in Chinese)

Chapter 47
Train Stops Setting Based on a Quantum-Inspired Particle Swarm Algorithm

Xuelei Meng, Limin Jia, Yong Qin and Jie Xu

Abstract To improve the train line plan quality and meet more transportation requirements, a model is presented to solve the train stops setting problem. We analyze the factors on the train setting problem and define the passenger transport efficiency. Then, an optimization model to improve the transport efficiency is constructed. The quantum particle swarm optimization algorithm is hired to solve the problem. Computing case based on Shanghai–Hangzhou high-speed railway proved the rationality of the model and the high performance of the algorithm. It is a new approach to design train stop plans which also offers constructive support for the managers of the railway bureau.

Keywords Train stops setting · Quantum-inspired · Passenger transport efficiency · Particle swarm algorithm

47.1 Introduction

Train planning plays a critical role in operating and managing railroad systems. Line planning is the tactical step of the whole planning process, which follows the basic step–demand estimation and capacity calculation. And the line planning problem can be divided into three steps, train pathing, train frequency setting, and train stops setting.

Chang et al. develops a multiobjective model for the track utilization optimization [1]. Barrenaa et al. built a train operation chart design model, considering the dynamic passenger demand [2]. Lin and Ku proposed a linear programming

X. Meng
School of Traffic and Transportation, Lanzhou Jiaotong University, No. 88 Anning West Road, Lanzhou, Gansu Province, China

L. Jia (✉) · Y. Qin · J. Xu
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, No. 3 Shangyuancun, Beijing, China
e-mail: jialm@vip.sina.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_47

model to design the line plan and developed an advanced algorithm to improve the computational accuracy [3]. Finger et al. summarized the advantages and challenges of implementation of integrated regular interval timetables (IRIT for railway passenger services and derived the main requirements for the successful introduction of IRIT [4]. Meng et al. proposed an approach to generate line plan in emergencies for railway networks to complete the passenger transportation [5]. Park et al. addressed two line planning problems and presented a genetic optimization algorithm for it [6]. Goossensa et al. introduced a series of models to solve line planning problems, considering the several halting patterns [7]. Mandl tackled the line plan and proposed a heuristic algorithm to construct a service network [8].

In this paper, we focus on the train stops setting problem on the high-speed railway. In the first section we introduce the related works and the train stops setting problem. In the second section, we build a mathematical model to describe train setting problem. And in the third section, we put forward the quantum particle optimization algorithm to solve the model. In the fourth section, we give a computing example and does some analysis work. In the last section, we draw the conclusion.

47.2 Train Stops Setting Model

Today, the most important goal in the railway transportation is to improve the transport speed with the development of high-speed railway. The quality of a service plan can be evaluated by two indices, transferring time and summary traveling time. The goal to reduce the traveling time is conflicting incompatible with the goal to reduce the summary transferring time, which is related to the number of the stops of the trains. Here, we consider the transport efficiency as the optimization goal, which is defined as

$$V = \frac{\sum_{j=1}^{N-1} \sum_{i=1}^{j} n_{ij}}{\sum_{k=1}^{M} t_k} \tag{47.1}$$

where n_{ij} is the number of passengers from station i to station j. t_k is the consumed time of train k on the high-speed railway. N is the number of the stations on the railway line. M is the number of the trains.

$$t_k = t_{\text{travel}} + \sigma(t_{\text{dwell}} + \Delta_{\text{start}} + \Delta_{\text{stop}}) \tag{47.2}$$

where t_{travel} is the actual running time of train k; if it runs without any stop. σ is the number of the stops. t_{dwell} is dwelling time of a train. We assume that the all the dwelling trains all the stops are equal. Δ_{start} and Δ_{stop} are the additional time consumed at the train starting process and at the stopping process.

In the passenger transportation, we must transport every passenger to his destination. So, the constraint is to meet all of the OD flows.

$$\sum_{i=1}^{K} \xi_{ij} \varphi_{ij} L_k \geq n_{ij} \tag{47.3}$$

where ξ_{ij} is a flag variable. If a train stops at both station i and station j, ξ_{ij} is 1; otherwise ξ_{ij} is 0. φ_{ij} is a parameter which indicates train k can reserver for n_{ij}. L_k is the seating capacity of train k.

47.3 Quantum PSO for Train Stops Setting Model

47.3.1 Quantum Particle Optimization Algorithm

The quantum particle swarm algorithm is different from the typical particle swarm algorithm in that the positions of the particles are determined by the different methods. First, the original positions of the particles are determined by the probability amplitude of the quantum.

$$Q_i = \begin{bmatrix} \cos\theta_{i1} & \cos\theta_{i2} & \cdots & \cos\theta_{iN} \\ \sin\theta_{i1} & \sin\theta_{i2} & \cdots & \sin\theta_{iN} \end{bmatrix} \tag{47.4}$$

$\theta_{ij} = 2\pi \times \text{rnd}$, rnd is a random number which is between 0 and 1. $1 < i \leq m$. i is the particle number. $1 \leq j \leq n$. j is the dimensionality of a particle. If the argument is assigned, the initialization of the particles is completed.

It is obvious that the initial values of the particles are between 0 and 1, which has many limitations in the real computation. The initial values of the particles positions must be transferred into the feasible solution. Set $[L_j, U_j]$ to be the range of the jth dimension of a particle, where $L_j < U_j$. The jth quantum position of particle Q_i is $Q_i = \begin{bmatrix} \cos\theta_{ij} \\ \sin\theta_{ij} \end{bmatrix}$. Then, the jth quantum position in the solution space can be described as

$$S_{gs}^{j} = \frac{1}{2}[b_j(1 + \cos(\theta_{ij})) + a_j(1 - \cos(\theta_{ij}))] \tag{47.5}$$

$$S_{es}^{j} = \frac{1}{2}[b_j(1 + \sin(\theta_{ij})) + a_j(1 - \sin(\theta_{ij}))] \tag{47.6}$$

Then, the ground state $|0\rangle$ corresponds to the solution of S_{gs}^{j} and the excitation state $|1\rangle$ correspond to the solution of S_{es}^{j}.

Thus, the particle positions can be changed by changing the arguments of the quantum. And the deciding variables can be adjusted.

Set Q_{il} to be the current best solution of the optimization problem and Q_g to be the global best solution of the optimization problem.

$$Q_{il} = (\cos(\theta_{il1}), \cos(\theta_{il2}), \ldots, \cos(\theta_{ilN})) \tag{47.7}$$

$$Q_g = (\cos(\theta_{g1}), \cos(\theta_{g2}), \ldots, \cos(\theta_{gN})) \tag{47.8}$$

Then, the change of argument of particle Q_i can be described as

$$\Delta\theta_{ij}(t+1) = \omega\Delta\theta_{ij}(t) + c_1(\Delta\theta_l) + c_2(\Delta\theta_g) \tag{47.9}$$

where

$$\Delta\theta_l = \begin{cases} 2\pi + \theta_{ilj} - \theta_{ij}(\theta_{ilj} - \theta_{ij} < -\pi) \\ \theta_{ilj} - \theta_{ij}(-\pi \le \theta_{ilj} - \theta_{ij} \le \pi) \\ \theta_{ilj} - \theta_{ij} - 2\pi(\theta_{ilj} - \theta_{ij} > \pi) \end{cases} \tag{47.10}$$

$$\Delta\theta_g = \begin{cases} 2\pi + \theta_{gj} - \theta_{ij}(\theta_{ilj} - \theta_{ij} < -\pi) \\ \theta_{gj} - \theta_{ij}(-\pi \le \theta_{gj} - \theta_{ij} \le \pi) \\ \theta_{gj} - \theta_{ij} - 2\pi(\theta_{gj} - \theta_{ij} > \pi) \end{cases} \tag{47.11}$$

After a derivation process, the new positions of particle θ_i are as follows:

$$\theta_{ig} = (\cos(\theta_{i1}(t) + \Delta\theta_{i1}(t+1)), \cos(\theta_{i2}(t) + \Delta\theta_{i2}(t+1)), \ldots, \cos(\theta_{iN}(t) + \Delta\theta_{iN}(t+1))) \tag{47.12}$$

$$\theta_{ie} = (\sin(\theta_{i1}(t) + \Delta\theta_{i1}(t+1)), \sin(\theta_{i2}(t) + \Delta\theta_{i2}(t+1)), \ldots, \sin(\theta_{iN}(t) + \Delta\theta_{iN}(t+1))) \tag{47.13}$$

Then, we can see that a particle corresponds with two solutions in the solution space. We can change the two positions in the solution space by moving the particle with quantum rotation gate. The coding of quantum can enlarge the searching space and we can improve the algorithm precision.

47.3.2 QPSO for Solving Train Stops Setting Problem

The QPSO algorithm for train stops setting is as follows.

Step 1 To initiate all the variables of the model;
Step 2 To calculate the fitness value of the model according to Eq. (47.1);

Step 3 To calculate the change of argument of the particles according to Eq. (47.9);
Step 4 To calculate the positions of the particles according to Eqs. (47.12) and (47.13);
Step 5 To calculate the fitness value of the model according to Eq. (47.1);
Step 6 To select the optimal local and global particle from the two swarms;
Step 7 To check if the computational accuracy satisfies the requirements. If yes, stop; otherwise, go to step 3.

47.4 Computing Case and Results Analysis

47.4.1 Basic Data of the Computing Case

The data is from Shanghai–Hangzhou high-speed railway. The passenger OD flow on the high-speed railway is shown in Table 47.1. Figure 47.1 shows the stop plan of all the trains.

In Table 47.1, SHHQ is abbreviation for Shanghaihongqiao, SJN for Songjiangnan, JSB for Jinshanbei, JSN for Jiashannan, JXN for Jiaxingnan, TX for Tongxiang, HNX for Hainingxi, YH for Yuhang, HZD for Hangzhoudong. There are 49 trains running on Shanghai–Hangzhou high-speed railway at the down-going direction.

47.4.2 Computing Results and Its Analysis

t_{travel} is 68 min. t_{dwell} is set to be 2 min. Δ_{start} is set to be 1 min and Δ_{stop} is set to be 1 min. According to the approach given in Sect. 47.3, we optimize the train stop

Table 47.1 O-D flows of passenger between stations

	SHHQ	SJN	JSB	JSN	JXN	TX	HNX	YH	HZD
SHHQ	–	709	664	1805	2524	1460	1351	1213	3620
SJN		–	154	312	510	250	214	185	631
JSB			–	368	543	320	201	192	752
JSN				–	245	210	650	123	1020
JXN					–	856	245	102	2034
TX						–	214	185	423
HNX							–	587	1201
YH								–	1254
HZD									–

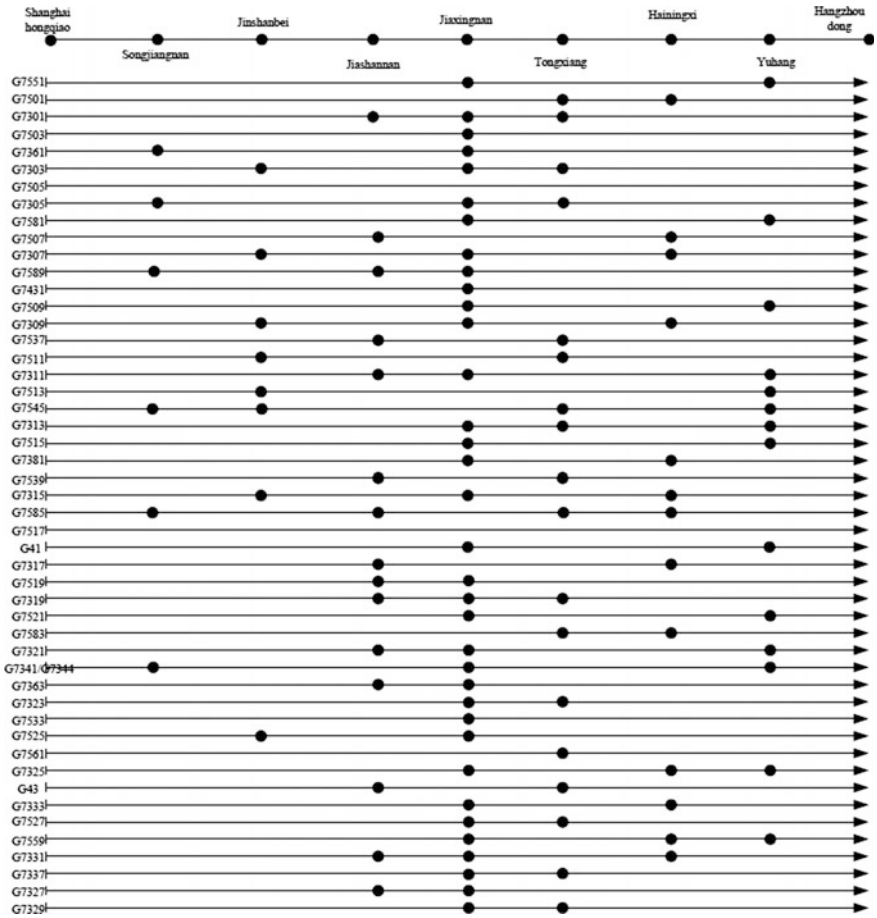


Fig. 47.1 Stop plan of the down-going trains on Shanghai–Hanzhou high-speed railway

plan of the trains. After 123 computing iterations, the optimal solution is attained. The fitness value is reduced to $\frac{27327}{68\times49+104\times(2+1+1)} = 7.291$. We can see that an additional stop is added at Songjiangnan station for G7525. G7525 adds a stop at Songjiangnan, G7545 adds a stop at Jinshanbei, G7315 and G7589 add a stop at Jiashannan, respectively. G7589 and G7315 cancel the stop at Jiaxingnan. G7319 cancels the stop at Tongxiang, G7313 cancels the stop at Yuhang. Hainingxi and Jiashannan do not change the stop plan. The optimized stop plan is shown in Fig. 47.2.

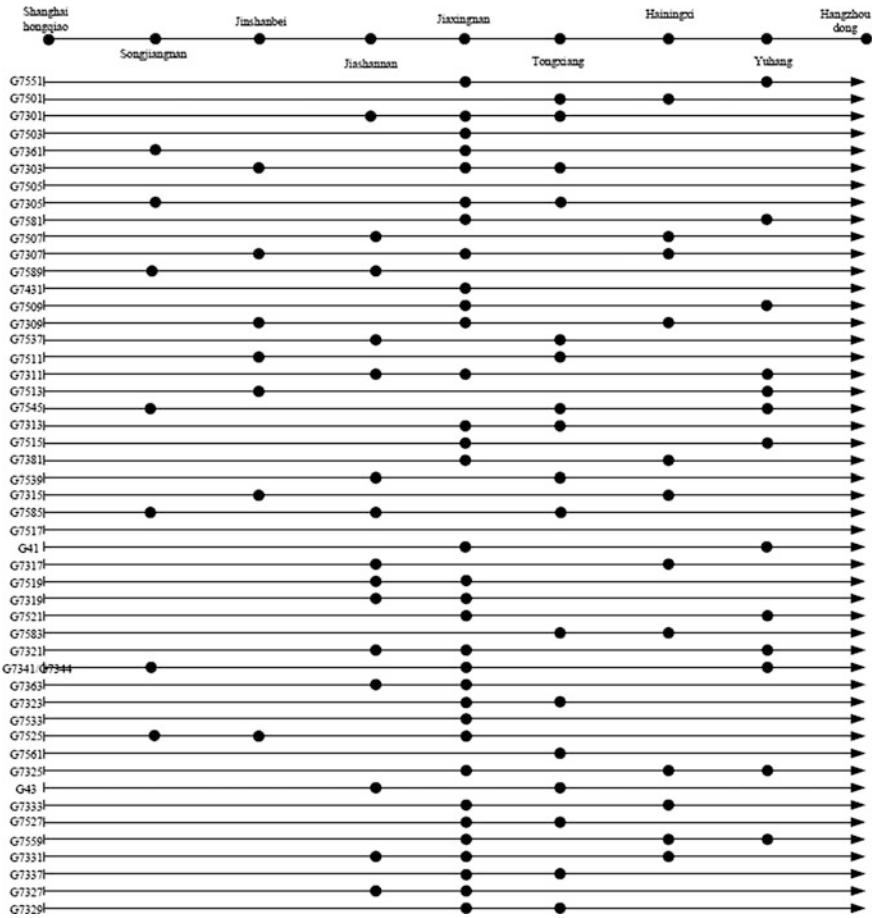


Fig. 47.2 Optimized stop plan of the down-going trains on Shanghai–Hanzhou high-speed railway

47.5 Conclusion

We take one of the line planning for railway passenger transportation-stops setting as the study focus. Stops setting is affected by the passenger OD flow and the train number. This paper defines the passenger transportation efficiency, considering the passenger OD, traveling time, dwelling time and the additional time for starting and stopping. It is a different viewpoint to optimize the stops of the trains. The QPSO has distinguished performance in this kind of optimization work. Better stop plan with higher passenger transport efficiency can be generated with the method in this paper. It is valuable for the managers OD the railway bureaus when they design the passenger train service plan.

Acknowledgment This work is supported by the National Natural Science Foundation of China (Grant 61263027), the State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2015K004), Beijing Jiaotong University, Natural Science Foundation of Gansu Province (Grant 1310RJZA068).

References

1. Chang Y-H, Yeh C-H, Shen C-C (2000) A multiobjective model for passenger train services planning: application to Taiwan's high-speed rail line. Transp Res Part B 34:91–106
2. Barrenaa E, Cancac D, Coelhoa LC (2014) Single-line rail rapid transit timetabling under dynamic passenger demand. Transp Res Part B 70:134–150
3. Lin D-Y, Ku Y-H (2014) An implicit enumeration algorithm for the passenger service planning problem: application to the Taiwan railways administration line. Eur J Oper Res 338(3): 863–875
4. Finger M, Haller A, Martins SS, Trinkner U (2014) Integrated timetables for railway passenger transport services. Competition Regul Netw Ind 15(1):78–107
5. Meng X, Cui B, Jia L (2014) Line planning in emergencies for railway networks. Kybernetes 43(1):40–52
6. Park BH, Seo YI, Hong SP, Rho HL (2013) Column generation approach to line planning with various halting-application to the Korean high-speed railway. Asia Pac J Oper Res 30(4):1350006
7. Goossensa J-W, van Hoesela S, Kroonb L (2006) On solving multi-type railway line planning problems. Eur J Oper Res 168(2):403–424
8. Mandl CE (1979) Evaluation and optimization of urban public transportation networks. Eur J Oper Res 5:396–404

Chapter 48
Optimizing Power for Train Operation Based on ACO

Zhuoyue Li, Xiukun Wei, Hui Wang and Limin Jia

Abstract The energy consumption of freight train is very huge. Among all kinds of the energy consumptions, the energy consumption of traction is the largest, which leads to a great cost of freight trains. The optimization of the speed curve plays an important role in energy-saving operation. In this paper, a speed curve optimization method based on Ant Colony Optimization (ACO) is established. A simulation model combined with the characteristics of train operation is built. The related data of energy consumption and the DYNAMIS are applied to analyze the train operation. The effectiveness of the optimization algorithm is demonstrated by the simulation results.

Keywords Energy consumption · Speed curve optimization · ACO

48.1 Introduction

Railway locomotive is the source of the traction of a train, its energy consumption accounts for 60–70 % [1] of railway transportation energy consumption. Therefore, reducing the locomotive traction energy consumption is significant. Among the various energy-saving methods, the speed curve optimization is a very effective one.

In recent years, many domestic and foreign scholars have conducted the research on the speed curve optimization on energy saving. In [2, 3], the author proved the existence of optimal operation method of a single train. In [4], the author proposed an improved GA to improve the arriving and departing efficiency of metro trains. In [5], the author divided the problem into a two-stage optimization, and the control variables could be either continuous or discrete. In [6], the Max–Min ant colony algorithm was used to study the train energy-saving optimization curve under the condition of fixed block. In [7], the author proposed a two-stage optimization, using

Z. Li (✉) · X. Wei · H. Wang · L. Jia
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing 100044, China
e-mail: xkwei@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_48

the Max–Min Ant System (MMAS) as the core algorithm. In [8], the author proposed a distance-based train trajectory searching model where some representative optimization algorithms are applied.

In the study of train energy-saving optimization, DP and GA have been widely used, but the use of ant colony algorithm (ACO) in the railway freight traffic is not that common. In this paper, the discrete nodes of the line and speed are divided according to the characters of freight trains. Determining the velocity correlation matrix of adjacent discrete points and generating the train speed curve. ACO is applied to study the train energy-saving optimization problems.

48.2 Model of Train

In the study of train energy-saving optimization, the forces are simplified as three broad categories, which are the train traction force (T), braking force (B), and resistance (R). After a detailed analysis of the forces, there are four modes which are traction, cruise, coasting, and braking.

In order to store data more easily and to search optimization path more quickly in the simulation, this model is built based on unit distance. Within a distance interval, the acceleration is regarded as a constant, the process of obtaining state $i + 1$ from state i can be illustrated by Fig. 48.1 (Δs is a known constant)

In Fig. 48.1, the initial state of the train operation is known. According to Newton's second law, acceleration within each distance interval and each state of the train can be derived.

Energy consumption of each distance interval was showed as follows (M is the total quality; Δh is the altitude difference between the current distance and the next distance; g is the acceleration of gravity):

$$e_i = \left(M \cdot \left(v_{i+1}^2 - v_i^2 \right)/2 + \int_{s_i}^{s_{i+1}} \mathrm{Rds} - \Delta h \cdot M \cdot g \right) \tag{48.2.1}$$

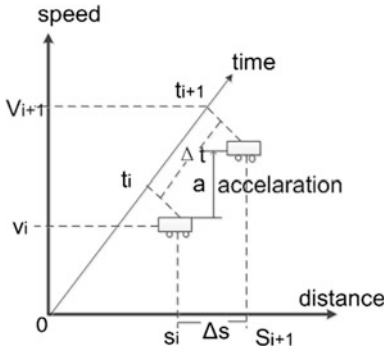


Fig. 48.1 State switch between two states

Total energy consumption:

$$E = \sum e_i \tag{48.2.2}$$

48.3 Basic Principle of Energy Optimization

In order to incorporate the algorithm to the train energy-saving optimization problem, the model which discretizes the train energy-saving optimization problem will be built first [7].

Step 1: Discretize the line into different points:

$$S \in (S_1, S_2, \ldots, S_i, \ldots S_m)$$

Near to the speed turning points, the discrete intervals are distributed more densely, guaranteeing the running speed changes smoothly at the speed turning points.

Step 2: Discretize the target speed into different points at each line discrete point:

$$v(S_i) \in \left(v_{(s_i,1)}, v_{(s_i,2)}, \ldots, v_{(s_i,j)}, \ldots v_{\left(s_i,n_{s_i}\right)}\right)$$

The maximum velocity at a discrete point is determined by (1) the speed upper limit; (2) the maximum traction of the train; (3) maximum velocity at the previous discrete point; (4) speed limit which needs to satisfy the condition of taking emergency brake at the discrete points. The train speed can be reduced to zero before the finish line. The minimum velocity at a discrete point is determined by (1) the speed lower limit; (2) the brake force of the train; (3) minimum velocity at the previous discrete point.

Step 3: Speed selection procedure based on the distance trajectory searching:

The speed discrete points at a distance discrete point are calculated and stored in advance according to the speed discrete points at the previous distance discrete point, which corresponds to different strategies like traction, cruise, coasting or braking. In order to calculate accurately, traction mode is set to 30, 60, and 100 %.

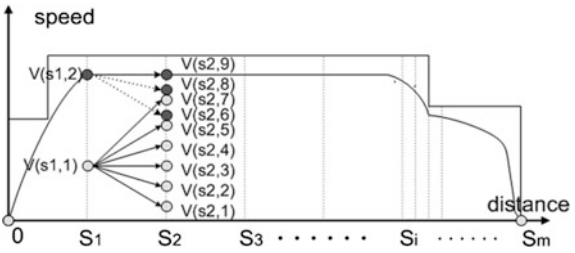


Fig. 48.2 Speed selection procedure [8]

In Fig. 48.2, assuming that there are two speed choices ($v(S_1, 1)$ and $v(S_1, 2)$) at the distance discrete point S_1. The speed selection at S_2 is from $v(S_2, 1)$ to $v(S_2, 9)$. If the train whose previous speed is $v(S_1, 1)$ takes braking, coasting or cruise, its speed will reduce to $v(S_2, 1)$, $v(S_2, 2)$, $v(S_2, 3)$, respectively; if it takes 30, 60 or 100 % traction, its speed will increase to $v(S_2, 4)$, $v(S_2, 5)$, $v(S_2, 7)$, respectively. The same rule can be applied to $v(S_1, 2)$. Because of the speed limit, the train cannot take traction.

Step 4: The establishment of correlation matrix:

Velocity correlation matrix is established as follows.

Constructing the query table of speed selection, the space size of the query table is $S_m \times n_{s_{\max}} \times n_{s_{\max}}$ ($n_{s_{\max}}$ is the maximum of $n_{s_i} (1 \leq i \leq m)$) [7].

The optional speed range is set in advance; the value of the connected velocity correlation is set to 0.01. Otherwise, the value is set to 0.

In Fig. 48.3, assuming that $n_{s_{\max}} = 6$, $1 < j < 6$ at S_i. If the starting speed is $v(S_i, j)$, it will arrive to the speed of $v(S_{i+1}, 1) \sim v(S_{i+1}, 6)$, in which $v(S_{i+1}, 1)$ and $v(S_{i+1}, 6)$ have no connection with $v(S_i, j)$, so the value of $(i, j, 1)$, $(i, j, 6)$ are set to 0, and others are set to 0.01.

Based on the above speed link in three-dimensional space, the energy consumption query table, time consumption query table and pheromone query table can be calculated. The corresponding value of $e(i, j, 1)$, $e(i, j, 6)$ are set to infinity, others are set to calculation values. The time table can be obtained in the same way. For the query table of pheromone, each connected velocity between two adjacent distance discrete points can be selected due to the pheromone information originally

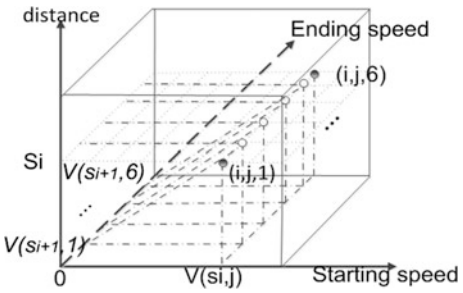


Fig. 48.3 Speed link in three dimension space

imposed. The stronger the linkage between two nodes is, the higher value is given to the pheromone correlation matrix corresponding points.

According to the Pontryagin minimum principle, there is a continuous speed curve, making energy consumption lowest under the constraint condition. The main characteristics of ACO are (1) use the distributed control; (2) ants can change the environment; (3) self-organization; (4) probabilistic global search. Thus, such a non-deterministic algorithm can have more opportunity to get the global optimal solution.

At each discrete point, ant *"k"* chooses the speed at the next predetermined point using random proportional theory. Refer to Fig. 48.3, assuming that the ant *"k"* is currently at the speed index j and the possibility of speed index Z (Z = 1, 2, 3, 4, 5, 6) to be selected is defined as follows [8]:

$$p_{(i,j,z)}^{k} = \begin{cases} \dfrac{[\tau(i,j,k)]^{\alpha}\cdot[\eta(i,j,k)]^{\beta}\cdot[\psi(i,j,k)]^{\gamma}}{\sum_{z\subset \text{allowed}_k}[\tau(i,j,k)]^{\alpha}\cdot[\eta(i,j,k)]^{\beta}\cdot[\psi(i,j,k)]^{\gamma}}, z \in \text{allowed}_k \\ 0, \text{others} \end{cases} \tag{48.3.1}$$

$allowed_k$ is the next candidate speed to be selected; α is information heuristic factor; β, γ are expect heuristic factors; $\eta(s_i,j,z)$ is heuristic function, $\psi(s_i,j,z)$ is heuristic function, $\tau(i,j,k)$ is pheromone value.

$$\eta(s_i,j,z) = 1/e(s_i,j,z) \tag{48.3.2}$$

$$\psi(s_i,j,z) = 1/t(s_i,j,z) \tag{48.3.3}$$

$$\tau(i+1,j,z) = (1-\rho)\cdot\tau(i,j,z) + \Delta\tau(i,j,z) \tag{48.3.4}$$

$$\Delta\tau(i,j,z) = \sum_{k=1}^{K}\Delta\tau(i,j,z)^{k} \tag{48.3.5}$$

K is the total number of ants, ρ is pheromone volatilization coefficient, $1-\rho$ is pheromone residual factor, $\Delta\tau(i,j,z)$ is the pheromone increment, $\Delta\tau(i,j,z)^{k}$ is the information left on the speed link of (i,j,z) by ant *"k"* in this circulation.

The train's actual running time needs within the stipulated time, the following is constraint.

Assumes that the schedule time is T_s, the total journey length is S. The average speed v_a equals to S/T_s [8]. For each distance discrete point, the elapsed time corresponding to the time schedule can be defined as follows:

$$T_{s_i} = S_i/v_a \tag{48.3.6}$$

The actual time for each distance discrete point s_i which represented as $T(i,j,z)$ can be calculated from the query table of time consumption.

The time margin ΔT_i for each distance discrete point s_i can be defined as follows:

$$\Delta T_i = T_{S_i} - T(i,j,z) \tag{48.3.7}$$

$\Delta T_i \geq 0$ means that the train runs fast enough, the parameter γ should be controlled in the range of a low value, especially when reached a larger value. If the train runs fast enough, the parameter γ can be set as 0. $\Delta T_i \leq 0$ means that the train is not fast enough and the time becomes important for the train. Thus, the parameter γ should be controlled in the range of a higher value.

48.4 Simulation Results

The simulation data comes from Shenshuo railway line. Track profile is shown in Fig. 48.4.

Based on MATLAB, the simulation undertakes 100 iterations. The results are shown as Fig. 48.5, the train speeds up to a certain speed with a larger acceleration, and then maintains constant speed. When moving uphill from 2100 to 2600 m, for example, the train takes traction mode, but the train slows down because the gradient is large. When moving downhill from 5000 m, for example, the train takes coasting mode, as the gradient is steep, the train speeds up. From 6000 to 7000 m,

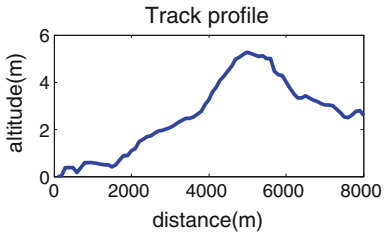


Fig. 48.4 Track profile

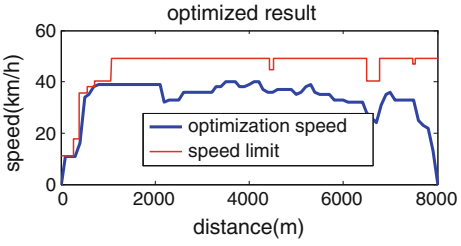


Fig. 48.5 Simulated optimization of energy-saving curve

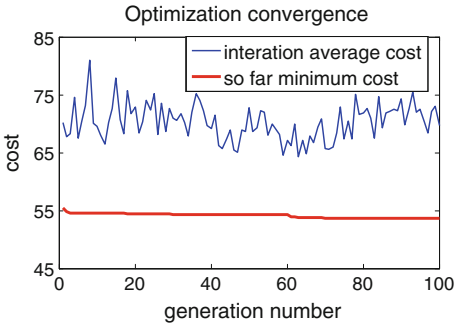


Fig. 48.6 Optimization convergence (a)

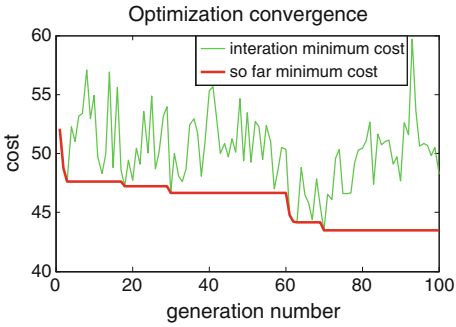


Fig. 48.7 Optimization convergence (b)

because of the influence of the speed limit and gradient, the speed decreases. Before the end stage, the train decelerates, and the deceleration became larger. Throughout the whole trip, the train accelerates with great acceleration in the start-up phase; decelerates with great deceleration before the finish line. When reaching a certain speed, the train takes cruise mode. In the whole operation process, it conforms to the drivers' driving habits.

Based on Figs. 48.6 and 48.7, Table 48.1 can be obtained.

Final convergence of minimum energy consumption saves 88 %, 9 %, 9 %, respectively, compared to the first iteration average cost, minimum cost and the so far minimum cost. When it comes to the seventeenth iteration, values are 78 %, 5.5 % and 4.5 %, respectively.

Table 48.1 Energy consumption analysis table

Iterations	Iteration average cost	Iteration minimum cost	So far minimum cost
1	82.4	47.7	47.7
17	77.9	46.1	45.6
70	71.6	43.7	43.7

48.5 Simulation Results with DYNAMIS

DYNAMIS has been developed in the Institute for Transport, Railway Construction and Operation of the University of Hanover since the mid-1980s to be used in calculation of the train. Strict physical algorithms are used to determine the precise dynamics model of a running train. It enables the user to answer questions regarding energy consumption, braking distance or dimensioning of power supplies. Importing the raw data and the data optimized by ACO into DYNAMIS.

The simulation results with raw data of the line are shown in Figs. 48.8 and 48.9; Figs. 48.10 and 48.11 show the simulation with data optimized by ACO; in order to increase the comparative, the simulation with data in the process of ACO optimizing are shown in Figs. 48.12 and 48.13.

Energy consumption analysis: based on the same parameters set by DYNAMIS, using the raw data as a guide speed curve for train, energy consumption is 33.8 (the default energy consumption unit of DYNAMIS); with the data in the process of

Fig. 48.8 Simulated operation curve with raw data

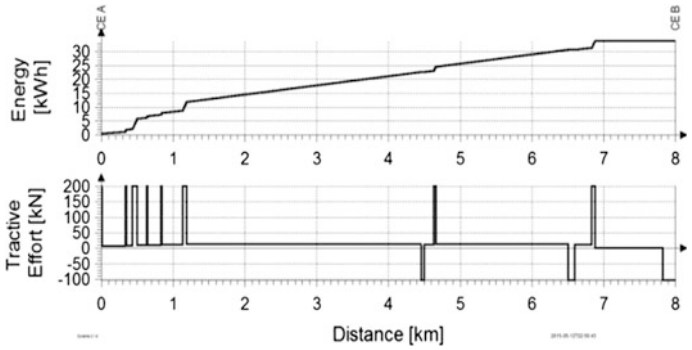


Fig. 48.9 Traction effort and energy consumption with raw data

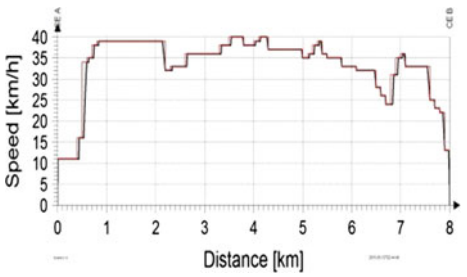


Fig. 48.10 Simulated operation curve with optimized data

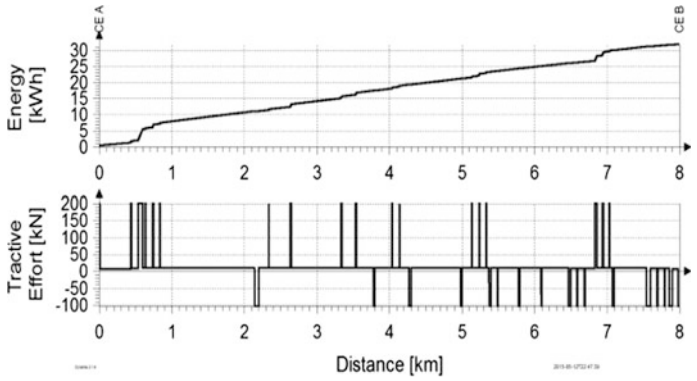
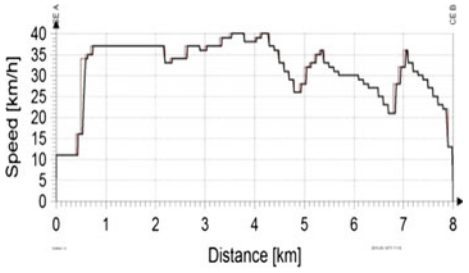


Fig. 48.11 Traction effort and energy consumption with optimized data

Fig. 48.12 Simulated operation curve with data during optimization process



ACO optimizing as a guide speed curve, energy consumption is 32.3; energy consumption is 31.8 when using the data optimized by ACO. Compared with the former two date, it saves 6.3 and 1.6 % energy, respectively.

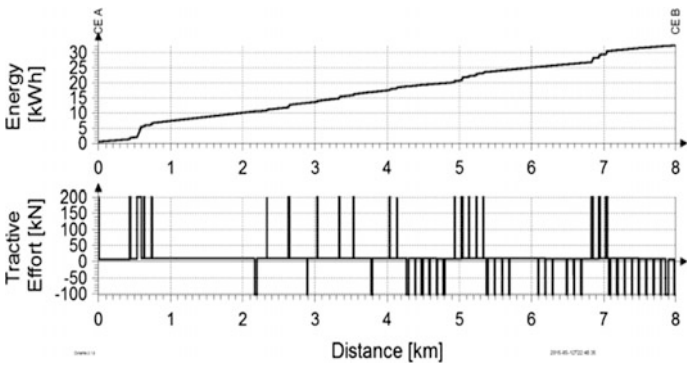


Fig. 48.13 Traction effort and energy consumption with data during optimization process

48.6 Conclusion

In this paper, the speed curve optimization problem of freight train based on ACO is investigated. The simulation results show that the train operation based on the ACO curve optimization has a better energy-saving effect. The future work will focus on three aspects. Firstly, improving ACO computational speed will be done. Secondly, ant colony algorithm for real-time online dynamic optimization will be studied further. Finally, more reality situation of the railway line will be considered to add to the simulation model.

Acknowledgments This work is supported by Chinese National Key Technologies R&D program (Contract No. 2013BAG24B03).

References

1. Wang Q (2014) Research on freight train energy-saving operation optimization. Southwest Jiaotong University (in Chinese)
2. Howlett PG (2000) The optimal control of a train. Ann Oper Res 98:65–87
3. Howleet PG (1990) An optimal strategy for the control of a train. J Austr Math Soc Ser B 3l:454–471
4. Hei X, Li Y, Wang L, Ma Q (2013) A search process for appropriate running of adjacent metro trains within stations based on genetic algorithm. Trans Electr Electron Eng 8(2):173–181 (in Chinese)
5. Xu X, Antsaklis PJ (2004) Optimal control of switched systems: new results and open problems. IEEE Trans Autom Control 4:2683–2687
6. Ke BR, Chen MC, Lin CL (2009) Block-Layout design using max-min ant system for savingenergy on mass rapid transit systems. IEEE Trans Intell Transp Syst 10(2):226–235
7. Yu X (2012) Energy-saving optimization and energy evaluation of train rail transit. Beijing Jiaotong University (in Chinese)
8. Lu S, Hillmansen S, Ho TK, Roberts C (2013) Single-train trajectory optimization. IEEE Trans Intell Transport 14(2):743–750

Chapter 49
Passenger Flow Forecast of Metro Station Based on the ARIMA Model

Shuai Feng and Guoqiang Cai

Abstract By the end of 2014, 83 metro lines with a length of over 2500 km in total had been constructed in 22 metropolitan cities in mainland China. A series of worth exploring and pondering problem arises in the construction process, and the passenger flow prediction analysis of metro station is one of them. This paper built an ARIMA model which is a kind of short-time traffic forecasting model with high precision. The detailed data of historical passenger flow in section in a typical station are fitted in this paper. On the basis of this, the passenger flow in the next day is forecasted and analyzed. The fitting is with the help of statistical software called SPSS. Finally, the model of ARIMA (3, 0, 2) is built up. The results showed that the ARIMA model prediction has certain accuracy. It can solve the problem of modeling about non-stationary time series prediction.

Keywords ARIMA model · Time series · Metro station · Passenger flow prediction

49.1 Introduction

The passenger flow prediction of metro station is an important part of intelligent transportation system. In present operational management process, it is lacking in quantitative prediction methods. Most of the station staff rely on intuition and experience to recognize the change of passenger flow. This will lead to a great uncertainty of the prediction. Sometimes, it is far different from the real results.

With the development of the random theory, time series analysis has become more and more important in economic statistics and prediction technology. Time series modeling methods usually contain the Box–Jenkins modeling method and the Pandit–Wu modeling method [1, 2]. Box–Jenkins modeling method is based on

S. Feng · G. Cai (✉)
State Key Laboratory of Rail Traffic Control and Safety,
Jiaotong University, Beijing, Haidian, China
e-mail: guoqiangcai@163.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_49

time series of autocorrelation function and partial autocorrelation function's statistical properties, finding out the suitable model, and then estimating the model. Usually the models we can think of are the auto-regressive moving average model (ARMA), auto-regressive-integrated moving average model (ARIMA), and multiplicative seasonal model. Pandit–Wu demonstrates the stationary time series of uniform interval sample and puts forward the idea of a dynamic system modeling, by gradually increasing order of the model until the increasing order cannot significantly reduce residual sum of squares. In the actual cases, the random data sequence is often not smooth. The Box–Jenkins model is very fit for the actual prediction that is not easy to build accurate mathematical model and system modeling with uncertainty, through the study of the difference of random data series, logarithmic transformation operations, using ARIMA modeling and forecasting analysis model.

49.2 The Principle of ARIMA Model

The traditional time series analysis methods, called ARIMA, that is auto-regressive integrated moving average model which, contain three parameters [3]. The parameters are p, d, q. It puts forward by Box–Jenkins in 1970. AR means auto-regressive model, and p is the auto-regressive coefficient. The MA represents moving average model, q is the moving average number, and d is the number of differences required to give a stationary series. The ARIMA model is defined by the following restriction, converting non-stationary time series into a stationary time series, fitting to time series data either to better understand the data or to predict future points in the series prediction. The step of prediction is like this. First, a trial model was given based on the time series. Secondly, a lot of experiments will be tried, identified, and estimated until a suitable model is built up. This is the process from identifying and modeling to try again.

The ARIMA (p, d, q) series can be defined by the following equation, assuming that d root lie in the unit circle [4, 5]:

$$A(B)\nabla^d y_t = C(B)_{\in t} \tag{49.1}$$

$$A(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p \tag{49.2}$$

$$C(B) = 1 - \beta_1 B - \beta_2 B^2 - \cdots - \beta_q B^{\mathrm{q}} \tag{49.3}$$

where $A(B)$ and $C(B)$ are p-order polynomials in B, $|\alpha_1| \leq 1, \ldots |\alpha_p| \leq 1$ and $|\beta_1| \leq 1, \ldots |\beta_{\mathrm{q}}| \leq 1$, and where B is the lag operator. $\in t$ is a white noise process for stationarity, the roots of $A(B)$ should lie outside the unit circle and for invertibility, the roots of $C(B)$ should lie outside the unit circle. This requirement that the

roots of both $\boldsymbol{A}(B)$ and $C(B)$ lie outside the unit circle $\nabla^d = (1 - B)^d$ is the dth power of the differencing operator. The differenced series $w_t = \nabla^d y_t$ is a stationary ARIMA (p, d, q) series.

49.3 The Forecasting Steps About ARIMA Model

The time series data must be stationary before we use ARIMA model to predict. That is to say, influential factors of the research must be the same within a specified time. If the given data series is not stationary or smooth, then data series should be processed in advance, thus making time series smooth. After those steps, the ARIMA model will be built up. The specific process is shown in Fig. 49.1 [6].

1. Time series processed in advance: It is also called stationary time series. By means of differential and logarithmic transformation, the original dataset is transformed into a stationary time series.
2. Testing the time series: We calculate the autocorrelation coefficient and partial autocorrelation coefficient of the time series and then draw autocorrelation plots and partial autocorrelation plots of the time series.
3. Selecting the suitable ARIMA model in fitting, estimating the model parameters, where p is the auto-regressive coefficient; d is the number of differences needed to give a stationary series; q is the moving average number of items.

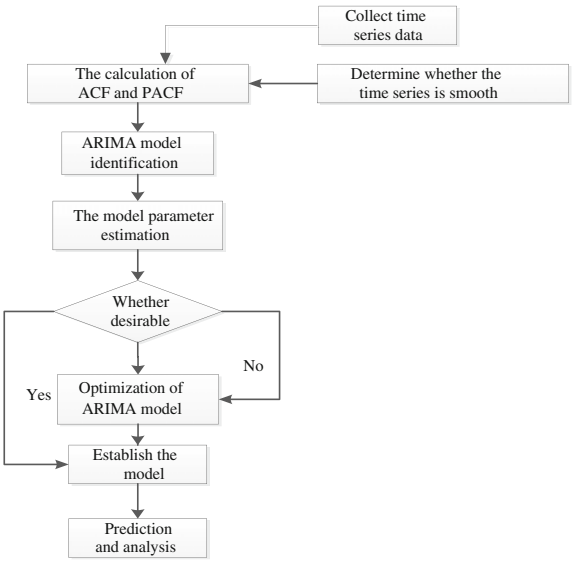


Fig. 49.1 The flowchart of ARIMA model prediction

4. Testing model validity: If the model cannot be tested, we still go back to the previous step. Model identification will be carried out until the parameters p, d, q are suitable.
5. Predicting the future trend of the time series based on the fitting model.

49.4 Modeling and Forecasting

49.4.1 The Initial Data

In this paper, taking the passenger flow from the detailed data of historical passenger flow in section in a typical station as the original data, that is June 1–18 from Pingguoyuan Station to Guchenglu Station of Beijing Subway Line 1. The observation period is 18 days. Data collected are at 30 min intervals with a total of 594 numbers. The normal mode of the passenger flow behavior is shown in Fig. 49.2. The passenger flow data come from Beijing Traffic Commanding Center (TCC).

Considering the factors of weather, passenger's mentality, and others, the behavior of the passengers is indeterminate. But if we are not in the case of special events, such as the concert, the sports meet, differences in variation of passenger flow weekly are very small. However, it is determined by the work schedule of the city enterprises and government institutions.

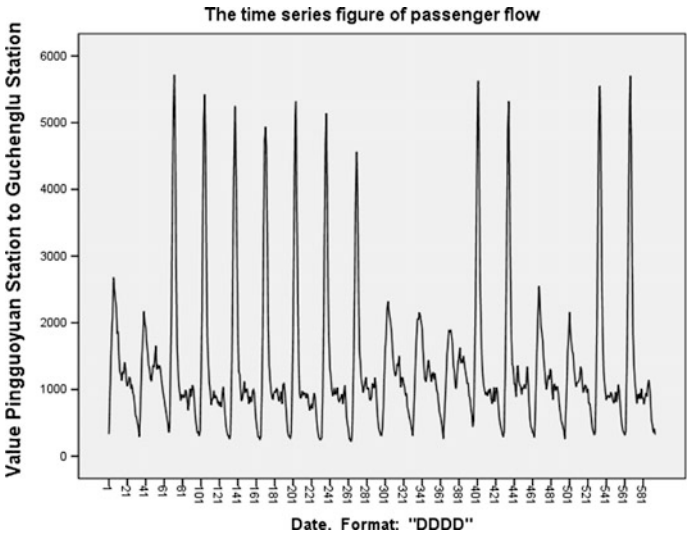


Fig. 49.2 Normal mode of the passenger flow behavior from Pingguoyuan to Guchenglu Station

49.4.2 Time Series Stationarized

Before modeling, first of all, we must ensure that the time series is smooth. In Fig. 49.2, we can see that the series are not stationary. So the time series processed in advance is needed. We use the natural logarithm to convert the data series in this paper. As shown in Fig. 49.3, after the conversion, the time series has no obvious upward or downward trend. This shows that the time series is stationary after the logarithmic transformation. We draw autocorrelation plots and partial autocorrelation plots of the time series by SPSS software. The correlation coefficient is negative exponential decay, and with the increasement of lag number, autocorrelation function (ACF) is close to zero. Also, partial autocorrelation function (PACF) shows "sharp" attenuation; PACF tends to zero by adding the lag number. The autocorrelation and partial autocorrelation plots are shown in Fig. 49.4. We can see that the time series data are stationary after conversion.

49.4.3 Model Establishment and Forecasting

According to the theory of time series, this paper chooses ARIMA model to design and to predict the next-day passenger flow from Pingguoyuan Station to Guchenglu Station.

AIC is the Akaike Information Criterion [7]. It is given by a Japanese scholar called Akaike, who puts forward a model selection of statistical model that is widely used. It is suitable for the order of the ARMIA model [8]. Using SPSS,

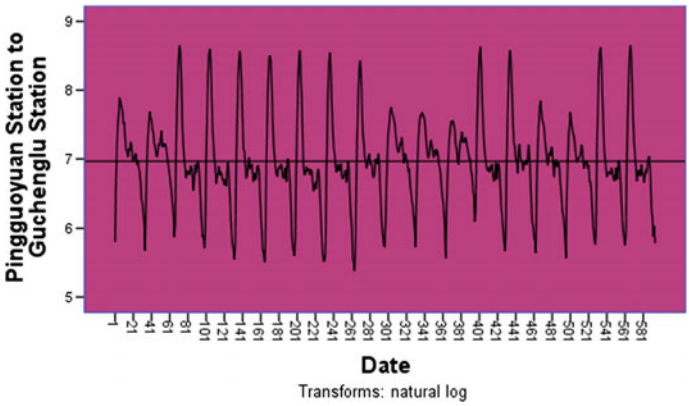


Fig. 49.3 The converted time series by natural log

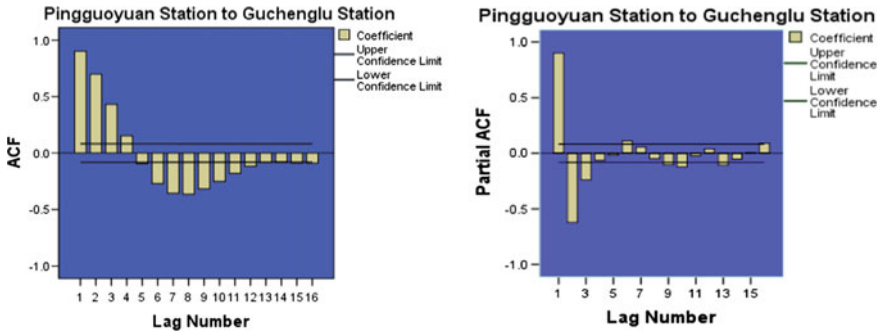


Fig. 49.4 Autocorrelation and partial autocorrelation plots

trying it step by step until three parameters of ARIMA (p, d, q) are suitable. According to the AIC, the ARIMA (3, 0, 2) is the best fitted model as the optimal fitting model of stationary time series. The model prediction and the fitting plot are shown in Fig. 49.5. The prediction and measured values of passenger flow at different time periods in one line are shown in Fig. 49.6.

From Figs. 49.5 and 49.6, we can see that most of the periods are consistency, except the period of 5:30 a.m.–9:00 a.m. and 7:00 p.m.–9:30 p.m. The prediction has certain of credibility. From the passenger flow prediction data all day, the relative error is 12.88 %. The prediction model is of certain reference value to the whole line.

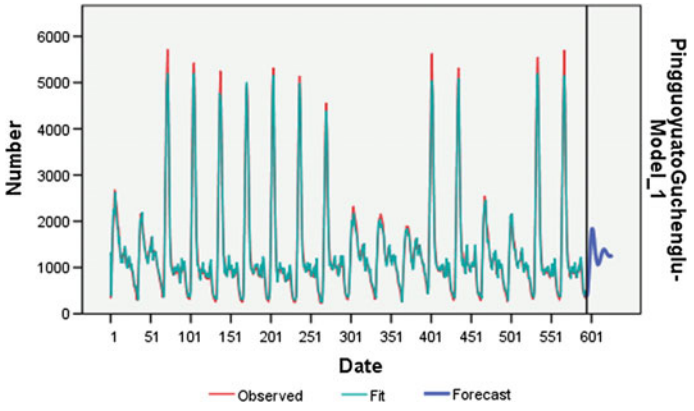


Fig. 49.5 Model prediction and fitting

Fig. 49.6 Comparisons of predictive values and real values

time bucket	predictive values	real values	elative erro
5:00–5:30	368	393	6. 36%
05:30–06:00	472	1085	56. 50%
06:00–06:30	655	2118	69. 10%
06:30–07:00	927	3673	74. 70%
07:00–07:30	1258	5037	75. 00%
07:30–08:00	1575	5485	71. 30%
08:00–08:30	1790	4559	60. 70%
08:30–09:00	1856	2805	33. 80%
09:00–09:30	1787	1823	2. 00%
09:30–10:00	1634	1302	20. 30%
10:00–10:30	1457	1087	25. 40%
10:30–11:00	1297	970	25. 20%
11:00–11:30	1175	808	31. 20%
11:30–12:00	1100	808	26. 50%
12:00–12:30	1069	899	15. 90%
12:30–13:00	1078	1019	5. 50%
13:00–13:30	1119	926	17. 20%
13:30–14:00	1181	1029	12. 90%
14:00–14:30	1251	882	29. 50%
14:30–15:00	1315	969	26. 30%
15:00–15:30	1364	828	39. 30%
15:30–16:00	1389	865	37. 70%
16:00–16:30	1391	964	30. 70%
16:30–17:00	1374	840	38. 90%
17:00–17:30	1345	1012	24. 70%
17:30–18:00	1312	1130	16. 10%
18:00–18:30	1282	992	29. 20%
18:30–19:00	1259	760	65. 60%
19:00–19:30	1245	671	85. 50%
19:30–20:00	1241	426	191. 30%
20:00–20:30	1245	394	215. 90%
20:30–21:00	1256	346	263. 00%
21:00–21:30	1270	330	284. 80%

49.5 Conclusion

By using the time series theory, the ARIMA model is established which can predict the short-term passenger flow. And it has a certain degree of accuracy. The fact shows that the modeling process is relatively simple and also has certain accuracy.

In order to forecast short-time passenger flow, there is a lot of work to be done, such as the online estimation of the model parameters, the general smoothing methods of time series. With the rapid development of urban rail transit, further research is essential to the new information technology.

Acknowledgment The author thanks the anonymous reviewers for their insightful and detailed comments. This paper was supported by the Comprehensive State Urban Rail Transfer Station Perception and Business Collaboration Topics (2012AA112403).

References

1. Hipel KW, Mcleod AI, Lennox WC (1977) Advances in Box-Jenkins modeling: 1. Model construction. Water Resour Res 13(3):567–575
2. Rui XU, Huang DF, Zhou LT et al (2009) The application of the time series analysis to GPS residuals. Sci Surveying Map 34(2):58–60 (in Chinese)
3. Gu Y, Han Y, Fang X (2011) Method of hub station passenger flow prediction based on ARMA model. Comput Commun 02:5–9 (in Chinese)
4. El Hag HMA, Sharif SM (2007) An adjusted ARIMA model for internet traffic. In: Computers/control engineering, pp 947–952
5. Chao H, Su S, Chenghong W (2004) A real-time short-term traffic flow adaptive forecasting method based on ARIMA model. J Syst Simul 07:1530–1532 (in Chinese)
6. Yajima Y (1985) Estimation of the degree of differencing of an ARIMA process. Ann Inst Stat Math 37(1):389–408
7. Webster R, Mcbratney AB (1989) On the Akaike information criterion for choosing models for variograms of soil properties. J Soil Sci 40(3):493–496
8. Wang S, Li Y (2010) Research of ARIMA model in buoy pressure error measurement. Comput Measure control 09:2054–2056 (in Chinese)

Chapter 50
Research on Method of Highlight Elimination Based on Color Space Conversion in Metal Images

Weibo Yu, Liming Zheng, Yanhui Ma and Keping Liu

Abstract In the process of three-dimensional detection with computer vision, there is a highlighted phenomenon in metal artifact surface. It leads to the interference problem in subsequent detection. To solve this question, a method of image highlight elimination is proposed which is based on color space conversion technology. Firstly, the RGB color space is converted into the Y′CbCr color space, the luminance value is adjusted, and then, the highlighted part of the image with luminance adjustment polynomial that obtained using cubic spline interpolation is disposed. Finally, the removed highlighted image in the RGB color space is obtained. Experimental results show that this method can not only effectively remove the highlighted metal in the image, but also have a better performance on preserving the texture information of the original image.

Keywords Metal images · Color space conversion · Highlight elimination · Luminance adjustment

50.1 Introduction

Computer vision technology is widely used in advanced manufacturing industry. It is one of the important aspects to obtain the three-dimensional images' information from two-dimensional images of artifact. The images obtained using camera are affected by many factors including camera's performance, lens distortion, light conditions, and objects themselves. The surfaces of most objects are diffused surfaces. But, the specular reflection happens on the metal objects' surface occurs in the light, which leads to highlighted phenomenon. Highlighted phenomenon makes images with high brightness. It will affect the detection system to obtain the

W. Yu (✉) · L. Zheng · Y. Ma · K. Liu
College of Electrical and Electronic Engineering, Changchun University of Technology, No. 2055 Yanan Street, Changchun, Jilin, China
e-mail: yu_weibo@126.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_50

matching points, which will result in a large error when detecting artifact size. Therefore, it is very important to eliminate highlighted part in metal object images.

The highlight removal methods of object surface are mainly as follows: the backlight lighting system, image inpainting method, dichromatic reflection model method, and light constraint complementary color method. The backlight lighting system [1] needs to be adjusted according to the actual manufacture environment. Compared with the light source designed in laboratory, it will produce a large deviation. Image inpainting method [2] requires manual intervention. It will slowly process the data because of large amount of information. Direct inpainting leads to lose of original image features in highlighted area. Dichromatic reflection model method [3] is a method that the highlighted area and the diffuse reflection area are, respectively, processed by using principal component analysis based on dichromatic reflection template [4]. This method can quickly remove the highlighted area. But the highlighted pixel clusters distort due to geometric shape and roughness of objects surface. It is not precise enough to estimate the color of the light source. Those make it not easy to be applicable. Light constraint complementary color method [5] overcomes the shortcomings that the general complementary method cannot maintain subtle changes on objects surface. But this method is not suitable for the real-time image processing due to manual intervention, complicated algorithm, and time-consuming. Therefore, it is necessary to find a simple and effective method that is applicable to eliminate the highlight of metal artifact surface image.

50.2 The Basic Principle of Color Space Conversion

50.2.1 Profile of Color Space

Color space [6] is a method used to express color by mathematics. It is used to specify and generate colors. From technical point, the color space can be divided into the following three categories. (1) Computer graphics color space (RGB-type color space) is a color space based on the principle of three primary colors composed of red, green, and blue, to express color image [7]. The color and luminance information contains in the three primary colors. They are correlative with each other, which makes it difficult to remove hightlight. (2) CIE color space [8] (XYZ-type color space) is the international standard of color space, generally served as the basic measurements of color. It can be used as a transitional color space, if it is unavailable to convert one color space to another color space directly. (3) Television system color space (YUV-type color space) is a television transmission color space in which luminance and chrominance are separated. In YUV color space, Y is luminance, and U and V are different components. Y′CbCr color space is derived from YUV color space, in which Y′ is the color luma. Cb and Cr are blue and red concentration offset components [9].

50.2.2 Color Space Conversion

In order to meet different application requirements, it is required to convert between different color spaces. When we detect the size of metal artifact, there is the highlighted phenomenon on its surface. Highlight elimination is to remove the highlighted area in images. Because the images are in RGB mode in the system of metal artifact size detection, here, we adopt computer graphics color space and television system color space.

In different standards, the relationships of different color spaces are different. International Telecommunication Union-Radiocommunication Sector (short for ITU-R) gives us a color space conversion in ITU-R BT. 601 [10]. In television system color space, R′G′B′ images come from tristimulus RGB images after gamma correction. Then those images are processed by color space conversion. Images are in RGB model in the system of metal artifact size detection. Luma value of three primary colors components is between 0 and 255 in those RGB images.

In the ITU-R BT. 601 standard, the Eq. (50.1) represents the relationship between RGB color space and Y′CbCr color space [11].

$$\begin{bmatrix} Y' \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R'_{255} \\ G'_{255} \\ B'_{255} \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{50.1}$$

where the value of Y′ is from 16 to 235, and the value of Cb and that of Cr are both from 16 to 240.

The relationship from Y′CbCr color space to RGB color space is represented as the Eq. (50.2) which comes from the Eq. (50.1).

$$\begin{bmatrix} R'_{255} \\ G'_{255} \\ B'_{255} \end{bmatrix} = \begin{bmatrix} 1.164 & 0 & 1.596 \\ 1.164 & -0.392 & -0.813 \\ 1.164 & 2.017 & 0 \end{bmatrix} \begin{bmatrix} Y'-16 \\ Cb-128 \\ Cr-128 \end{bmatrix} \tag{50.2}$$

50.3 Method of Eliminating Highlight Based on Color Space Conversion in Metal Image

Based on the above discussion, the range of luminance Y′ is 16–235. So, the value of every pixel's luminance is not 0. In general, the eyes feel soft if the value of the luminance is less than 170. This range is called the diffuse range. The eyes feel discomfort if the value of Y′ is between 200 and 255. At this time, the images present the highlighted phenomenon, which causes the texture feature of metal artifact weaken or even disappear. So the value of highlight needs to be converted to the value of diffuse area. The diffuse area should be reserved as much as possible at the same time.

50.3.1 The Specific Process of Highlight Elimination Algorithm

The method of highlight elimination is based on color space conversion. This process is divided into five steps as follows:

Step 1: Convert RGB color space to Y′CbCr color space with Eq. (50.1).
Step 2: Normalize the value of luminance Y′. That is to map [16, 235] to [0, 1], which can simplify the subsequent calculation and make the curve clearer.
Step 3: Histogram equalizes the artifact image. This makes the image brightness to be distributed in the whole range of gray instead of in some certain sections. It just redistributes the histogram density without destroying the original histogram.
Step 4: Determine the polynomial function to adjust the luminance with mathematical methods after considering the speed and accuracy.
Step 5: The new brightness value instead of the original brightness in 0–1 is converted to 16–235. The image is obtained after converting Y′CbCr color space to RGB color space by conversion Eq. (50.2).

50.3.2 Determine the Polynomial Function to Adjust Brightness

Determine the polynomial function of new brightness, after the original brightness normalization and equalization. The polynomial is obtained with cubic spline interpolation in this paper, because it is simple to calculate, high stability, excellent in convergence, and easy to implement on the computer.

If Y' is between 0 and 0.5 which is low luminance value, we take $Y'_1 = Y'$. If Y' ranges 0.5–1.0 which is high luminance value, we take Y'_1 from 0.5–0.65. Within this interval, the polynomial is obtained with cubic spline interpolation in the natural boundary conditions. The following piecewise polynomial is used to adjust highlight brightness.

$$Y'_1 = f(Y') = \begin{cases} Y' & (0 \le Y < 0.5) \\ -0.1778Y'^3 - 0.0444Y'^2 + 0.6778Y' + 0.1944 & (0.5 \le Y' \le 1) \end{cases} \tag{50.3}$$

The adjustment ratio (the ratio of new and original brightness value) polynomial is given as follows, that is representation of the conversion rate.

$$y = Y_1'/Y' = \begin{cases} 1 & (0 \le Y' < 0.5) \\ -0.1778Y'^2 - 0.0444Y' + 0.6778 + 0.1944/Y' & (0.5 \le Y' \le 1) \end{cases} \tag{50.4}$$

When Y' is between 0 and 0.5, y is 1, which is the representation of low conversion rate. When Y' ranges from 0.5 to 1.0, Y_1' is from 0.5 to 0.65 increasingly, which represents conversion rate is gradually increasing. This ensures not only eliminating image highlight in highlighted area, but also keeping low conversion rate in diffuse area.

50.4 Experiments and Results

Image of ring metal artifact is acquired with industrial CCD camera. It is in RGB model and its size is 720 × 480. There are some highlighted regions on metal artifact under white light source. The textural feature in highlighted regions is not obvious or even disappeared. Figure 50.1a is the original image. Figure 50.1b is the image after histogram equalization. Figure 50.1c is the histogram of original image. Figure 50.1d is the histogram of histogram-equalized image. The brightness value that is more than 170 in the histogram represents highlighted area in image. From Fig. 50.1, the luminance of the original image is relatively concentrated, while the luminance spacing of the equalization image increases obviously. The details are more clearly in equalization image than original image.

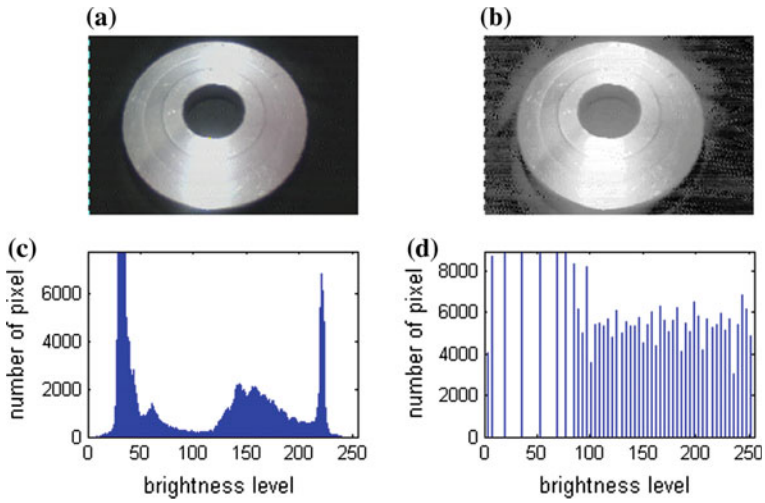


Fig. 50.1 Original image, equalization image, and their histograms: **a** original image, **b** equalization image, **c** histogram of original image, **d** histogram of equalization image

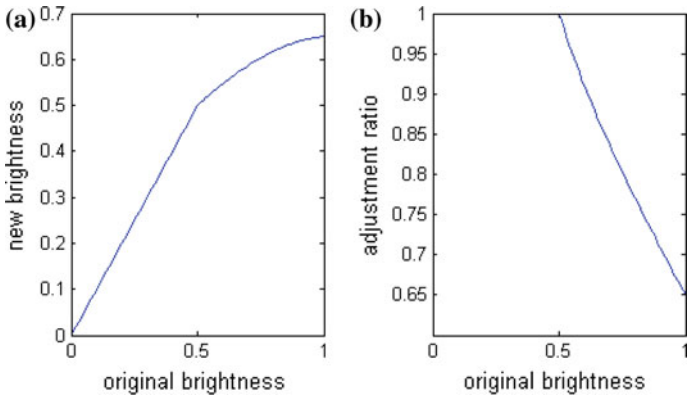


Fig. 50.2 New brightness curve and brightness adjustment ratio curve: **a** new brightness curve, **b** brightness adjustment ratio curve

Figure 50.2a is new brightness $Y'_1 = f(Y')$ curve obtained by Eq. (50.3). Figure 50.2b is the adjustment ratio curve. The adjustment ratio is large in low luminance area, which means the low conversion rate. The adjustment ratio is small in highlighted area, which means the high conversion rate.

Convert Y′CbCr color space to RGB color space with Eq. (50.2), after converting the new luminance Y′ into 16–235. So we get the metal artifact image of highlight elimination. The image is shown in Fig. 50.3a. Figure 50.3b is the histogram of highlight elimination image. It shows that the pixel in the highlighted area of original image has almost been eliminated. The pixels that the luminance value is larger than 170 have almost been eliminated. Figure 50.3a preserves the texture feature information of original image.

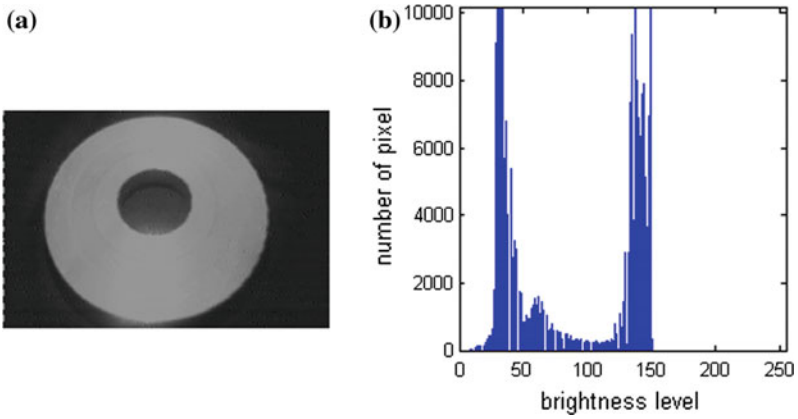


Fig. 50.3 Highlight elimination image and its histogram: **a** highlight elimination image, **b** histogram of highlight elimination image

50.5 Conclusion

We have proposed a method based on color space conversion and luminance adjustment polynomial to eliminate image highlight which is produced in the process of detecting metal artifact. Firstly, the image is converted from RGB color space to Y′CbCr color space. Then, the luminance values are normalized and equalized. New luminance value is obtained with piecewise polynomial of adjustment brightness that obtained with cubic spline interpolation. Finally, after converting Y′CbCr color space to RGB color space, we get the highlight elimination image. Experiments show that this method not only lowers the luminance value in the highlighted region, but also does not change the luminance value in diffuse region. It preserves the texture feature information of original image at the same time. Therefore, it can eliminate the highlighted area in metal images with this method, to achieve the purpose of weakening highlight.

Acknowledgments This work is supported by Jilin Provincial Science and Technology Development Project (20140204018GX) and Jilin Provincial Youth Leading Talent and Team Project of Science and Technology Innovation (20150519009JH). Weibo Yu is the corresponding author.

References

1. Yu K (2006) Research on image with highlight based on delamination and its application of assembling angle position of intravenous needles. Zhejiang University of Technology (in Chinese)
2. Marcelo B, Guillermo S, Vicent C et al (2000) Image inpainting. In: Computer graphics proceedings, pp 417–424
3. Wang C, Zhu F (2007) Removing highly illuminated region from a single image. J Shanghai Univ (Natural Science Edition) 13(2):151–154 (in Chinese)
4. Yang Q, Wang S, , Ahuja N (2010) Real-time specular highlight removal using bilateral filtering. In: European conference on computer vision, Greece, pp 87–100 (in Chinese)
5. Tan P, Yang J, Lei B et al (2014) Illumination-constrained inpainting for single image highlight removal. J Software 15(1):33–40 (in Chinese)
6. Lin F (2009) Multimedia tutorial. Tsinghua University Press, Beijing (in Chinese)
7. Wu L, Chen X, Wang F et al (2013) The application of color space transformation in video sampling system. Shanxi Electron Technol 3:51–52, 85 (in Chinese)
8. Zhang C, Zhang F, Fu S et al (2015) Study on the toning mechanism of old newsprint deinked pulps. Paper Paper Making 34(1):23–27 (in Chinese)
9. Zhang G, Yu J, Chen X (2010) Face detection based on skin color segmentation and improved gabor filter. Comput Measure Control 18(1):136–138, 141 (in Chinese)
10. ITU-R (2011) Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. International Radio Consultative Committee
11. Tan T, Ma D, Huang K et al (2015) Power-efficient image blending engine design based on self-adaptive pipeline. J Zhejiang Univ (Eng Sci) 49(1):27–35, 53 (in Chinese)

Chapter 51
The Design of a Novel Digital Asymmetric HVI Code Generator

Yuan Yang and Lei Yuan

Abstract The code generator of asymmetric high-voltage impulse (HVI) track circuit is used to generate asymmetric HVI to break down the rust pollution layer of rail surface, in order to deal with the undesirable train shunting phenomenon caused by the pollution. After pointed out the problems in existing analog asymmetric HVI track circuit code generator, we use DSP TMS320F2812 to convert analog circuit into digital one, constant current source to supply power, and two IGBTs instead of thyristor to reduce the bad impacts caused by changes of circuit state. New capacitance value calculation and monitoring methods, and new output impulse calculation and monitoring methods are presented in this paper. These new methods make the code generator digitalized and the online fault monitoring function goes well.

Keywords Asymmetric HVI · Digitalized code generator · Fault monitoring

51.1 Introduction

Track circuit is used to check whether there is a train on some track sections, whether a track section is complete, and whether there is an equipment operation. When there is a train on a track circuit section, the track relays should be shunt release; this state is called track circuit shunt state. The correct and timely shunt is an important condition to ensure train safety. Different railway lines undertake different tasks such as some lines or sections will appear rail surface rust, rail surface dust accumulation, wheels wear, and some other bad phenomena. These can cause bad contact between wheels and rail surface and also result in a train occupation that cannot be detected. The bad shunting phenomenon seriously affects transport efficiency and even safety.

Y. Yang (✉) · L. Yuan (✉)
School of Automation and Information, Xi'an University of Technology,
No. 5 South Jinhua Road, Xi'an, Shaanxi, China
e-mail: yuanleixut@126.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_51

The safe shunt voltage in old and surface severity oxidized sections recommended by the International Union of Rail (UIC) European Rail Research Institute (ERRI) is 50 V (peak value) [1].

In consideration of bad rail surface environment and safe rail shunt voltage requirements, asymmetric HVI track circuit appears for the detection of train wheel's occupation. It well solves poor shunting and fast action problems caused by track circuit shunt sensitivity and high-resistance wheels, and was widely used in DC and AC electrified stations and sections [2].

51.2 The Characteristics of Asymmetric HVI

In any part of track circuit, enough peak voltage and enough current that can reduce resistance to limit value are guarantees for reliable shunt. Also the choice of impulse repetition frequency is very important in order to maintain the relay correct attracting or falling state between the adjacent impulse, and at the same time, letting relay reaction's time as short as possible after the track circuit shunting. The impulse repetition frequency still need to be reduced to some extent to limit the sending impulse power. The impedance of track circuit is very low. In order to get lower consumption of track circuit, high-voltage and continuous large current power supply cannot be used. But high-voltage, large current are necessary for track circuits to break down oil, rust, and semiconductor insulating layer. According to the above, choose asymmetric HVI that has long time interval and short time impulse. The asymmetric HVI consists of positive and negative impulses; the positive impulse amplitude is far greater than the negative impulse amplitude. And the positive impulse amplitude width is much smaller than the negative impulse. The asymmetric HVI period is 3 or 3.5 Hz, and the ideal asymmetric HVI is shown in Fig. 51.1.

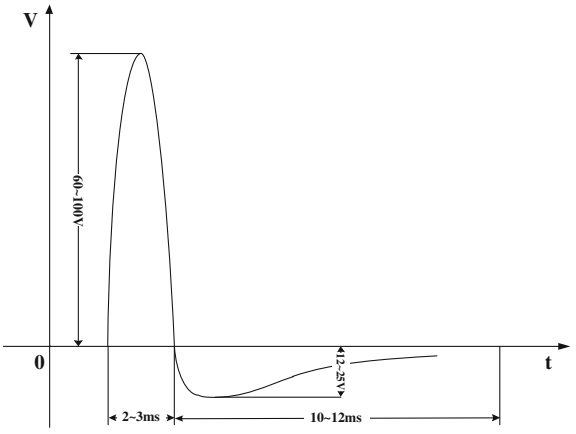


Fig. 51.1 The ideal high-voltage asymmetric pulse waveform

The asymmetrical HVI track circuit code generator uses capacitor charging and discharging technology, that is, a long time charging, instantaneous discharging to the rail. Discharging process forms a high impulse. When at the end of the capacitor discharging, due to the load choke transformer and the track circuit's inductance magnetic to electric conversion, the impulse tail forms in the load. So get the asymmetric HVI of 1 % full duty ratio. Asymmetrical HVI capacitor C = 50 μF, and charging terminal voltage is 300–500 V. Assuming that voltage is 400 V, capacitor *C*'s storage capacity is given as:

$$W = \frac{CU^2}{2} = \frac{40 \times 10^{-6} \times 400^2}{2} = 3.2 \text{ J}$$

The maximum instantaneous power of impulse can be up to 10,000 W. Asymmetric HVI has short power period and long inactivity period; each track circuit's average power consumption is less than 60 W. At the same time, asymmetric HVI track circuit's power consumption is unrelated to train occupation and changes of track circuit's load. It depends only on the code generator's capacitor energy storage.

51.3 Problems of Traditional Code Generator

Asymmetric HVI track circuit consists of code generator, receiver, and track relay. Code generator is the source generating part. To generate the correct asymmetric HVI signal (which has right polarity, appropriate amplitude, impulse width, frequency and steepness of wave edge) through the process of charging and discharging to 50 μF capacitor is crucial to the entire track circuits[3].

The traditional code generator consists of analog circuit, and the principle of the circuit is shown in Fig. 51.2. Since the drift of analog circuit is large the output impulse's amplitude stability is poor. Analog circuits also have disadvantages such as large size, expensive price, complex production process, and operation and unstable work. So digitalized code generator is an inevitable trend [4].

At present, the analog code generator use 50 Hz, 220 V AC power supply, and series voltage stabilizing transformer scheme. In the design of code generator, capacitor's charging and discharging rely on the realization of controlling code generator loaded or shorten in frequency of 3 or 3.5 Hz. Though the short circuit time is very short, the transformer's temperature can rise quickly, causing heating, insulation aging, and damage and taking great impact to power [5].

Under the current power supply mode, 50 μF capacitor will have serious heating phenomenon after code generator works for a long time. This will reduce the life of capacitor and then will make code generator unable to send correct asymmetric HVI waveform, or even cause code generator unable to work completely. These phenomena will directly affect the rail vehicle's occupation check and driving safety. At the same time, AC power through the transformer to the capacitor for charging

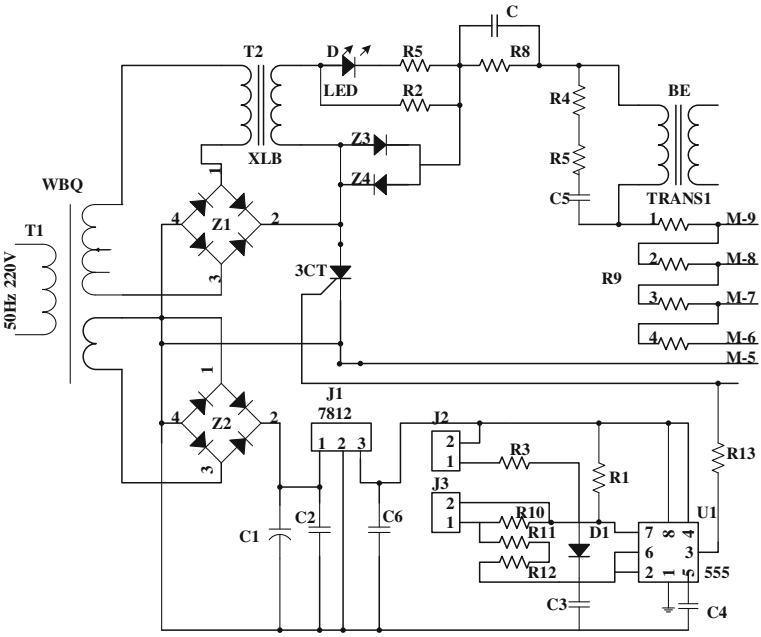


Fig. 51.2 The schematic diagram of traditional analog code generator

and discharging makes the capacitor difficult to be monitored, so the analog code generator will not monitor the capacitor.

The analog code generator output 3 or 3.5 Hz square signals to trigger thyristor's on and off through 555 oscillator to control charging and discharging of the capacitor. When there are some problems in 555 oscillator and/or thyristor, they will not send trigger impulse. This can lead to the failure of asymmetric HVI signal output and impact the whole system. And no impulse trigger self-test function will be as time- and human-consuming when the code generator gets damaged and needed to find damage parts [6]. Not monitoring on the output impulse also makes track circuit work incorrect because of the wrong output impulse.

51.4 Improved Schemes

According to the above problems, improved digital code generator's structure is shown in Fig. 51.3.

For hardware circuit, the following are used: constant current source to provide high-voltage which capacitor's charging and discharging circuit needed; two IGBT instead of the original semi-active controlled thyristor tube; and TMS320F2812 DSP instead of 555 oscillator to generate periodic square wave

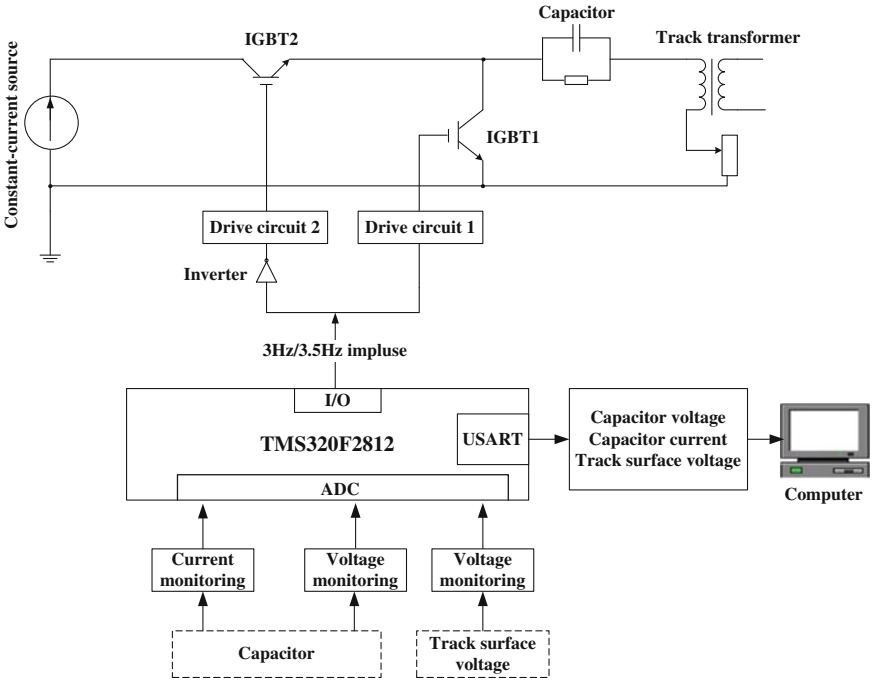


Fig. 51.3 Structure diagram of digital code generator

signal to trigger IGBT's on and off. And at the same time, DSP undertake tasks of signal acquisition, filtering and calculation processing.

DSP produces two ways' same 3 or 3.5 Hz square wave trigger signal, one way through drive circuit 1 to control IGBT1 on and off to charge and discharge the capacitor, and this trigger signal feedback to DSP for checking trigger impulse frequency and driving circuit by itself. The other way is through inverter and drive circuit 2 to control IGBT2 on and off. When IGBT1 is on, capacitor will discharge to the whole circuit, at the same time, IGBT2 is closed and break the connection between discharge circuit and power supply. This solves the problem of short circuit of the power supply caused by capacitor discharging.

For capacitors, charging time is long and discharging time is short; current at the beginning of capacitor's charging under the 50 Hz, 220 V power supply is too large. Long-term repeated charging and discharging can cause capacitor's overheating and then affect working life. Constant current source power supply maintains constant current through the capacitor. According to capacitor's relation of voltage and current, $i = C\frac{\mathrm{d}u}{\mathrm{d}t}$. Integrating both ends of the equation at the same time interval, because the capacitance value C is fixed, we get the final equation to calculate capacitance value:

$$C = i\frac{\Delta t}{\Delta u} \tag{51.1}$$

Slight jitter noise of capacitance waveform will has great impact on the final capacitance calculation result, so take moving average filtering method to pretreat capacitor voltage waveform.

Moving average filtering is the most common filtering method in digital signal processing. In spite of its simplicity, it is optimal for reducing random noise. This makes it the premier filtering method for time domain signals [7]. Moving average filtering operates by averaging a number of points from input signal to produce each point in the output signal. In equation form, this can be written as:

$$y[i] = \frac{1}{M}\sum_{j=0}^{M-1} x[i+j] \tag{51.2}$$

where $x[\ ]$ is the input signal, $y[\ ]$ is the output signal, and M is the number of points in the average. Equation (51.2) only uses points on one side of the output sample being calculated. As an alternative, the group of points from the input signal can be chosen symmetrically around the output point, that is:

$$y[i] = \frac{1}{M}\sum_{j=-\frac{M-1}{2}}^{\frac{M-1}{2}} x[i+j] \tag{51.3}$$

Symmetrical averaging requires that M be an odd number. This filtering method is mainly used to denoise. Often M is smaller, the waveform we can get will smoother, but the signal-rising edge will be destroyed and slowed down. The bigger the M is, the bigger the high-frequency loss is. After the acquisition of voltage smoothing preprocessing, calculate capacitance value after moving average filtering and then compare the difference ratio between calculated value and standard value to determine whether the capacitor is working in normal range according to Eq. (51.1).

It is important for code generator to send qualified asymmetric HVI signal to rail surface.

DSP get head and tail peak voltages of track surface through voltage acquisition. And at the same time, calculate the width of head and tail. There are many differences among asymmetric HVI and typical triangular wave, square wave. While they all have periodic and other same characteristics, the dynamic changes in each periods are not the same. Also the high impulse increases the difficulty of measurement. Usually, we choose the average value method, but its measurement results usually have no direct physical meaning. And peak value method cannot express the most useful part of asymmetric HVI [8, 9].

So we select true root mean square (TRMS) method for measurement of non-sinusoidal wave, and it can reflect the true useful function of measured items in the circuit. The definition of TRMS is shown in (51.4):

$$V = \sqrt{\frac{1}{T}\int_0^T |U^2(t)|\mathrm{d}t} \tag{51.4}$$

However, we can use TRMS value conversion chips, such as AD637, AD536, with a number of auxiliary components to achieve transformation from original voltage to TRMS value. The improved digital code generator has used a microprocessor, so directly use DSP to calculate the TRMS value.

51.5 Experimental Results

Under the constant current source, the method of using DSP to send trigger impulses to control IGBT1's on and off then to control capacitor's charging and discharging is feasible. This method can output correct asymmetric HVI and it is shown in Fig. 51.4.

Take $M = 31$ in Eq. (51.3) to calculate average value for the acquisition voltage of the capacitor. The simulation comparison before and after moving average filtering is shown in Fig. 51.5. TMS320F2812 selects two point intervals of 70 ms between capacitor's charging processes after moving average filtering. According

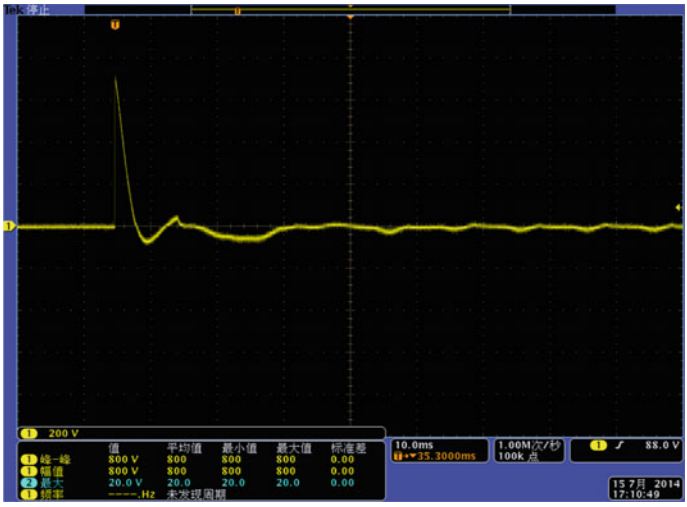


Fig. 51.4 Asymmetric HVI waveform of digital code generator

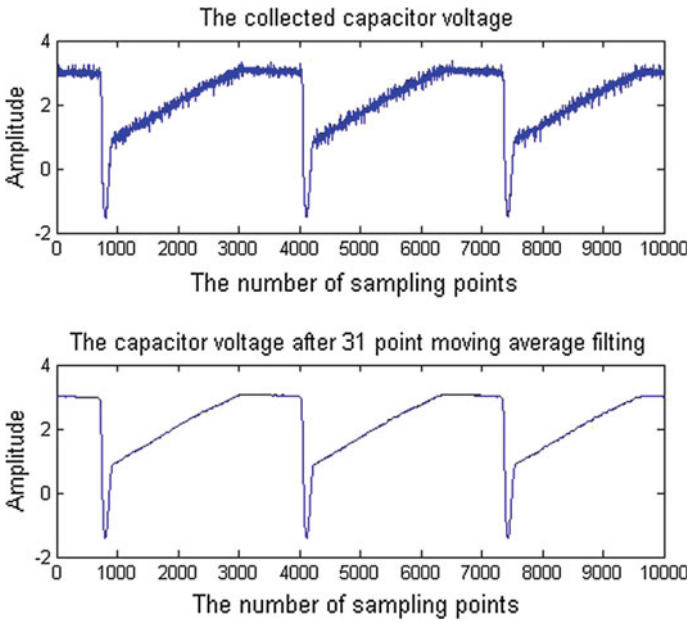


Fig. 51.5 Capacitor voltages before and after moving average filtering

to Eq. (51.1), calculate the capacitance value. Upload calculation results and the acquisition current value to PC and compare them with each standard value for real-time monitoring the capacitor's performance. TMS320F2812 serial port output calculated capacitance values and standard 50 μF value kept errors within ±0.5 %. It is proved that the method is effective and feasible.

Using TMS320F2812 instead of TRMS conversion chip to calculate the TRMS value of asymmetric HVI will make full use of microprocessor, and at the same time will reduce the hardware cost. Taking anti-shake and other anti-interference measures in software, using digital display and real-time monitoring by computer will improve the accuracy of measurement and reading.

51.6 Conclusion

Based on problems in the traditional analog asymmetric HVI track circuit, this paper proposes digital improvement on TMS320F2812 DSP, using constant current source and using two IGBTs to solve power impact problems when capacitor is discharging, and also proposes new calculation and monitoring methods on trigger impulses, capacitor, voltage of rail surface, and other key points. Each part verified by experiments was consistent with expected. The digital asymmetric HVI track circuit still has very large development space in the future. With the continuous

progress of science and technology, making asymmetric HVI track circuit becomes lightweight, convenient maintenance, real-time monitoring, stable working, digitization, and networking will be a trend.

References

1. Yongsheng H (2008) Improvement plan research of track circuit bad shunting. Railw Signal Commun 44(5):24–26 (in Chinese)
2. Jixiang G (1998) The basic operation of the railway signal, 1st edn. China Railway Publishing House, Beijing (in Chinese)
3. Jiayu L, Jin L, Fuzheng Z (2000) An electronic receiver of high-voltage impluse track circuit with dynamic detection. J Shanghai Tiedao Univ 21(12):48–52 (in Chinese)
4. Fuqiang G (2011) Analysis of distributed railway signal generator. Sci Technol Inf 35:52 (in Chinese)
5. Yanyan W (2014) Design of high-voltage pulse track circuit system of stabilizing power supply. Electron Des Eng 22(24):100–102 (in Chinese)
6. Jiuhua X (2008) Study on track circuit flash fault. Railw Signal commun 44(7):29 (in Chinese)
7. Smith SW (1997) The scientist and engineer's guide to digital signal processing. California Technical Pub, PP 277–284
8. Yi C (2014) Discussion about true root mean square value measuring on the asymmetric high voltage impulse signal. Railw Signal Commun 50(2):18–20 (in Chinese)
9. Jing W (2011) Measurement of voltage effective value. J Xi'an Univ Posts Telecommun 16(S1):113–116 (in Chinese)

Chapter 52
Study on Train Operation Adjustment Based on Collaborative Particle Swarm Algorithms

Yazheng An, Yong Qin, Xuelei Meng and Ziyang Wang

Abstract The paper establishes the operation adjustment model in accordance with constraints of the train operation based on the characteristic of high real-time and dynamic and the fitness function aiming at getting the least level of delay time and the number of delay times. In order to solve the established operation adjustment model, the method of cooperation is introduced to improve the typical particle swarm algorithm, which improves the efficiency of operation and the global optimum. Finally, taking the train operation adjustment of dedicated passenger line between Beijing and Shenyang as an example, the model and algorithm are verified. It can be seen that the method is efficient and feasible and it has a good guiding significance for train operation adjustment.

Keywords Operation adjustment · Delay times · Cooperation · Particle swarm algorithm

52.1 Introduction

Train operation adjustment is the core of dispatching, which plays an important role in the stability of traffic order in railway transport organization. The train diagram is the basic documents of the train operation as well as the basis of dispatching work. Train operation adjustment is to adjust the trains not in accordance with the provisions of the diagram based on the existing diagram.

Y. An · Y. Qin (✉) · Z. Wang
Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing 100044, China
e-mail: qinyong2146@126.com

X. Meng
School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_52

As early in 1973, Szpigel [1] firstly proposed the problem on how to determine the optimal scheduling scheme. Then, in 1983, Sauder proposed a model for the optimization of train scheduling problems and applied it to the adjustment and optimization of the single track, which transferred the research from theory to practice for the first time. In 1993, Professor Cao Jiaming put forward the optimized train operation adjustment model of single-track railway [2], which laid the foundation for the optimization modeling of single-track railway train operation adjustment and solved the problem of real-time constraint and complexation of train operation adjustment. In the subsequent study, artificial intelligence algorithm is widely used in solving the train operation adjustment model to meet the requirement of real-time operation adjustment. On the basis of the theoretical research above, strategy of train operation adjustment method transferred from the original relying on subjective knowledge and experience of dispatchers into a scheme of auxiliary computer monitoring for optimal decisions [3–5].

52.2 The Establishment of Train Operation Adjustment Model

52.2.1 Description of the Constraints of Train Operation Adjustment

The paper studies the operation adjustment problems mainly aspects of station operations and interval running time and determines the constraints of time conditions of stations in an appointed section concretely. Assuming that starting point of the interval is t_0 and $t_0 = 0$. The length of period is A (minute) and the final time is $t_L = t_0 + A$. There are M stations in the selected section and the number of trains is N in the period. $L_{i,s}$ indicates the actual departure time of train i from station s and $L^0_{i,s}$ is the planning. $d_{i,s}$ indicates the actual arriving time of train i to station s and the planning is $R^0_{i,s}$. $t^z_{i,s}$ is the minimum operating time of train i in station s and $t^i_{s,s+1}$ is the minimum pure running time of train i from station s to station $s + 1$. $a_{i,s}$ can be described by formula 52.2.1 as follows:

$$a_{i,s} = \begin{cases} 1, & L_{i,s} \neq R_{i,s} \\ 0, & L_{i,s} = R_{i,s} \end{cases} \tag{52.2.1}$$

Then, define a variable as follows:

$$f(a,b) = \begin{cases} 1, & a > b \\ 0, & a \leq b \end{cases} \tag{52.2.2}$$

For trains in double line with automatic block, the following constraints should be taken in consideration:

(a) The limitation of earliest departure time

$$L_{i,s} - L_{i,s}^{0} \geq 0 \tag{52.2.3}$$

(b) The constraint of the running time in the section

$$L_{i,s} + a_{i,s}t_{s}^{q} + t_{s,s+1}^{i} + a_{i,i+1}t_{s+1}^{t} \geq d_{i,s+1} \tag{52.2.4}$$

(c) Constraint of operational time in station

$$R_{i,s} + a_{i,s}t_{i,s}^{z} \leq L_{i,s} \tag{52.2.5}$$

In addition to ensuring the operating conditions of single train, we should also take the restrictions between the trains into consideration.

(a) The constraint of tracing time interval

$$L_{i+1,s} - L_{i,s} \geq I_{z,s} \tag{52.2.6}$$

$$R_{i+1,s} - R_{i,s} \geq I_{z,s} \tag{52.2.7}$$

(b) The constraint of departure and arriving between the trains from the same direction

$$R_{i,s} + \tau_{s}^{RL} \leq L_{i+1,s} \tag{52.2.8}$$

$$R_{i+1,s} - R_{i,s} \geq I_{z,s} \tag{52.2.9}$$

Due to the limited capacity of the equipment in the station, the constraints should be described as follows:

(a) The capacity of the railway lines in the station

$$\sum_{i} f(L_{i,s}, t) - \sum_{i} f\left(R_{i,s}, t\right) \leq l_{s}, \quad \mathrm{t} \in (t_0, t_0 + A) \tag{52.2.10}$$

52.2.2 Establishment of the Fitness Function

The goal of train operation adjustment is to make sure that the trains out of the graph restore to the planning graph as soon as possible, which can be measured by the following aspects:

(a) Minimum total delaying time of arriving

$$g_1 = \min \sum_{s=1}^{M} \sum_{i=1}^{N} \left(R_{i,s} - L_{i,s}^0 \right) \tag{52.2.11}$$

(b) Minimum total delaying time of departing

$$g_2 = \min \sum_{s=1}^{M} \sum_{i=1}^{N} \left(L_{i,s} - L_{i,s}^0 \right) \tag{52.2.12}$$

(c) Minimum delaying times

$$g_3 = \min \sum_{s=1}^{M} W_s \tag{52.2.13}$$

$$W_s = \sum_{i=1}^{N} f\left(R_{i,s}, R_{i,s}^0 \right) \tag{52.2.14}$$

α, ρ, and β represent the degree of importance of the index, respectively. In the actual condition, the delaying time of different level of the train dose difference effect on the operation. Thus, we describe the levels trains as μ_i and the μ_i of common, general fast and fast trains are valued of 3, 2, and 1, respectively. Then, the final objective function can be expressed as follows:

$$g = \min \sum_{s=1}^{M} \sum_{i=1}^{N} \mu_i \left[\alpha \left(R_{i,s} - R_{i,s}^0 \right) + \beta W_s + \rho \left(L_{i,s} - L_{i,s}^0 \right) \right] \tag{52.2.15}$$

52.3 Collaborative Particle Swarm Optimization in Train Operation Adjustment

The basic particle swarm algorithm is more likely to fall into the local optimum and it takes more iteration to get the final solution. Following the collaborative symbiotic species in nature, we divide the particle into several groups under the

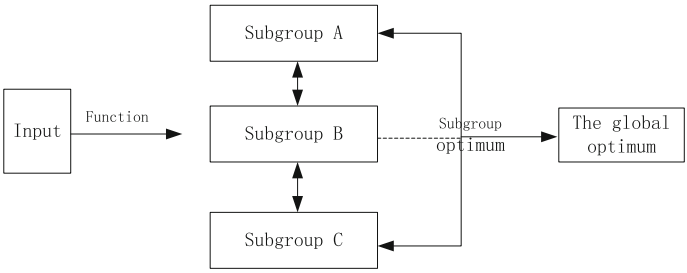


Fig. 52.1 Diagram of cooperative particle swarm algorithm principle

principle of mutual benefit and reciprocity and each of groups has the different characteristics. Information is exchanged between particle and particle group, individuals and individuals, and the global optimal solution is determined finally [6–8]. That is the basic idea of the cooperative particle swarm algorithm.

Each of the subgroup searches, respectively, in the target space of D dimensions according to different mechanisms, and the position of each particle of the subgroup is on behalf of the result of the problem. As the independent particles evolution happened, information is exchanged and shared among the subgroups; thus, the global optimal solution is obtained by means of cooperation. The basic form of the algorithm can be shown in Fig. 52.1.

Assuming that particle population size is M and the searching space is in N dimensions. $X_i^t = x_{i1}^t, x_{i2}^t, \ldots, x_{in}^t, \quad i = 1, 2, 3, \ldots, M$ are the coordinates of particle i, and $V_i^t = (V_1^t, V_2^t, \ldots, V_{mi}^t)$ is the velocity. Thus, the position and velocity can be described as follows:

According to the basic idea of particle swarm optimization algorithm, the position and velocity updating formulas of cooperative particle swarm algorithm in this paper can be described as follows:

$$V_{ij}^t = \omega v_{ij}^{t-1} + c_1 r_1 \left(p_{ij}(t-1) - x_{ij}^{t-1}\right) + c_2 r_2 \left(g_j^a(t-1) - x_{ij}^{t-1}\right) + c_3 r_3 \left(g_j^b(t-1) - x_{ij}^{t-1}\right) + c_4 r_4 \left(g_j^c(t-1) - x_i^{t-1}\right) \tag{52.3.1}$$

$$v_{ij}^M(\mathrm{t}) = \begin{cases} v_{\max}, & v_{ij}^t > v_{\max} \\ v_{ij}^t, & v_{ij}^t \le v_{\max} \end{cases} \tag{52.3.2}$$

$$x_{ij}^M(t) = x_{ij}^{t-1} + v_{ij}^t \tag{52.3.3}$$

where

$g_j^a(t-1)$ is the present optimal location of individuals in subgroup A after $t-1$ iteration.

$g_j^b(t-1)$ is the present optimal location of individuals in subgroup B after $t-1$ iteration.

$g_j^c(t-1)$ is the present optimal location of individuals in subgroup C after $t-1$ iteration.

c_2, c_3, c_4 are learning factors, which, respectively, denote the effect on whole group by subgroup A, subgroup B, and subgroup C.

$v_{ij}^M(\mathrm{t})$ is the velocity vector of the main group after t iteration.

$x_{ij}^M(t)$ is the optimal position of the main group in the Jth dimensional after t iteration.

As same as the determination of position vector, the amount of change of position vector in each space is corresponding to the velocity vector of the particle in the corresponding space. The velocity vector of the particle can be described as follows:

$$V_t = \begin{bmatrix} v_{1,1}^t & v_{1,2}^t & v_{1,3}^t & \cdots & v_{1,2MN}^t \\ v_{2,1}^t & v_{2,2}^t & v_{2,3}^t & \cdots & v_{2,2MN}^t \\ v_{3,1}^t & v_{3,2}^t & v_{3,3}^t & \cdots & v_{3,2MN}^t \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ v_{m,1}^t & v_{m,2}^t & v_{m,3}^t & \cdots & v_{m,2MN}^t \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ v_{K,1}^t & v_{K,2}^t & v_{K,3}^t & \cdots & v_{K,2MN}^t \end{bmatrix} \tag{52.3.4}$$

The velocity vector and the position vector in single dimension update according to formula 52.3.4 after the $t+1$th iteration. Velocity vector of particles represents refers to the changing of the time in the train operation adjustment model.

52.4 The Process of Algorithm Implementation

The parameters of the algorithms can be set according to the actual situation of the problem and the steps of the algorithm are as follows:

Step1: Initialize the position and velocity of each particle in the three subgroups in allowable range and set group A as a master group. (The setting of master group is just reference, and we can get the same effect whatever subgroup

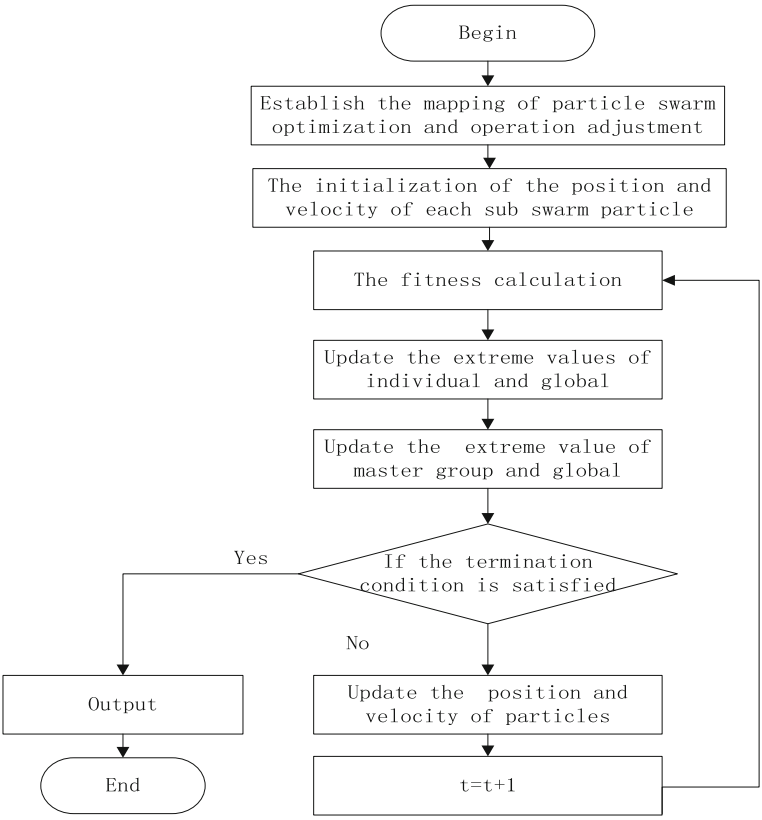


Fig. 52.2 Flow process diagram of the cooperative particle swarm algorithm

sets as the master group, and finally, all the particles converge with the optimal region.)

Step2: Compare the fitness corresponding to the position of each particle with the fitness of its best ever position at present and mark it as the better position which corresponding to a better fitness.

Step3: Update the position of particle in each subgroup;

Step4: Update the position of each subgroup according to the velocity formula and the location formula of the basic particle swarm algorithm, respectively;

Step5: Check whether a termination condition is satisfied and if it did terminate the loop and the given output of the optimal solution set and the optimal fitness value, or return to step2 and continue the cycle.

Step6: The flow process diagram of the cooperative particle swarm algorithm in this paper can be shown in Fig. 52.2.

52.5 Analysis on Actual Problem

Taking the section of Beijing–Shenyang passenger dedicated line as the background, we can solve the operation adjustment model with the cooperative particle swarm algorithm (Table 52.1).

There are 9 docked along stations in the section of Beijing–Shenyang of Beijing Shanghai passenger dedicated line, which divide the section into 8 smaller sections. In order to verify the train operation adjustment model and the feasibility of the algorithm, the paper chooses 4 h (15:00–19:00) as the adjustment object.

The scale of each group is set as 50 and the initial values for the parameters in the algorithm are shown in Table 52.2.

The result shows that the trains are delayed for 23 times in total. The delay of arrival time is 557 min and the delay of departure is 560 min. Thus, the value of fitness function is g = 247.5. In the process of train operation adjustment, the

Table 52.1 The original timetable

Timetable (Beijing–Shenyang)								
Stations	F/D	D27	D23	D51	D7	D1	D3	D13
Beijing	D	–	–	–	–	–	–	–
	F	15:15	15:33	17:13	17:45	18:17	18:23	19:01
North of Tangshan	D	16:29	16:46	18:24	18:54	19:32	19:36	20:00
	F	16:29	16:50	18:30	18:58	19:32	19:40	20:04
Qinhuangdao	D	17:34	17:55	19:36	20:04	20:37	20:55	21:07
	F	17:34	17:59	19:40	20:08	20:37	20:59	21:11
North of Huludao	D	18:34	18:56	20:45	21:07	21:40	21:53	22:12
	F	18:38	19:00	20:48	21:07	21:44	21:57	22:16
South of Jinzhou	D	18:54	19:14	21:05	21:24	21:59	22:11	22:31
	F	18:54	19:14	21:05	21:26	21:59	22:11	22:31
North of Panjin	D	19:18	19:36	21:30	21:50	22:24	22:34	22:54
	F	19:15	19:40	21:30	21:55	22:26	22:38	22:54
North of Shenyang	D	20:10	20:30	22:20	22:43	23:10	23:20	23:34
	F	20:13	20:35	–	–	–	–	–

Table 52.2 Value of parameters

Parameters	Values
ω	0.9
c_1	1.5
c_{2s}	1.5
c_3	0.5
c_4	0.5

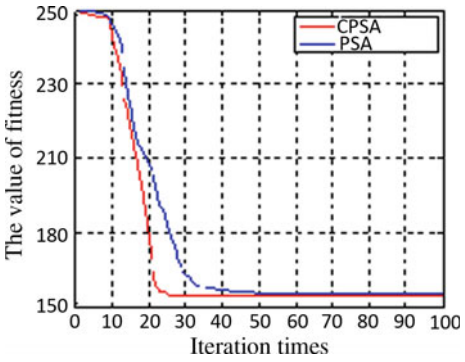


Fig. 52.3 The fitness changes with iteration

punctuality of originating should be ensured and the delayed time should be caught up with. Thus, the result is consistent with the basic principles of operation adjustment.

Figure 52.3 depicts the value of fitness function got by cooperative particle swarm (CPSA) algorithm and basic particle swarm algorithm (PSA) with the increasing of the iterations. It can be seen that when the cooperative particle swarm algorithm is used, the fitness value tends to be stable after 25 iterations. Therefore, the value tends to be stable after 50 iterations if we use the basic particle swarm algorithm in this research. For the perspective of the optimal degree of the optimal solution, an improved cooperative particle swarm algorithm performs more advantage.

Figures 52.4, 52.5, and 52.6 are reactions of changes of the single targets with the increasing of the iterations. As can be seen from the graph, the overall variation is adapted to three objective functions. Thus, a certain relationship in the objective causality was existed between the functions and the algorithm is applicable.

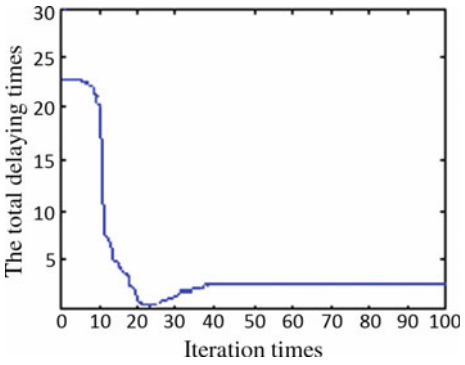


Fig. 52.4 The total delaying times changes with iteration

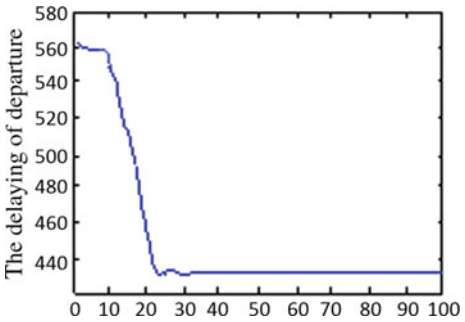


Fig. 52.5 The delaying time of departure changes with iteration

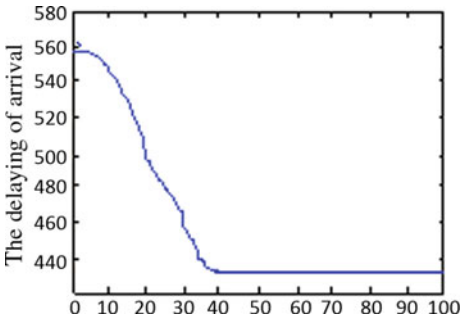


Fig. 52.6 The delaying time of arriving changes with iteration

52.6 Conclusion

The paper discusses the core of scheduling problem—operation adjustment problem according to realities rail transport scheduling. The constraints of operation adjustment are proposed and the corresponding objective function is established by the analysis of the basic process of train operation adjustment and basic principles and the common strategy for train operation adjustment. In addition, the paper presents a highly efficient intelligent optimization algorithm—particle swarm algorithm for the characteristic of the problem. The idea of cooperation was introduced on the basis of the basic particle swarm algorithm, which improves the basic particle swarm algorithm makes it more efficient.

After verification, cooperative particle swarm algorithm proposed in this paper performs better in solving the operation adjustment problem than basic particle swarm algorithm dose. The improved algorithm can converge to the optimal solution and get the better fitness value with less iteration.

Acknowledgment This work was supported in part by the National Natural Science Foundation of China under Grant (61374157).

References

1. Szpigel B (1972) Optimal train scheduling on a single line railway. In: Proceedings of conference on operational research. Amsterdam, the Netherlands:[s,n]
2. Jiaming C (1994) The optimization model and its algorithm for adjusting train diagram on single track railway lines. J China Railw Soc 3(16):72–78
3. Wen C, Li B (2013) Train operation adjustment based on conflict resolution for high-speed rail. J Transp Secur 6(1):77–87
4. Chen Y, Yu J, Zhou L, Tao Q (2013) Study on the algorithm for train operation adjustment based on ordinal optimization. Adv Mech Eng
5. Yong-Jun C, Lei-Shan Z (2010) Study on train operation adjustment algorithm based on ordinal optimization. J China Railw Soc 32(3):1–8
6. Jiang Y-S, Chen Y-R, Pu Y (2005) Study on dynamically integrated model and algorithm for problem of train operation adjustment based on multi-agent system. 26(1):114–118
7. Chuanjun J, Siji H, Yudong Y (2006) Study on the particle swarm optimization algorithm for train operation adjustment. J China Railw Soc 28(3):6–11
8. Zhang J, Dai D (2012) A particle swarm optimization algorithm based on the pattern search method. Adv Mater Res 532:1664–1669

Chapter 53
Traffic Evacuation Plan Research of Dongli Square

Xue Han, Xi Zhang and Yongming He

Abstract With the rapid development of the economy and the improvement of the urbanization level, our city is increasing traffic volume, in which the traffic problem is appearing gradually, especially in the important city or in the big city. This article mainly aims at the example of transformation of Harbin city Dongli roundabout based on the collection of data, carries on the analysis to the traffic situation and put forward ideas and strategies to improve the optimization of the intersection, optimization ideas and improving measures will make the traffic simulation. Finally, I hope the analysis of Harbin Dongli roundabout improvement examples can help other similar circular intersection improvement.

Keywords Roundabout · Traffic organization · Optimization · Traffic simulation

53.1 Intersection Research Status of Dongle Square in Harbin

53.1.1 Road Situation

Around of square of Dongli in Harbin is the intersection of San da Dong Li road, Hexing road, Haping road, and Heping road which is also the important node in second of Harbin. Around the island, they are Songlei mall, Lesong square, Paris square, Middle of underground shopping malls, and other large shopping, entertainment places.

Central island diameter of Dongli square intersection is about 100 m, four-lane ring, and width is about 15 m, widened import way is horn.

X. Han · X. Zhang (✉) · Y. He
School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China
e-mail: xizhang@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_53

Fig. 53.1 Roundabout of Dongli square in Harbin

1. The road section type. The four corners of the world of import way is two-way six lanes and central separation by the police. Motor vehicles and non-motor vehicles are separated by planting green zone. The direction of west and east (2) has a span bridge, and the direction of north and south has a wear under the underground passage for vehicles straight of the four corners. The underground passage is two-way four lanes. Near the four directions around the island of Dongli square intersection, all have switching lanes, reducing the pressure on the roundabout.
2. The traffic organization situation: the intersection is an roundabout, no signal control intersections. The specific form is shown in Fig. 53.1.

In conclusion, Dongli square traffic problems can be summarized as follows [1, 2]:

1. Traffic organization is poor and traffic capacity is low,
2. Non-motor vehicles, pedestrians, motor vehicle interference are serious, and
3. Road network planning is not reasonable.

53.1.2 Traffic Volume Survey

1. Survey plan
   The survey relies on the entrance ramp and four ring road sections as the observation section including 12 investigators to investigate and each entry shall be charged by two people who, respectively, statistic all imports into the loop flow and traffic on your left. Four sections are charged, respectively, by four

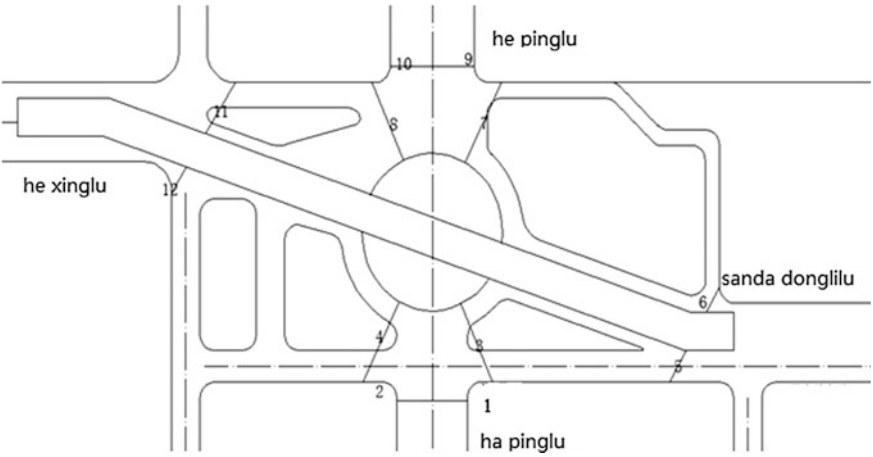


Fig. 53.2 Distribution of the intersection location survey data

Table 53.1 Summary of intersection traffic survey data

	Entrance number			
	1	2	3	4
Q in i	2909	2389	2605	2771
Q right i	1004	720	492	660
Q interweave i	5537	4726	5187	5868
Q out i	3240	3200	2144	2090

where Q_i the volume of the section i

people for investigation and each investigator is responsible for a section of each in the morning rush hour 7:30–8:30 continuous investigation for three days as shown Fig. 53.2.

2. Survey data (Table 53.1)

53.1.3 Dongli Square Intersection Capacity Calculation

Capacity measured intersection traffic capacity index, is whether intersection can satisfy the requirement of traffic index and the basic requirement of intersection design, the certain service level. Thus, in the intersection design or improve, the capacity of research is necessary [3–5].

Table 53.2 Weaving period of utilization

Interlaced section	Design capacity (Car/h)	Rush hour traffic flow (Car/h)	Saturation rate
Northeast	3128	5187	1.66
Southeast	4461	5537	1.24
Southwest	2947	4726	1.60
Northwest	4927	5868	1.19
The total	15463	21318	1.38

Interim formula of the Environment Agency

$$C_{\mathrm{m}} = \frac{160w\left(1 + \frac{e}{w}\right)}{1 + \frac{w}{l}}(/h)$$

where

C_{m} Weaving section on maximum capacity
C_{d} ($C_{\mathrm{d}} = 0.8C_{\mathrm{m}}$); weaving segment design capacity
l Weaving length (m)
W Weaving section width (m)
$e1$ Entrance ramp width (m)
$e2$ The ring section width (m)
e The average width of the entrance ramp (m)

It is concluded that the weaving period of utilization, as shown in Table 53.2.

Table 53.2 shows that the capacity of various weaving sections at peak hours of Dongli square intersection has reached the saturated state, and the saturated rate is 1.38. Now, the roads taking weaving segment into consideration simply at rush hour cannot meet the normal operation, so the improvement of the intersection is urgent.

53.2 Dongli Square Traffic Evacuation Plan

Now, the buildings near the Dongli square intersection have been basically formatted, and there are important buildings and large shopping malls, and there has built up across the bridge of east and west and the underground passage of north and south. However, traffic congestion problem is still not solved [6]. According to the situation around the Dongli square intersection, increasing the circle area is not desirable. Therefore, the measures to solve the roundabout can only be changed roundabout to the signal control level crossing or using around the streets and the traffic management measures of the roundabout to evacuate.

Table 53.3 Traffic signal timing design calculation table

1		2	3	4	5	6	7	8	9	10	11	12	13
West	Left	913	1	1800	0.13	Phase 1: 0.25	0.8	12	122	110	34	0.9	32
	Straight	992	2	2000	0.25								
	Right	1004	1	1550	0.16								
East	Left	517	1	1800	0.14	Phase 2: 0.14					20		18
	Straight	1596	2	2000	0.2								
	Right	492	1	1550	0.16								
North	Left	432	1	1800	0.12	Phase 3: 0.22					29		27
	Straight	1679	2	2000	0.21								
	Right	660	1	1550	0.22								
South	Left	984	1	1800	0.2	Phase 4: 0.2					27		25
	Straight	685	2	2000	0.16								
	Right	720	1	1550	0.11								

The chart: *1* entrance lane; *2* traffic volume *Q*; *3* the lane channelizing scheme; *4* saturation flow *Q*; *5* flow ratio *Y*; *6* phase maximum flow ratio; *7* the sum of flow ratio; *8* the total loss of time *L*; *9* cycle length *T*; *10* total effective green light time Ge; *11* effective green light time Ge; *12* spilt λ; *13* green light show time *g*

53.2.1 Dongli Square Roundabout Changes into Signal Control Intersection

According to the survey, the problem about Dongli square intersection, Loxonin roundabout transformation is to dismantle center around the island to signal control intersection, construct across the bridge, and build turn left to area. According to the actual situation, to design the import and export of shunt measures and plan a pedestrian crossing.

Dongli square intersection for intersection traffic flow graph is shown as Table 53.2 based on field survey which concludes that intersection traffic signal timing design calculation table, as Table 53.3.

The intersection cycle is 122 s, phase is 4, digital phase loss calculation is $L_s = 3$ s, total loss is $L = 12$ s, total effective green light time is 110 s, and then, the intersection signal timing diagram is shown in Fig. 53.3.

53.2.2 Dongli Square Intersection Using Traffic Management Measures to Improve It

Dongli square carry the traffic flow from Ha Ping road, He Ping road, and San Da Dong Li road. At the same time, it connects Dong Yuan street and Ti Yu Tou Dao small street. But small street traffic is smaller. The main improvement measures of Dongli square intersection are to use the perfect traffic facilities and the computer

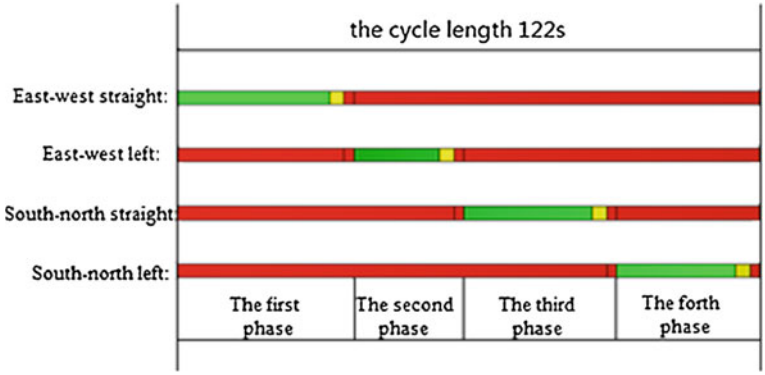


Fig. 53.3 Signal timing diagram

technology and the necessary traffic organization and management measures, unified management of the area of the whole road network, through the roundabout traffic diversion, to achieve the entire road network the smooth flow of traffic, and balance state.

53.3 Harbin Dongli Square Intersection Traffic Simulation

VISSIM is a kind of microcosmic, based on the time interval and driving behavior simulation modeling tool, used for modeling and analysis of various traffic conditions (lane, traffic composition, traffic signal, bus stops, etc.). The circulation of the urban traffic and public transport is the evaluation of transportation engineering design, and urban planning is an effective tool [7, 8].

Using VISSIM software, the current situation of Dongli square intersection is simulated in Harbin as shown in Fig. 53.4.

53.3.1 The Traffic Simulation of the Improved Dongli Square Intersection

According to the second chapter, the no signal control intersection changes into signal control scheme, using the street near the intersection. For example, the intersection of Ha Ping road and Xing Fu road and the intersection of He Ping road and Nong Lin street can disperse some traffic flow. The application of VISSIM traffic simulation software is shown in Fig. 53.5.

Fig. 53.4 Dongli square intersection traffic status simulation diagram

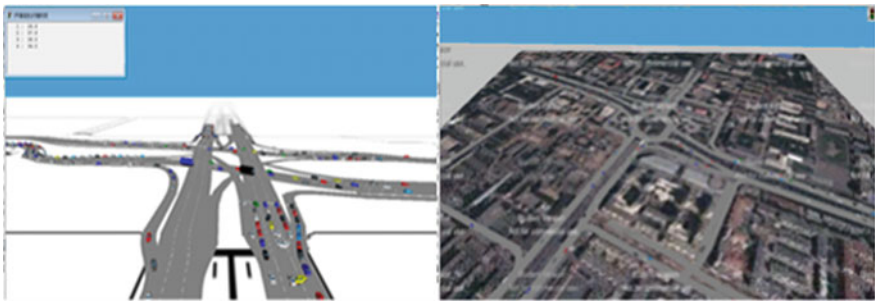


Fig. 53.5 The traffic simulation of the improved intersection

53.3.2 Dongli Square Intersection Improvement Scheme Comparison

According to the improvement plan of the third chapter on Dongli square intersection, to process traffic simulation and get the traffic simulation contrast. Detailed traffic simulation analysis is shown in Table 53.4.

From the simulation parameters, it can be seen that after the two options, travel time has significantly reduced. It proves that improving the efficiency of vehicles is increased.

From the demonstration, it can be seen that traffic order is improved substantially, vehicle delay time is reduced, the time of vehicle parking in the intersection is reduced, traffic interference phenomenon is disappeared, and vehicle becomes more secure, security cost is less, and fast transportation is provided.

To sum up, from the point of the simulation parameters and demonstration singly, the change from roundabout to signal control intersection is more

Table 53.4 Traffic simulation comparison chart

Section travel (time/s)	Ha Ping road to He Ping road	He Ping road to Ha Ping road	He Xing road to Sanda Dongli road	Sanda Dongli road to He Xing road
Present condition	149.7	87.4	127.4	172.6
Improvement 1	24.8	47.8	33.4	34.0
Improvement 2	45.8	76.5	68.8	102.4

where *improved 1* change into signal control intersection
Improved 2 using traffic management measures to improve after

reasonable. But, for the above two kinds of schemes, when I was in traffic simulation, I just use the streets around the circle part of shunt evacuated through the traffic circle, which do not join some traffic facilities, such as variable message signs and road signs marking, so it remains to be further research.

53.4 Conclusion

According to the scheme comparison, when the intersection traffic conflict, just increasing the general light controls traffic capacity of intersection or the improvement of service level which cannot have obvious improvement effect, should not be commonly used. From the point of the simulation parameters and demonstration, the change from roundabout to signal control intersection is more reasonable. But, for the above two kinds of schemes, when I was in traffic simulation, I just use the streets around the circle part of shunt evacuated through the traffic circle, which do not join some traffic facilities, such as variable message signs and road signs marking, so it remains to be further research. According to the comparison of quantity and economic conditions, it can select a reasonable plan: the no signal intersection upgrade to signal control intersection.

References

1. Fan L (2007) The development of the annular plane intersection design theory and its application in our country. Private Sci Technol 3:115 (in Chinese)
2. Yang X (2002) Urban road traffic design guidelines. Beijing People Publishing House, pp 144–148 (in Chinese)
3. Gao X, Zhu R, Jin W (2009) The reformation design of Urban road roundabout. In: The 18th annual academic essays of the national urban road society (in Chinese)
4. Thai van M-J (2000) The design of Roundabouts in France: historical context and state of the art. TRB, Washington, DC
5. Yuan Jingyin, Yuan Zhenzhou (2006) Comparative analysis of signal intersection capacity calculation method. Technol Road Transp 5:123–128 (in Chinese)

6. Shen L (2007) The simulation study of urban traffic flow based on Aimsun. Xi'an the master's thesis in Changan University (in Chinese)
7. Zhu J, Suni Q (2010) Big city intersection traffic improvement strategy analysis-take the improvement of the intersection of Chongqing urban area as an example. Highw Automot Appl (1):61–63 (in Chinese)
8. National Research Council Highway Capacity Manual (2000) Transportation research board, Washington, DC, pp 16-1–16-9

Chapter 54
Cost Optimization of Empty Shipping Container Allocation Based on Integer Programming

Mingming Zheng and Guoqiang Cai

Abstract This paper builds empty shipping container allocation model on the base of integer programming on the premise of meeting heavy container transportation. In this model, the cost of allocation includes transportation cost, container leasing costs, and loading–unloading costs. The aim of objective function is to make the allocation cost minimum. In the circumstances of constraint conditions, many factors such as ship transport capacity, transport time, empty container quantities at supply ports, and quantities at demand ports are considered synthetically. Finally, the model is verified by an actual example and data, using LINDO software to calculate and determine the optimal scheme of empty container allocation.

Keywords Empty container allocation · Cost optimization · Integer programming · LINDO

54.1 Introduction

At present, the phenomenon of imbalanced freight flow exists in almost all areas in the world caused by asymmetric trade and seasonal changes of freight transport, and then, problem of empty container allocation was produced, which is one of the most difficulties that liner shipping companies faced. However, the import and export container distribution at China's coastal ports is extremely uneven, so how to realize the optimization of empty container allocation and arrange it reasonably is of great significance for reducing the cost of container operating.

The optimization problem of empty marine container allocation draws scholars' attention at home and abroad in recent years. Cheung and Chen [1] constructed dynamic and random two stage of marine container allocation network model. In the first stage, supply and demand of empty container and ship's remaining capacity

M. Zheng · G. Cai (✉)
State Key Lab of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing, China
e-mail: guoqiangcai@163.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_54

are regarded as deterministic variables; in the second stage, change these parameters to stochastic dynamic ones which are closer to actual decision than deterministic model. Choong et al. [2] considered prior principle of heavy container in practice, then set up integer programming model of empty container allocation based on minimum cost, and finally, they gave an empty container allocation case in Mississippi River and discussed how much the planning period length affect the decision making. Li et al. [3] did some research on the problem of empty container dispatching based on inventory theory and built a new inventory model which combined the penalty function features to solve the problem of imbalanced demand of empty containers. Lam et al. [4] solved the problem of empty container allocation by dynamic programming. They built dynamic stochastic model based on the dispatching system of two ports and two routes and conducted a simulation to verify the validity of the approximate optimal solution. Meng and Wang [5] combined liner route design with empty container allocation and investigated problem of marine empty container allocation in the situation which its route is uncertain. Cheang and Lim [6] considered problem of empty container allocation combining the third-party container leasing strategy. A decision support system was given to solve the allocating problem by using network flow method. Chang [7] determine the number of empty containers a port demanding, which is often uncertain, by means of economic order quantity (EOQ). Chen and Hsieh [8] proposed a model, called as fuzzy inventory decision making, to determine empty container optimal quantities at a port.

54.2 Model of Empty Container Allocation Establishment

54.2.1 Assumptions of Model

Due to the actual business of container transport that is extremely complex and a lot of variable factors, many difficulties are posed that we can not solve them easily. In order to build the model conveniently, some optimization assumptions are needed for marine container transportation system.

Assumption 1: Supply and demand of empty container at each ports has been known, and supply and demand ports have been determined.
Assumption 2: All are same type containers, without considering container type restriction.
Assumption 3: Not considering problems of repairing containers and abandoned containers.
Assumption 4: Empty container can be promptly leased at each port, only considering empty container leasing costs.
Assumption 5: Empty container transport directly from the supply to the demand.

Assumption 6: All of the empty container allocations occur in working time period T. That is to say, no empty container flows before period of $k = 1$, and no empty container flows after period of $k = T$, too. Working cycle divides into T stages, one stage represents one day.

54.2.2 Model Establish

For establishing the model, the following symbols are introduced.

i Supply port;
j Demand port;
T Planning period span;
S Assemblage of supply ports within the planning period;
D Assemblage of demand ports within the planning period;
$x_{ij}^{kk'}$ Number of empty containers, being shipped from port i to j with the time from k to k';
$c_{ij}^{kk'}$ Transportation cost per empty, being shipped from port i to j with the time from k to k';
t_{ij}^{k} Required time, empty containers from port i to j;
n_{i}^{k} Amount of empty containers at port i in time k;
c_{li}^{k} Loading cost per empty, at port i in time k;
$n_{j}^{k'}$ Quantity demanded of empty containers, at port j in time k';
$c_{ulj}^{k'}$ Unloading cost per empty, at port j in time k';
$c_{rj}^{k'}$ Leasing cost per empty, at port j in time k';
$v_{ij}^{kk'}$ Maximum capacity of ships from port i to j with the time from k to k';

Two time series generators are defined: $D\left(x_{ij}^{kk'}\right) = k, M\left(x_{ij}^{kk'}\right) = k'$; the returning value of U is the earliest time of shipment for $x_{ij}^{kk'}$, The returning value of M is the latest unloading time for $x_{ij}^{kk'}$.

In this paper, we consider three kinds of costs, which are transportation cost, loading–unloading cost, and leasing cost.

Transportation cost:

$$f_1 = \sum_{\substack{k,k'=1,\ldots,T \\ k' \geq k}} \sum_{i \in S, j \in D} x_{ij}^{kk'} \times c_{ij}^{kk'}$$

Loading–unloading cost:

$$f_2 = \sum_{\substack{k,k'=1,\ldots,T \\ k' \geq k}} \sum_{i \in S, j \in D} x_{ij}^{kk'} \times \left(c_{ij}^{kk'} + c_{uli}^{k'} \right)$$

Leasing cost:

$$f_3 = \sum_{\substack{k,k'=1,\ldots,T \\ k' \geq k}} \sum_{i \in S, j \in D} c_{rj}^{k'} \times \left(n_j^{k'} + x_{ij}^{kk'} \right)$$

In summary, the aim of objective function in the model is to make the total cost minimized, and objective function is as follows:

$$\min Z = f_1 + f_2 + f_3$$

Constraint condition:
Amounts ≤ Maximum capacity of ships.

$$x_{ij}^{kk'} \leq v_{ij}^{kk'}$$

Shipping out amounts ≤ Empty container amounts at a port.

$$\sum_{\substack{k,k'=1,\ldots,T \\ k' \geq k}} \sum_{i \in S, j \in D} x_{ij}^{kk'} \leq n_i^k$$

The time of all containers entering and leaving ports is feasible, that is to say, all containers enter and leave ports just in working time period T.

$$U\left(x_{ij}^{kk'}\right) + t_{ij}^{U\left(x_{ij}^{kk'}\right)} - M(x_{ij}^{kk'}) \leq 0$$

Nonnegative constraint:

$$x_{ij}^{kk'} \leq 0 \text{ and integer}, \quad i \in S;\ j \in D;\ k' \geq k;\ k, k' = 1, 2, \ldots, T$$

Among each formula above, k stands for setting out from supply port in time k, k' stands for arriving at demand port in time k', and thus, each formula concluded k and k' requires $k \leq k'$.

54.3 Algorithm Description and Computation

54.3.1 Integer Programming

Integer programming is a branch of mathematical programming, studying a class of optimization problems which require some or all of variables rounded to integers. If all of independent variables are rounded to integers, we call it all integer programming or pure integer programming problem. If a part of variables are integers, we call it mixed integer programming problem. Integer programming is called as IP problem for short. The major discussed problem of this paper is integer linear programming, called ILP problem for short.

General frameworks of solving an integer programming problem are as follows: generating a derivative problem of one original problem and then producing another problem, called a slack problem, which is easier than the derivative problem. (The derivative problem is called original problem of the slack problem.) Through the solution of slack problem, we can determine the fate of its original problem, that is, original problem has been solved or generating one or several its derivative problems which replace itself. Then, choose one derivative problem of original problem that has been not discarded and solved and repeat the above step until unsettled derivative problem not remains. We have developed some common integer programming algorithms since the 1960s, such as all kinds of cutting plane method, branch and bound method, implicit enumeration method of solving 0-1 programming, decomposition method, group theory method, and dynamic programming method. Some people have developed near-computer and computer analogy approach in the last decade, which achieves great effect.

54.3.2 Computation Description

Most data in this computation case are obtained by researching on shipping enterprise, and some data are referred to corresponding magazines; thus, this computation case is close to actual circumstance.

Four supply ports and five demand ports are assumed in this computation case, five days as a cycle of working time, that is, $T = 5$; we divided T into five stages. $c_{tij}^{kk'}$, c_{li}^{k}, $c_{ulj}^{k'}$, $c_{rj}^{k'}$ and $v_{ij}^{kk'}$ remain unchanged, unrelated to time; t_{ij}^{k} is not related to original time. The information such as supply and demand data of empty containers in the system is known, determining how to combine empty container allocation with leasing so that making the total cost of empty container transportation minimized (Tables 54.1, 54.2, 54.3, 54.4, 54.5, 54.6).

Table 54.1 Transport cost between ports (RMB/TEU)

Port	D_1	D_2	D_3	D_4	D_5
S_1	65	98	65	156	98
S_2	78	72	182	163	65
S_3	104	65	98	130	91
S_4	208	91	156	98	104

Table 54.2 Loading cost (RMB/TEU)

Port	S_1	S_2	S_3	S_4
C_{li}	26	13	20	33

Table 54.3 Unloading cost (RMB/TEU)

Port	D_1	D_2	D_3	D_4	D_5
C_{ulj}	20	7	13	13	26

Table 54.4 Leasing cost (RMB/TEU)

Port	D_1	D_2	D_3	D_4	D_5
C_{rj}	65	104	98	130	78

Table 54.5 Empty containers transport capacity between ports (TEU)

Port	D_1	D_2	D_3	D_4	D_5
S_1	40	130	60	80	40
S_2	90	80	70	50	40
S_3	70	45	120	40	110
S_4	40	50	50	60	50

Table 54.6 Transport time from i to j (day)

t_{ij}	D_1	D_2	D_3	D_4	D_5
S_1	0	0	2	4	3
S_2	0	3	4	2	0
S_3	3	4	3	1	1
S_4	2	1	3	4	4

54.3.3 Results and Analysis

Integer programming is realized by LINDO software, which inputs the linear constraints mentioned above in the LINDO software interface, and then, the software run the program to get the optimal results.

The optimal results of LINDO software operated after consolidation (Tables 54.7 and 54.8).

Table 54.7 Allocation volumes between ports

Time ($t = k$)	Allocation volumes (the number in brackets represent allocation volumes)
1	$S_1 \rightarrow D_1(0)$; $S_1 \rightarrow D_2(10)$; $S_2 \rightarrow D_5(20)$; $S_2 \rightarrow D_1(40)$; $S_4 \rightarrow D_4(30)$
2	$S_1 \rightarrow D_1(30)$; $S_2 \rightarrow D_5(20)$; $S_2 \rightarrow D_1(60)$; $S_1 \rightarrow D_2(80)$; $S_4 \rightarrow D_4(60)$; $S_3 \rightarrow D_4(15)$; $S_3 \rightarrow D_5(75)$; $S_4 \rightarrow D_2(30)$
3	$S_1 \rightarrow D_1(12)$; $S_2 \rightarrow D_5(30)$; $S_2 \rightarrow D_1(46)$; $S_1 \rightarrow D_2(35)$; $S_4 \rightarrow D_4(52)$; $S_3 \rightarrow D_4(16)$; $S_3 \rightarrow D_5(65)$; $S_4 \rightarrow D_2(8)$; $S_1 \rightarrow D_3(2)$; $S_2 \rightarrow D_4(28)$; $S_4 \rightarrow D_1(35)$
4	$S_1 \rightarrow D_1(20)$; $S_2 \rightarrow D_5(33)$; $S_2 \rightarrow D_1(80)$; $S_1 \rightarrow D_2(48)$; $S_4 \rightarrow D_4(50)$; $S_3 \rightarrow D_4(18)$; $S_3 \rightarrow D_5(68)$; $S_4 \rightarrow D_2(25)$; $S_1 \rightarrow D_3(28)$; $S_2 \rightarrow D_4(34)$; $S_4 \rightarrow D_1(3)$; $S_1 \rightarrow D_5(2)$; $S_2 \rightarrow D_2(70)$; $S_3 \rightarrow D_1(40)$; $S_3 \rightarrow D_3(81)$; $S_4 \rightarrow D_3(23)$
5	$S_1 \rightarrow D_1(15)$; $S_2 \rightarrow D_5(28)$; $S_2 \rightarrow D_1(77)$; $S_1 \rightarrow D_2(56)$; $S_4 \rightarrow D_4(21)$; $S_3 \rightarrow D_4(22)$; $S_3 \rightarrow D_5(81)$; $S_4 \rightarrow D_2(44)$; $S_1 \rightarrow D_3(38)$; $S_2 \rightarrow D_4(43)$; $S_4 \rightarrow D_1(25)$; $S_1 \rightarrow D_5(30)$; $S_2 \rightarrow D_2(70)$; $S_3 \rightarrow D_1(45)$; $S_3 \rightarrow D_3(65)$; $S_4 \rightarrow D_3(44)$; $S_3 \rightarrow D_2(40)$; $S_4 \rightarrow D_4(35)$

Table 54.8 Leasing container volumes between ports

	D_1	D_2	D_3	D_4	D_5
$k = 1$	250	407	239	391	346
$k = 2$	15	91	185	647	517
$k = 3$	219	8	248	54	207
$k = 4$	227	663	69	54	176
$k = 5$	278	413	0	260	0

54.4 Conclusion

This paper tries to approach the actual circumstance; in this model, we have fully considered transportation costs, loading and unloading costs, leasing costs, and transportation time. Integer programming model of empty container allocation was established under the premise of meeting heavy container transportation. Combining concrete computation and using LINDO software to solve the integer programming model, we can arrive at a conclusion about optimal allocation and leasing scheme. Although this paper tries to approach the actual circumstance when establishing the model, it still exists some disparities with reality. Many factors in reality need to be considered.

References

1. Cheung RK, Chen CY (1998) Two-stage stochastic network model and solution methods for the dynamic empty container allocation problem. Transp Sci 32(2):142–162
2. Choong ST, Cole MH, Kutanoglu E (2002) Empty container management for intermodal transportation networks. Transp Res Part E 38:423–438

3. Li JA, Liu K, Leung SCH, Lai KK (2004) Empty container management in a port with long-run average criterion. Math Comput Model 40(1–2):85–100
4. Lam SW, Lee LH, Tang LC (2007) An approximate dynamic programming approach for the empty container allocation problem. Transp Res Part C 15(4):265–277
5. Meng Q, Wang SA (2011) Liner shipping service network design with empty container repositioning. Transp Res Part E 47:695–708
6. Cheang B, Lim A (2005) A network flow based method for the distribution of empty containers. Int J Comput Appl Technol 22(4):198–204
7. Chang SC (1999) Fuzzy production inventory for fuzzy product quantity with triangular fuzzy number. Fuzzy Sets Syst 107:37–57
8. Chen SH, Hsieh CH (1999) Optimization of fuzzy backorder inventories models. In: Proceedings of 1999 IEEE international conference on systems, man and cybernetics, SMC'99, Tokyo, Japan, pp 425–431

Chapter 55
An IP Geolocation Algorithm Using Bandwidth Measurement and Distance Estimation

Minzheng Jia, Yuanzhong Zhu and Hongyu Yang

Abstract IP geolocation technology can locate the geographic location of users and equipments in the network under the condition of only knowing the IP address of the users or the devices. This paper describes the IP geolocation technology and then proposes an IP geolocation algorithm which locates the target IP address according to the relationship of paths by measuring bandwidth and determining path distance. Some experiments are carried out using this method. The results show that this method can geolocate IP address more accurately.

Keywords Geolocation · Bandwidth measurement · IP address · Path

55.1 Introduction

According to the reports in the Internet Live Stats, the number of Internet users in the world is more than 3 billion and the number of sites is more than 1 billion. Internet is developing from breadth to depth, while network application is changing people's life from the point to the whole. The penetration of Internet into Internet users' life is increasing. With the rapid development of Internet, network application has proposed new requirements, one of which is to distinguish the network users geographically. The goal of IP geolocation technology is to locate the geographic location of users and equipments in the network under the condition of only knowing the IP address of the users or the devices [1].

IP geolocation technology can provide geographic information, so it has great application value in terms of services associated with geography, such as targeted advertising, network security, network optimization, and instant messaging.

M. Jia (✉) · Y. Zhu · H. Yang
Beijing Polytechnic College, No. 368#, Shimen Road, Shijingshan District, Beijing, People's Republic of China
e-mail: jmz@bgy.org.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_55

55.2 Related Work of IP Geolocation

With the appearance of network entity geographic location technology in recent years, IP geolocation technology aims to give a relatively accurate estimate only with IP address. According to different technology and methods adopted, IP geolocation technology can be divided into two categories: Geographic positioning method based on information query and geographic positioning method based on measurement.

- Geographic positioning method based on information query

Geographic positioning method based on information query accesses to domain name, host name of IP address, organization name of registered address block, street address, and so on. But for a large ISP or geographically dispersed organization, the registered address is typically different from the real location of its host [2]. Because it is dependent on registration information deeply, geographic positioning method based on information query can only access a part of information of IP address. Moreover, when the registration information is not updated in time, the error may be relatively large.

- Geographic positioning method based on measurement

Geographic positioning method based on measurement estimates the host location by measuring delay from target host to the point of measurement and converting it into distance. In order to improve positioning accuracy, people usually combine network topology information to locate [3]. The typical algorithms are GeoPing, CBG, TBG, and Spotter.

GeoPing [4] algorithm is based on the assumption that the round-trip delay from hosts of similar geographic location to other hosts is similar. The algorithm divides all the nodes in the network into three types: measuring nodes, target nodes, and datum nodes.

CBG [5] is a geographic location algorithm based on the constraint. The algorithm assumes that it is a linear relationship between a network delay and the geographic distance. To locate the target host, Ping is used to measure the round-trip delay from the target host to each measurement reference point and transform the round-trip delay to geographic distance. In order to locate the target host, the positioning method in sensor networks is used and the intersection of geographic range is calculated. But the measurement accuracy should be further improved in actual measurement.

TBG [6] is a geographic location based on network topology. The algorithm locates routing nodes and target nodes through converting the network topology and delay data into a set of constraints. TBG treats the near-node position on the network path as the location of target node. And TBG based on the link delay estimation and accurate topology discovery discovers topology of the network by using Traceroute. The measurement cost of the algorithm is higher than the cost of CBG and GeoPing.

Different from the above algorithms, the Spotter [2] algorithm assumes that delay and distance follow the law of uniform distribution for each measurement point. The location of host can be estimated by normal distribution parameter estimation method on the basis of delay and distance following the standard normal distribution. This method is a model based on the probability, and so its accuracy is mainly influenced by the probability model.

Moreover, there is also geographic positioning algorithm by using a variety of techniques together. Its accuracy is improved, while the cost of measurement increases.

55.3 A New IP Geolocation Algorithm

In the network, the data packets reach the destination from the source hop by hop through network path via route forwarding. If the network paths to destination and intermediate nodes in the path are clear, the geographic location of destination can be estimated by referring network distance. On the basis of this idea, an IP geolocation algorithm named as GBMDE is put forward.

55.3.1 Model

In network, time delay (t) of a path can be divided into four parts: queuing delay (T_q), transmission delay (T_t), processing delay (T_p), and propagation delay (T_g), which can be expressed as the following formula:

$$t = T_q + T_t + T_p + T_g \tag{55.1}$$

The queuing delay is affected by the background flow of the network. If the network is idle, there is no interference flow, and queuing delay is zero. The transmission delay is determined by the size of the transmitted packet (s) and the bandwidth (b), which is equal to the ratio of the two ($T_q = s/b$). If the size of the packet is fixed, the transmission delay will also be fixed. The processing delay is affected by the performance of equipment, but the effect is generally very small and can be neglected. The propagation delay is related to the transmission medium, and it is proportional to the propagation distance of the signal.

For the single hop path, if the packet is sent in fixed size without other flow, the time delay (t) can be expressed as Eq. (55.2).

$$t = \frac{l}{k} + g \tag{55.2}$$

l is the propagation distance, g is the sum of T_q, T_t, and T_p, which is fixed in this condition. k is the transmission rate of electromagnetic wave signal on transmission medium, which is 2.3 * 10^5 km/s in copper cables and 2 * 10^5 km/s in optical fibers. According to the assumption, there is a linear relationship between the delay and the geographic distance in the model.

55.3.2 GBMDE Algorithm

To locate the IP address of destination such as D, we take the following steps.

Step 1: we select a number of known nodes that are widely distributed over a geographic range as the measurement observation points. Then, we select one of the measurement observation points as the source to measure the path from the observation point to the destination D and record the round-trip delay of route for each hop and node set K. The same size data packets are sent to measure delay for many times in order to reduce the impact of queuing delay. The round-trip delay of the ith hop is taken as $\mathrm{RTT}_i = \min(\mathrm{RTT}_i^1, \ldots, \mathrm{RTT}_i^n)$. In the formula, n is the number of measurements.

Step 2: in the node set K, we can use several methods to determine the location of node A which is the last hop to the target D in the path. If the observation point reaches the destination address D after j hops, the delay time from the node A to D is $\Delta t = (\mathrm{RTT}_j - \mathrm{RTT}_{j-1})/2$.

Step 3: if the processing delay is neglected, then $g = T_q = s/b$. It is known from formula (55.2), $\Delta t = \frac{d_a}{k} + \frac{S}{b}$. d_a is the distance from node A to node D. The equation is a linear regression equation, in which Δt and s are two variables, and the rest are constants.

Step 4: using data packets of different sizes, and repeating the above steps 1–3, we can get a series of delay data Δt_1, Δt_2… from node A to node D. The size of the bandwidth b can be obtained according to a linear regression prediction and then the distance d_a from node A to node D can be determined. At the same time, the area of the circle $R(a)$ is located where A is the center and d_a is the radius.

Step 5: selecting another observation point, and repeating the above steps 1–4, we can also get the node B. And then, we determine the distance d_b from node B to node D and the area of the circle $R(b)$ where B is the center and d_b is the radius. Similar to node B, we can also get the node C. At the same time, the distance d_c from node C to node D and the area of the circle $R(c)$ where C is the center and d_c is the radius can also be obtained.

In order to geolocate D, the relationship of node A, node B, and node C should be discussed. There are three cases of node A, node B, and node C.

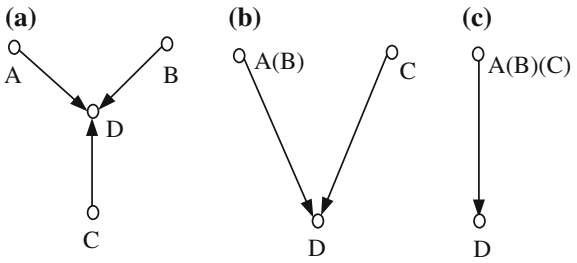


Fig. 55.1 The relationship of node A, node B, and node C

Case 1: if node A, node B, and node C are in different geographic locations, as shown in Fig. 55.1a and if $R(a) \cap R(b) \cap R(c) \neq \emptyset$, the location of the target IP address is $R(a) \cap R(b) \cap R(c)$. If there exists a crosspoint among the three nodes and if the distribution geographic coordinates of A, B, and C are (x_a, y_a), (x_b, y_b), and (x_c, y_c), we can get the coordinates (x, y) of D according to trilateration algorithms [7], as shown in Eq. (55.3)

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2(x_a - x_c) & 2(y_a - y_c) \\ 2(x_b - x_c) & 2(y_b - y) \end{bmatrix}^{-1} \begin{bmatrix} x_a^2 - x_c^2 + y_a^2 - y_a^2 + d_c^2 - d_a^2 \\ x_a^2 - x_c^2 + y_b^2 - y_c^2 + d_c^2 - d_b^2 \end{bmatrix} \tag{55.3}$$

If $R(a) \cap R(b) \cap R(c) = \emptyset$, we amplify $R(a)$, $R(b)$, and $R(c)$ with equal proportion until the crossregion occurs. Then, the above method can be used.

Case 2: if two nodes among A, B, and C superpose each other, only two nodes near D can be measured geographically. Suppose node A and node B superpose each other and $R(a) \cap R(c) \neq \emptyset$; then, the location of the target IP address is $R(a) \cap R(c)$. If there exits only one crosspoint between the above two nodes, as shown in Fig. 55.1b, distribution geographic coordinates of A and C are (x_a, y_a) and (x_c, y_c). In this case, we need to measure the distance d_{ac} between A and C according to the map. The mathematical model of the case is shown in Fig. 55.2. In order to get the coordinates of D, the values of t and h are firstly calculated, as shown in Eqs. (55.4), (55.5), and (55.6).

$$h = \frac{2\sqrt{p(p - d_{ac})(p - d_a)(p - d_c)}}{d_{ac}} \tag{55.4}$$

$$p = \frac{d_{ac} + d_a + d_c}{2} \tag{55.5}$$

$$t = \pm\sqrt{d_a^2 - h^2} \tag{55.6}$$

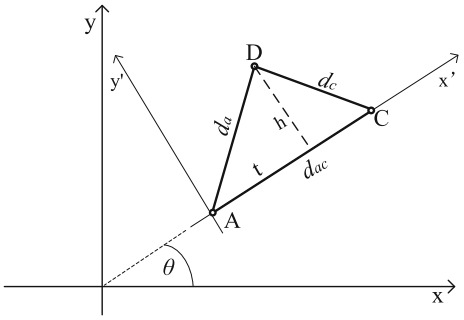


Fig. 55.2 Two-nodes model

A new coordinate system (x', y') with A as a new coordinate origin and AC as a new X axis is established. The corresponding relationship between the new coordinate system and the original coordinate system [8] can be obtained in Eqs. (55.7) and (55.8).

$$\begin{cases} x = x' \cos\theta - y' \sin\theta + x_a \\ y = x' \sin\theta + y' \cos\theta + y_a \end{cases} \tag{55.7}$$

$$\begin{cases} \sin\theta = \dfrac{y_c - y_a}{d_{ac}} \\ \cos\theta = \dfrac{x_c - x_a}{d_{ac}} \end{cases} \tag{55.8}$$

θ is the angle between the new coordinate system x' axis and the x axis of the original coordinate system.

In the new coordinate system (x', y') coordinates of D point are (t, h). The coordinates of D can be described in Eq. (55.9).

$$\begin{cases} x = t\dfrac{x_c - x_a}{d_{ac}} - h\dfrac{y_c - y_a}{d_{ac}} + x_a \\ y = t\dfrac{y_c - y_a}{d_{ac}} + h\dfrac{x_c - x_a}{d_{ac}} + y_a \end{cases} \tag{55.9}$$

If $R(a) \cap R(c) = \emptyset$,, we amplify $R(a)$ and $R(c)$ with equal proportion until the crossregion occurs. Then, the above method can be used.

Case 3: if node A, node B, and node C superpose each other, different measurement observation points in geography reach the destination node D through the same path as shown in Fig. 55.1c. This situation occurs in the case that the destination node D is the node of network terminal. The majority of users access network by China mobile, China telecom, and China Unicom in China and other operators. Node A is often the route node which urban operators access, so we use node A to represent the location of the target node.

In addition, in real network environment, IP address is usually distributed by using address prefix as scale. And usually a period of continuous IP address is also adjacent [9] in geography. By comparing the target IP address and the IP address with known location in the area of $R(a) \cap R(b) \cap R(c)$, we can further determine the position of the target IP, which can reduce the error and improve the positioning accuracy.

55.4 Experiments and Results

In order to verify the correctness of the proposed locating method, we prepare experimental tool with Wincpap in this paper to make delay measurement result accurate to the microsecond level, which improves the measurement accuracy, compared with Traceroute tool. More than 30 cities of China are selected as observation points of measurement for experimental verification. On each observation point, the IP address of target nodes is measured and recorded, and the collected data are finally calculated by locating server. The measurement system consists of measurement observation points, target nodes, and locating servers, as shown in Fig. 55.3.

In order to compare the locating results, we choose an IP database as a location database and select 1000 records as the IP address to locate and remove them from the database.

Locating error can be defined as the distance between the results from the measurement location algorithm and IP location from the IP database. Based on

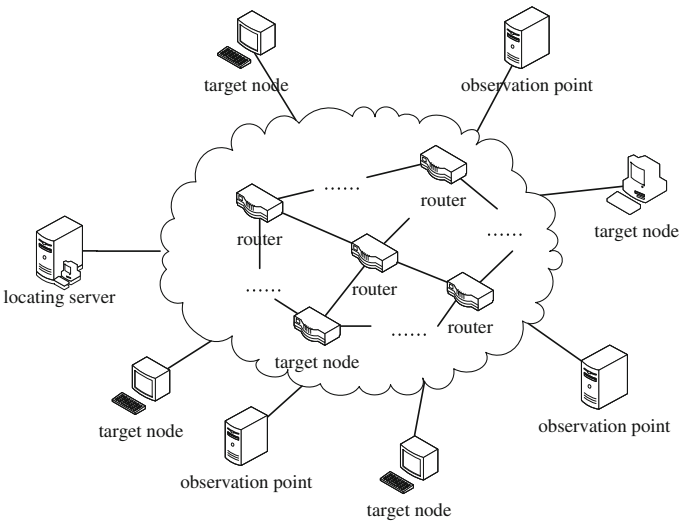


Fig. 55.3 Measurement system

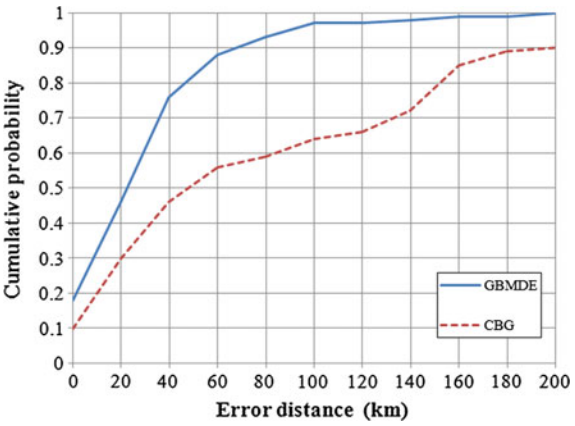


Fig. 55.4 Experimental results

repeatedly measuring method and the comparison with the above-mentioned CBG algorithm, the experiment has been carried out. Experiment results are shown in Fig. 55.4.

Measurement results show that compared with CBG algorithm, the GBMDE algorithm can locate IP address more accurately with positioning error median within 30 or 100 km and above. Through the analysis of these data, we find that the cause for larger error is that ISPs always change their IP from time to time and that IP database data cannot be real-time-updated. For further verification, more than 20 IP addresses with exact location are measured. And the target IP address can be located more accurately with the GBMDE algorithm, which verifies its validity.

55.5 Conclusion

With the appearance of new network applications, many applications need to know the geographic location of the network entity. This paper studies the existing IP geolocation technology and introduces geographic positioning method based on information query and geographic positioning method based on measurement. Also a measurement model is established. On the basis of this model and geographic positioning method for information query, GBMDE algorithm is proposed in this paper. Compared with CBG, GBMDE algorithm improves the accuracy of network node location estimation. Because the selection of measurement observation points is an important factor in IP geolocation, the future work is to study how to improve precision and accuracy of location by selecting more reasonable measurement observation points.

Acknowledgments This work is supported by Beijing Higher Education Young Elite Teacher Project (YETP1778) and Beijing Polytechnic College Scientific Research Task (bgzykyz201405).

References

1. Wei L (2012) Research and implementation of the IP geolocation technology. Master thesis, Beijing Jiaotong University (in Chinese)
2. Laki S, Mátray P, Hága P, Sebök T, Csabai I, Vattay G (2011) Spotter: a model based active geolocation service. In: Proceedings of the INFOCOM. IEEE Press, Shanghai, pp 3173–3181
3. Zhangfeng W, Jing F, Changyou X et al (2014) Research on the IP geolocation technology. J Softw 25(7):1527–1540 (in Chinese)
4. Padmanabhan VN, Subramanian L (2001) An investigation of geographic mapping techniques for Internet hosts. In: Proceeding of the ACM SIGCOMM. ACM Press, San Diego, pp 173–185
5. Gueye B, Ziviani A, Crovella M, Fdida S (2006) Constraint-based geolocation of internet hosts. ACM/IEEE Trans Netw 14(6):1219–1232
6. Katz-Bassett E, John J, Krishnamurthy A, Wetherall D, Anderson T, Chawathe Y (2006) Towards IP geolocation using delay and topology measurements. In: Proceedings of the IMC 2006. ACM Press, Rio de Janeiro, pp 71–84
7. Yanyan H (2014) Study on DV-HOP localization algorithm for wireless sensor networks. Master thesis, Southwest Jiaotong University (in Chinese)
8. Xiangyuan K (2010) Base of geodetic science. Wuhan University Press, Wuhan (in Chinese)
9. Yan L, Feng X, Junyong L (2007) Research on geographic mapping for Internet entities. Comput Eng 09:37–39 (in Chinese)

Chapter 56
Line-Based Traction Energy Consumption Estimation of Urban Rail Transit

Ying Tang, Enjian Yao, Rui Zhang and Xun Sun

Abstract With the rapid development of urban rail system, the associated energy consumption increases progressively. Therefore, it is significant to predict the energy consumption resulted from the rail system. This paper aims to establish a traction energy consumption (TEC) estimation model by considering the level of service (LOS) factors related to the energy consumption (e.g., ambient temperature and load factors). First, the TEC model of a special line is established after analyzing the influencing factors of TEC; then, based on the data collected in Beijing urban rail system, the model for a special line is estimated and examined; finally, three different TEC estimation methods, i.e., unit indicator method, time series model, and regression model, are introduced and compared with the proposed model. The results show that the proposed TEC model considering the LOS has higher precision, with MAPE of 1 %. Moreover, when the line operation length, etc., do not change in the near future, except for unit indicator method, other methods all can be used to estimate the monthly TEC.

Keywords Urban rail transit · Traction energy consumption · Estimation · Regression model

56.1 Introduction

Urban rail transit is regarded as an ideal solution to reduce the impact of urban mobility due to its high capacity, reliability, safety, and excellent environmental performance, which vigorously developed in recent years. Although the unit person-kilometer energy consumption is smaller than the conventional buses, private

Y. Tang (✉) · E. Yao · R. Zhang · X. Sun
School of Traffic and Transportation, Beijing Jiaotong University,
Shangyuancun, 3, Haidian District, Beijing 100044, People's Republic of China
e-mail: 14120884@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_56

cars, and other transport modes, there is a large amount of energy consumption in rail transit system with the expanding scale of transit network. Beijing Subway 2011–2013 annual energy consumption data show that the average annual energy consumption of rail transit is up to 10 billion kw h, where the traction energy consumption (TEC) shares 53–56 %, so it is necessary to estimate the TEC of rail transit.

At present, the study of urban rail transit TEC is mainly concentrated in three aspects: first, to explore the mathematical model and evaluation methods of energy consumption at system level [1, 2]; second, to get the calculation method or the simulation of TEC at technical level [3, 4]; and third, to use the unit indicator method, regression method, and other methods to build the estimation models for TEC based on statistical analysis of historical data [5–7]. However, the establishment of some models requires detailed parameters of lines and trains, and the process of modeling is complex. In contrast, the structures of other models are relatively simple, but they do not consider the relationship between level of service (LOS) and TEC.

This study analyzes the LOS indictors related to energy consumption and builds the TEC model with the consideration of the selected LOS indictors (i.e., ambient temperature, load rate) and passenger turnover volume; what's more, different estimation methods about TEC are compared to analyze the rationality of the proposed methodology.

56.2 Line-Based TEC Model

There are many factors to affect TEC of a train, including their own property (e.g., laying mode, vehicle factor, and regional factor), train operation plan (e.g., train intervals and train running time), passenger flow characteristics (e.g., passenger flow volume, passenger turnover volume, and the average transport distance), and seasonal factor [8]. Compared with the total TEC indicator, there are less factors to influence the person-kilometer TEC; therefore, in order to simplify the model, considering the relevance of the person-kilometer TEC and ambient temperature, load rate, and other LOS indicators, the regression model of the person-kilometer TEC is established, and then TEC value is calculated based on person-kilometer TEC and passenger turnover volume.

56.2.1 Model Construction

Taking into account the relevance of urban rail transit TEC, available degree of index data and fixity of a special line's technical equipment (e.g., train capacity, platform screen door set, and train speed), the ambient temperature, load rate, and train interval are selected as LOS factors related to the TEC of a special line.

Table 56.1 Correlation analyses between energy consumption and LOS indicator

Variable		Load rate (%)	UAT (°C)	AVOTD (°C)	Train interval (min)
Person-kilometer TEC (kw h/(10^4 person - km))	Pearson correlation	−0.312*	0.611**	−0.455**	−0.217
	Significance (bilateral)	0.050	0.000	0.005	0.185
	N	36	36	36	36

*Significant correlation at the 0.05 level (bilateral)
**Significant correlation at the 0.01 level (bilateral)

Table 56.1 is the correlation analysis between person-kilometer TEC and outside ambient temperature (UAT), the absolute value of the temperature difference (AVOTD), load rate, train interval.

As shown in Table 56.1, there is a certain correlation between person-kilometer TEC and load rate, UAT, AVOTD, train interval, but the first three factors have significant impact on person-kilometer TEC.

Because of person-kilometer TEC's seasonal, cyclical characteristic, the weak linear correlation between UAT and AVOTD, the UAT, AVOTD, and load rate are finally selected as explanatory variables of person-kilometer TEC to build the following multivariate linear regression model

$$y^i = \beta_0 + \beta_1 x_1^i + \beta_2 x_2^i + \beta_3 x_3^i + \varepsilon^i, \tag{56.1}$$

where y is the person-kilometer TEC, kw.h/(10^4km); x_1 and x_2 are UAT and AVOTD, °C, respectively ; x_3 is the train load rate; β_0 is regression constant; β_1, β_2, and β_3 are the regression coefficients of x_1, $x_{2,\text{ and }}$ x_3, respectively; ε is a random error term; i is the month, in which AVOTD is the temperature difference between UAT and train internal temperature as follows:

$$x_2^i = \left| x_1^i - x_0^i \right|. \tag{56.2}$$

Considering the air-conditioning cooling in summer and heating in winter, and combining with the control scheme of train air-conditioning, the air-conditioning temperature should be set to ensure the comfort perceived by the human, and also to take into account human health; therefore, the difference between interior temperature and the outside temperature is not too large. This paper references to the air-conditioning temperature control curve of metro vehicle proposed by Zhimei et al. [9], and combined with temperature changes of the particular season in Beijing, the train interior temperature of summer (June–September) was set to 27°; the interior temperature of winter (November to March) is set to 19°; the interior temperature of other months (4, 5, and 10 month) is calculated according to the

Table 56.2 Model parameters of line 1

Parameter	Coefficient	T	VIF
Constant	426.492	16.174	
UAT (°C)	3.947	9.831	6.011
AVOTD (°C)	1.874	3.662	6.126
Load rate (%)	−7.203	10.200	1.278

actual temperature (where the interior temperature is equal to the outside ambient temperature). Therefore, the interior temperature of different seasons is:

$$x_0^i = \begin{cases} 19 & i = 1, 2, 3, 11, 12 \\ 27 & i = 6, 7, 8, 9 \\ x_1^i & i = 4, 5, 10 \end{cases} \tag{56.3}$$

Then, according to the relationship among TEC, person-kilometer TEC, and passenger turnover volume, the TEC of a special line based on LOS can be calculated by the following equation:

$$E^i = y^i \times r^i, \tag{56.4}$$

where E is TEC, kw h; r is passenger turnover volume, 10^4 person–km. Taking the actual data in 2011–2013 of Beijing Subway line 1 as an example, we established person-kilometer TEC model and use multiple linear regression to solve the regression model parameters. The results are shown in Table 56.2.

As shown in Table 56.2, T absolute value of all parameters is greater than 1.96, indicating that the UAT, AVODT, and load rate of the model, respectively, have significant impact on person-kilometer TEC at 95 % confidence level. The variance inflation factor (VIF) of three variables was less than 7, attesting the three independent variables' collinearity is weak. In addition, the statistical value, $F = 81.999 > 2.92 = F_{0.05}(3, 32)$, states clearly the linear relationship between dependent variables and independent variables is overall significantly substantiated at 95 % confidence level; Durbin-Watson test value DW = 1.745 is in the vicinity of 2, signifying that there is no serial correlation of random error term. Adjusted R-squared is 0.874, which shows that the model has good fitting effect; therefore, the multiple linear regression model is valid. Moreover, according to the sign of the parameter, it can be seen that the smaller the load factor is, the larger UAT and AVOTD is, the greater the person-kilometer TEC is.

56.2.2 Model Verification

Mainly based on the actual operation data of Beijing Subway line 1, we verify the line-based TEC models' estimation ability of the proposed method. By comparing the estimated and actual TEC values of line 1 in 2014 January–August, the mean

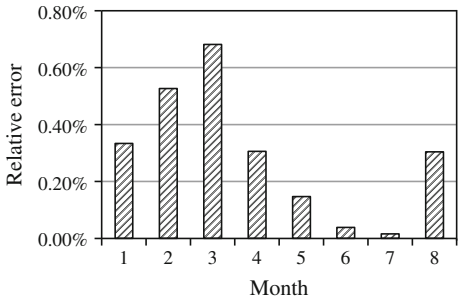


Fig. 56.1 Error analysis of TEC

absolute percentage error (MAPE) of the estimated and actual value is calculated to test accuracy.

$$\text{MAPE} = \frac{1}{n}\sum_{h=1}^{n}\frac{\left|E'_h - E_h\right|}{E_h} \times 100\ \%, \tag{56.5}$$

where E'_h is the estimated TEC value, E_h is the actual TEC value, and h is the number of estimated months.

The established TEC model on the above is calculated and analyzed. After calculation, we get the total TEC estimated value of Beijing Subway line 1 in 2014 January–August which is 65,500,382 kw h and the MAPE is 0.29 %, which reveals that the model has a very high accuracy. Figure 56.1 shows the TEC relative error of each month in 2014 January–August. As shown in Fig. 56.1, the maximum TEC relative error is in March, and the minimum relative error is in July. However, the relative error from January to August is all within 0.7 %, indicating this model has good estimation effect.

56.3 TEC Estimation Methods' Comparison

There are many estimation methods about TEC, and the following will compare several commonly used methods with this established method on the above.

56.3.1 Unit Indicator Method

Taking train ton-kilometer TEC as the unit indicator, the line-based TEC is calculated combined with the basic parameters and operation data of the subway line, and the ton-kilometer TEC is defined as:

$$\eta_{jk} = E_{jk}/\left(2n_j W_j l_j + mr_j\right), \quad k = 1, 2, \tag{56.6}$$

where η_{jk} is ton-kilometer TEC; kw h.(t km)$^{-1}$; E_{jk} is the TEC statistical value of the underground lines or surface lines; n_j is yearly trains pairs of line j; W_j is the train weight, t; l_j is the operation length, km; m is the average body weight of passengers, t (taking the value of 0.06 t); r_j is the passenger turnover volume, person–km.

After listing systematically the TEC data of Beijing Subway underground lines, the ton-kilometer TEC of underground lines is calculated to 0.0528 kw h.(t km)$^{-1}$, which is very close to the 0.0521 kw h.(t km)$^{-1}$ calculated by Zijia et al. [5], verifying the accuracy of the unit indicator method. Combined with the operation data and basic parameters of Beijing Subway line, we get the total TEC of line 1 in 2014 January–August which is 61,525,430 kw h, with the MAPE of 12 %.

56.3.2 Time Series Model

By using the SPSS software, we ultimately screen the optimal TEC model of Beijing Subway line 1 which is ARIMA (1,0,0) (1,1,0), where the model parameters $|T| = (2.056, 4.916, 3.55) > 1.96$, Ljung-Box Q statistical value $p = 0.704 > 0.05$, show that at the 95 % confidence level, the model has pass through the T test and the residual white noise test, and in addition, R-squared is 0.926, showing a high fitting degree of this model. Therefore, we use this model to estimate the TEC of Line 1 in 2014 January–August which is 68,257,787 kw h, with the MAPE of 6 %.

56.3.3 Other Regression Model

Zhen-ming et al. [7] considered two variables, the monthly average temperature (outside ambient temperature) and the total mileage, to establish regression model of TEC. Under such a model, an efficient TEC regression equation of line 1 can be set up via SPSS software, and we have

$$E = -2657996 + 80204x_1 + 24263x_4, \tag{56.7}$$

where x_4 is the total mileage, 10^4 vehicle-kilometers. Thus, according to the operation data of average temperature and the total mileage, the TEC of a special line can be obtained. The total estimated TEC value of Beijing Subway Line 1 in 2014 January–August is 65,831,879 kw h, with the MAPE of 5 %.

Table 56.3 Contrast of different methods

Methods	Actual value of TEC (kw h)	Estimated value of TEC (kw h)	MAPE (%)
The proposed method	64,421,825	65,500,382	0.29
Unit indicator method		62,574,158	12
Time series model		68,257,787	6
Other regression model		65,831,879	5

56.3.4 The Comparison of Various Estimation Methods

In summary, the actual and estimated values' comparison of various estimation methods is as follows.

According to Table 56.3, the MAPE of the proposed method, time series model, and other regression model is small in 6 %, while the larger of unit indicator method, describing the relative error of each month's actual and estimated TEC value using the former three methods, is less volatile, and all can be used to estimate the specific month's TEC while the unit indicator method is not suitable for the TEC estimation of specific month. Among them, the result of other regression model is the nearest with the method proposed in this paper, because both methods are established by considering several key indicators of operations. However, the unit indicator method results in large MAPE without considering the impact of seasonal factor on the ton-kilometer indicator, so it is recommended to divide the season to summer and winter, others to calibrate the ton-kilometer TEC indicator. Additionally, the time series model highlights the time series, but if there is a greater change in the external environment, the estimated error will be great; therefore, if the operation length and the number of stations TEC suddenly increase, the method is not available for pulling TEC estimation. It shows that the proposed method is more accurate and reasonable while compared with three other methods.

56.4 Conclusions

TEC is the major component of energy consumption related to urban rail system. Based on Beijing subway operation data, a TEC model incorporated in the LOS indicators is established specific to a certain line after analyzing the influencing factors of TEC. Specifically, the LOS indicators are connected with the person-kilometer TEC (e.g., ambient temperature, load rate) and passenger turnover volume. In addition, three other TEC estimation methods are introduced and compared with the proposed method. The results show that the model established in this paper has higher estimated accuracy (the MAPE is within 1 %) and more reasonable explanation when compared with other methods.

Acknowledgments This work is financially supported by National 973 Program of China (No. 2012CB725403).

References

1. Danziger NH (1975) Energy optimization for rail public transit system. Transp Res Rec 4:31–39
2. Gonzalez-Gil A, Palacin R, Batty P et al (2014) A systems approach to reduce urban rail energy consumption. Energy Convers Manag 80:509–524
3. Lijun Y, Yongsheng H, Lixia S (2005) Energy-saving track profile of urban mass transit. J Tongji Univ (Nat Sci) 5(4):20–26 (in Chinese)
4. Shuai S, Tao T (2014) Optimal train control for ATO system. J China Railway Soc 36(12):50–55 (in Chinese)
5. Zijia W, Feng C, Zhongheng S (2013) Prediction on medium and long term energy consumption of urban rail transit network in Beijing. China Railway Sci 34(3):133–136 (in Chinese)
6. Feng C, Yang Y, Ouyang L (2014) Estimation methods for traction energy consumption of urban rail transit. Urban Rapid Rail Trans 27(2):90–93 (in Chinese)
7. Zhenming Y, Yue Ji-guang Y, Xiaobao W et al (2010) Prediction of urban rail transit power consumption based on regression model. Urban Mass Trans 12:22–25 (in Chinese)
8. Jingya S, Yongqing S, Jiguang Y (2008) Analysis on factors influencing energy consumption in urban rail transit and the establishment of energy consumption evaluation system. Railway Transp Econ 30(9):46–49 (in Chinese)
9. Zhimei L, Ke Y, Tianzhi Y, Fengmin X (2013) The research of temperature control in metro vehicle air conditioning. Technol Market 20(5):29–30 (in Chinese)

Chapter 57
Reliability Analysis of Snow Depth Monitoring System Based on Improved FMECA

Jing Yang, Zongyi Xing, Jingbao Ren and Yong Qin

Abstract Aiming at the problems of low accuracy and strong subjectivity in reliability analysis based on FMECA, the fuzzy theory and VIKOR theory methods are applied in criticality analysis in FMECA technique. Firstly, fuzzy decision matrix of risk factors based on fuzzy theory is built, in which fuzziness is removed by barycenter method, and specific evaluation value of risk factors is obtained. Secondly, several specific parameters of evaluation value of risk factors are calculated based on VIKOR theory. Finally, the compromise solution of all failure modes is found out, and the weak links are determined by VIKOR method. The proposed modified FMECA method is conducted for the reliability analysis of snow depth monitoring system, and the results are compared with field data. The results show the efficiency of the proposed method in reliability analysis of snow depth monitoring system.

Keywords Reliability · Failure mode effects and criticality analysis (FMECA) · VIKOR · Fuzzy theory

57.1 Introduction

Highly reliable pre-disaster system for high-speed railway is necessary for preventing operation risk posed by natural environment. As an important part of pre-disaster system for high-speed railway, snow depth monitoring system and its reliability are of great significance to the safe operation of high-speed trains. FMECA, a well-developed method of reliability analysis, is widely applied in the

J. Yang · Z. Xing (✉) · J. Ren
School of Mechanical Engineering, Nangjing University of Science and Technology, Nanjing 210094, China
e-mail: xingzongyi@163.com; nust_yj6033@163.com

Y. Qin
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiao Tong University, Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_57

reliability analysis of various systems. Qiang [1] analyzes the failure mode of freight car bogies and calculates respective criticality, which offers meaningful reference for the maintenance decision making of bogies; a reliability framework, based on FMECA, proposed by Hanliang [2] is used to analyze the reliability of diesel engine.

All failure modes of snow depth monitoring system and respective effects on system can be obtained qualitatively by adopting FMECA method to conduct reliability analysis of snow depth monitoring system. However, objective results are hard to calculate, and weak links are hard to find correctly as a result of large-span RPN and strong subjectivity [3] when calculating respective RPN of each failure mode quantitatively; therefore, it is impossible to offer good technical guidance for routine maintenance of snow depth monitoring system.

To this issue, fuzzy theory and VIKOR theory are used to improve criticality analysis (CA) of FMECA in this paper. First, based on fuzzy theory, fuzzy decision matrix is constructed and gravity method is applied to remove fuzziness of fuzzy decision matrix. Second, several specific parameters of risk factors and criticality value of failure modes are calculated. Finally, the compromise solution of all failure modes is found, and weak links of snow depth monitoring system are determined by applying VIKOR method.

57.2 Snow Depth Monitoring System

Snow depth monitoring system, a part of pre-disaster system for high-speed railway, is responsible for collecting, processing, and analyzing snowfall data, showing snow depth value on maintenance terminal in real time and giving alarm when snow depth value is beyond the alert level, which assists dispatchers in traffic control and workers in track maintenance.

Snow depth monitoring system is composed of laser snow depth gauge meter, anti-thunder module, data-collecting unit, data transmission unit, and railway monitoring center, as shown in Fig. 57.1. The data gathered are transmitted by trackside cable to railway monitoring center where data are analyzed and alarm signal is emitted.

57.3 Theory of Improved FMECA

In the process of traditional FMECA, RPN value, by which each of essential components can be ranked and CA can be conducted, can be calculated according to the O, S, and D values of essential components after conducting failure mode and effect analysis (FMEA) on the system [4].

To overcome the disadvantages of large-span RPN and strong subjectivity, in the process of improved FMECA, firstly, fuzzy theory [5] is introduced to view three

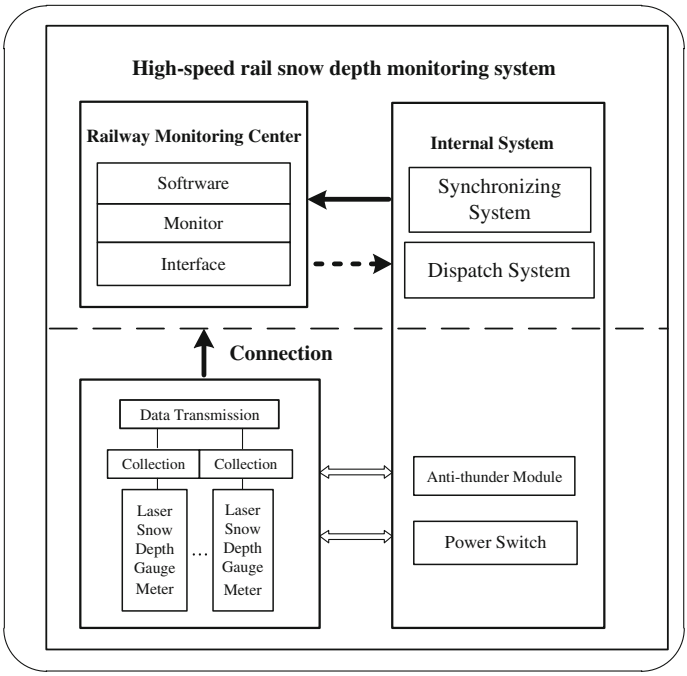


Fig. 57.1 Structure diagram of snow depth monitoring system

risk factors O, S, and D as fuzzy variable, fuzzy language is used to evaluate them, and fuzzy decision matrix of risk factor is constructed and its fuzziness is removed [6]. Then VIKOR method is applied to determine the positive ideal solution, negative ideal solution, optimal solution, and worst solution of failure mode.

57.3.1 Building Fuzzy Matrix

Suppose expert group contains K members, denoted as $\mathrm{TM}_1, \mathrm{TM}_2, \ldots, \mathrm{TM}_k$, the quantity of failure mode to be analyzed is m, denoted as $\mathrm{FM}_k(i = 1, 2, \ldots, m)$, and the quantity of risk factor to be considered is n, denoted as $C_j(j = 1, 2, \ldots, n)$. The fuzzy evaluation from experts TM_k for risk factor of failure mode and its weight are denoted as $x_{ij}^k = \left(x_{ij1}^k, x_{ij2}^k, x_{ij3}^k, x_{ij4}^k\right), w_j = \left(w_{j1}, w_{j2}, w_{j3}, w_{j4}\right)$, where $k = 1, 2, \ldots, K, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n$. The fuzzy evaluation for risk factor and its weight of each failure mode can be obtained by the following formulas:

$$x_{ij} = \left(x_{ij1}, x_{ij2}, x_{ij3}, x_{ij4}\right) \tag{57.1}$$

$$w_j = \left(w_{j1}, w_{j2}, w_{j3}, w_{j4}\right) \tag{57.2}$$

where

$$x_{ij1} = \min x_{ij1}^k,\ x_{ij2} = \frac{1}{K}\sum_{k=1}^{K} x_{ij2}^k,\ x_{ij3} = \frac{1}{K}\sum_{k=1}^{K} x_{ij3}^k,\ x_{ij4} = \min x_{ij4}^k$$

$$w_{j1} = \min w_{j1}^k,\ w_{j2} = \frac{1}{K}\sum_{k=1}^{K} w_{j2}^k,\ w_{j3} = \frac{1}{K}\sum_{k=1}^{K} w_{j3}^k,\ w_{j4} = \min w_{j4}^k$$

Suppose the fuzzy decision matrix of risk factor obtained by K experts conducting fuzzy evaluation for risk factor is $\tilde{D}$:

$$\tilde{D} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{x}_{m1} & \tilde{x}_{m2} & \cdots & \tilde{x}_{mn} \end{bmatrix} \tag{57.3}$$

57.3.2 Removing Fuzziness of Fuzzy Decision Matrix

Decision matrix of risk factor and the specific value of its weight can be gained after removing the fuzziness of the fuzzy evaluation of risk factor and its weight by using gravity method. The formula of using gravity method to remove fuzziness is as follows:

$$\tilde{x}_0(\tilde{A}) = \frac{\int x\mu_{\tilde{A}}(x)\mathrm{d}x}{\int \mu_{\tilde{A}}(x)\mathrm{d}x} \tag{57.4}$$

where $\tilde{x}_0(\tilde{A})$ is the specific value after removing fuzziness and $\mu_{\tilde{A}}(x)$ is the membership functions of fuzzy value.

57.3.3 Determination of Positive and Negative Ideal Solution

VIKOR method [7] is first proposed by Opricovic in 1998, and it is a multi-criteria decision making (MCDM) or multi-criteria decision analysis method. Assuming that compromise is acceptable for conflict resolution, the decision maker wants a solution that is the closest to the ideal, and the alternatives are evaluated according

to all established criteria. VIKOR ranks alternatives and determines the solution named compromise that is the closest to the ideal.

Considering risk factor is cost-type index, maximum and minimum can be used to determine positive ideal solution f_j^* and negative ideal solution f_j^- according to Formula (57.5):

$$f_j^* = \min x_{ij},\ f_j^- = \max x_{ij} \tag{57.5}$$

where x_{ij} is the specific value of $\widetilde{x_{ij}}$ after removing fuzziness.

57.3.4 Calculation of Optimal Solution and Worst Solution

During failure mode criticality analysis, S_i is the optimal solution of failure mode comprehensive evaluation:

$$S_i = \sum_{j=1}^{n} \frac{w_j\left(f_j^* - x_{ij}\right)}{f_j^* - f_j^-} \tag{57.6}$$

R_i is the worst solution of failure mode comprehensive evaluation:

$$R_i = \max_j \left(\frac{w_j\left(f_j^* - x_{ij}\right)}{f_j^* - f_j^-} \right) \tag{57.7}$$

where w_j is the specific value of each risk factor after removing fuzziness.

57.3.5 Calculating Criticality Value of Failure Mode

Criticality value of failure mode, Q_i, can be obtained by Formula (57.8):

$$Q_i = v\frac{S_i - S^*}{S^- - S^*} + (1 - v)\frac{R_i - R^*}{R^- - R^*} \tag{57.8}$$

where $S^* = \min S_i,\ S^- = \max S_i,\ R^* = \min R_i,\ R^- = \max R_i,\ S^*$ is the maximum utility of failure mode criticality group, and R^* is the minimum individual pity of failure mode criticality group. Coefficient v is the weight of decision mechanism of maximum group utility [8].

57.3.6 Determining a Compromise Plan

Rank alternative plan according to three sorted lists S_i, R_i, Q_i, and the closer to the top, the better the plan is. Failure mode is ranked as $A^{(1)}, A^{(2)}, \ldots, A^{(n)}$ according to increasing Q_{i} value, and $A^{(1)}$ needs to meet the following requirements when it is the least harmful failure mode.

C1 (Acceptable Advantage): $Q(A^{(2)}) - Q(A^{(1)}) \geq \mathrm{DQ}$, where $Q(A^{(1)})$ is the top Q_i value after ranking all failure mode according to Q_i and $Q(A^{(2)})$ is the alternative with second position in the ranking list by Q_i. DQ means acceptable threshold, $\mathrm{DQ} = 1/(m-1)$.

C2 (Acceptable Stability in decision making): The alternative $A^{(1)}$ must also be the best ranked by S and/or R. This compromise solution is stable within a decision-making process, which could be the strategy of maximum group utility (when $v > 0.5$ is needed, or "by consensus" v about 0.5, or "with veto" $v < 0.5$).

If one of the conditions is not satisfied, then a set of compromise solutions is proposed, which consists of alternatives $A^{(1)}$ and $A^{(2)}$ if only the condition C2 is not satisfied, or alternatives $A^{(1)}, A^{(2)}, \ldots, A^{(M)}$ if the condition C1 is not satisfied; $A^{(M)}$ is determined by the relation $Q(A^{(M)}) - Q(A^{(1)}) < \mathrm{DQ}$ for maximum M (the positions of these alternatives are "in closeness") [9].

57.4 Improved FMECA Analysis of Snow Depth Monitoring System

Failure mode and effect analysis (FMEA) is conducted on snow depth monitoring system to determine the failure mode of number 12 as follows (Table 57.1) .

The expert team of failure mode criticality analysis on snow depth monitoring system comprise 5 experts from design, assembly, maintenance, and other different

Table 57.1 Failure modes of critical components of snow depth monitoring system

No.	Failure mode
FM1	Broken monitoring host
FM2	Inner module failure of snow depth meter
FM3	Software failure of snow depth meter
FM4	Broken data transmission module
FM5	Communication failure of snow depth meter
FM6	Data access module misfunction
FM7	Power module misfunction
FM8	Snow depth meter misfunction
FM9	Power switch module misfunction
FM10	Broken cable terminal
FM11	Monitor failing to receive data
FM12	Broken anti-thunder module

departments. The experts are given different weight and conduct fuzzy analysis on risk factors of each failure mode.

According to fuzzy membership Formulas (57.1) and (57.2), fuzzy analysis of failure mode of critical components is turned into group fuzzy analysis of risk factors of failure mode of critical components. In order to obtain specific value of fuzzy analysis, fuzziness of risk factors of failure mode of 12 main components and fuzzy analysis of their weight are removed to get their specific value, according to trapezoidal fuzzy-removing rules in fuzzy membership Formula (57.3).

After determining the specific value of risk factors of failure mode of 12 main components, based on Formula (57.5), positive and negative ideal solutions of three risk factors can be achieved as follows:

$$f_o^* = 1.218, \quad f_s^* = 5.962, \quad f_D^* = 2.962$$
$$f_o^- = 7.146, \quad f_s^- = 8.537, \quad f_D^- = 8.000$$

The optimal solution S, worst solution R, and VIKOR value can be obtained by Formulas (57.6), (57.7), and (57.8). According to the optimal solution S, the worst solution R, and VIKOR value, each failure mode is ranked, respectively, which gives three different ranking results of failure mode, as shown in Table 57.2.

The sorting results of failure mode, lined up in increasing order of Q in Table 57.2, are FM1, FM7, FM6, FM3, FM10, FM4, FM9, FM11, FM8, FM12, FM5, and FM2. Failure mode FM1 ranks first in the order of Q, which is shown in sorting results, and when $m = 12$, acceptable threshold value $\mathrm{DQ} = 0.09$, $Q(\mathrm{FM7}) - Q(\mathrm{FM1}) = 0.093 \geq \mathrm{DQ} = 0$, which satisfies C1;

Table 57.2 Three ranking results of failure modes

Failure mode	Sorting by S	Sorting by R	Sorting by Q
FM1	1	1	1
FM2	12	11	12
FM3	4	2	4
FM4	6	6	6
FM5	11	10	11
FM6	3	3	3
FM7	2	4	2
FM8	9	9	9
FM9	7	7	7
FM10	5	5	5
FM11	8	8	8
FM12	10	12	10

FM1 ranks higher than FM7 in the order of S and R, which satisfies C2. Failure modes are determined by trade-off, and the ranking of criticality of failure mode of main components is FM2 > FM5, FM3 > FM12 > FM8 > FM11, FM9 > FM4, FM10, FM7, FM6 > FM1

According to criticality ranking of failure mode of main components, among known twelve failure modes, FM2 (inner module failure of snow depth meter) has the biggest criticality; FM5 (communication failure of snow depth meter) and FM3 (software failure of snow depth meter) have the same criticality, ranking second; the decreasing order of criticality of the rest failure modes is FM12 (anti-thunder module is broken), FM8 (the function is broken), FM11 (monitor fails to receive data), and FM9 (failure of power module); the failure mode of rest components has the same criticality. Components with level 1 failure mode shall be paid more attention during routine maintenance; components with level 2 failure mode shall be maintained according to site condition.

57.5 Conclusion

Traditional FMECA is improved by applying fuzzy theory and VIKOR method, and improved FMECA can effectively solve the problems of strong subjectivity which criticality analysis fails to solve and quantify objective evaluation indicators. Improved FMECA is adopted to analyze the reliability of snow depth monitoring system, and the analysis results are compared with research results from high-speed rail site, showing that the analysis results are corresponded with actual results. A conclusion can be drawn that improved FMECA method can be used to find weak links of snow depth monitoring system and offer technical support for routine maintenance of snow depth monitoring system.

Acknowledgments This study was supported by the Key Technologies Research and Development Program of China (2011BAG01B05). The author gratefully acknowledge the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially.

References

1. Qiang Z, Yongzeng Z, Xiaolong Z (2007) Failure mode analysis of side frame of freight truck based on FMECA. Railway Locomotive Car 3:35–37 + 67.(in Chinese)
2. Hanliang Z (2009) Reliability research on diesel based on FMECA. Univ Electron Sci Technol China (in Chinese)
3. Sankar NR, Prabh BS (2001) Modified approach for prioritization of failures in a system failure mode and effects analysis. Int J Qual Reliab Manag 18(3):324–336
4. Jun X (2014) Reliability analysis and its application of metro door system. Nanjing Univ Sci Technol (in Chinese)
5. Jun X, Jingbao R, Haiyan J (2014) Reliability analysis of metro door based on fuzzy FMECA. Mach Manufact Autom 02:184–187 (in Chinese)
6. Liu HC, Liu L (2011) Failure mode and effects analysis using fuzzy evidential reasoning approach and grey theory. Experts Syst Appl 38:4403–4415

7. Liu HC, Liu L, Liu N (2012) Risk evaluation in failure mode and effects analysis with extended VIKOR method under fuzzy environment. Expert Syst Appl 39(17):12926–12934
8. L HC, Mao LX (2013) Induced aggregation operators in the VIKOR method and its application in material selection. Appl Math Model 37:6325–6338
9. Chunyan D, Xiliang Z, Enchuang W, Zhen L (2012) Research on renewable energy technology evaluation based on VIKOR. Sci Decis Making 01:65–77 (in Chinese)

Chapter 58
Optimal Departure Time of Feeder Buses Under the Condition of Coordinated Dispatching

Shen Lyu, Feng Tian and Yikui Mo

Abstract According to the four key factors—the waiting time distribution of transfer passengers, the departure intervals of the feeder bus, departure time of the feeder bus, and the operational reliability of the feeder bus, we put forward a model about transfer passenger average waiting time in accordance with the normal distribution of the transfer passenger arriving time. In the model, the optimal departure time is determined with the differentiation theory in order to minimize the average waiting time. At the same time, the sensitivity analysis is also carried out to determine stability of the optimal departure time. Results indicate that the optimal departure time does not vary with the departure intervals. During the peak hours, impacts of both the average value of the transfer passenger arriving time and its standard deviation from the optimal departure time can be ignored. During the off-peak hours, the variation of the average value can be ignored, while the variation of the standard deviation must be considered. It can be concluded that the optimal departure time of feeder bus is stable during the peak hours.

Keywords Urban rail transit · Optimal departure time · Feeder bus · Coordinated dispatching · Sensitivity analysis

58.1 Introduction

With the development of intelligent transport system, the realization of coordinated dispatching of rail transit and feeder bus using advanced public transport system (APTS) is an effective way to reduce transfer and waiting time; it is also a hot research field which is widely focused by China and foreign countries. At present,

S. Lyu (✉) · Y. Mo
Civil Engineer College, ShenZhen University, No. 3688 Nan Hai Road,
Nan Shan District, ShenZhen, China
e-mail: lvshen@szu.edu.cn

F. Tian
ShenZhen Urban Transportation Planning Center Co. Ltd, Shenzhen, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_58

the study on coordinated dispatching optimization mainly focuses on optimal methods of departure interval of feeder bus. Under the coordinated dispatching, the common optimal method of departure interval was established by Lee [1, 2] and Schonfeld [3], Shoaib [4] expanded to coordinated dispatching of public transport system with the main axis of rail transit route. According to this thought, Zhang [5], Chen [6], He [7], and Teng [8] made studies on departure interval optimization between rail transit and conventional public transport under the coordinated dispatching. While in aspect of departure time optimization, Ceder [9] optimized departure time of different bus route vehicles in each station with maximization of vehicles arriving transfer junction synchronously. Zhou [10] put forward optimal method of departure time of conventional public transport with the synchronous transfer of rail transit under ideal condition. In general, optimization of departure interval was emphasized, while the study on departure time optimization is fewer. The main reason is that the current studies are based on the following assumption: Without consideration of transfer walking time, the arriving of sending vehicles and departure of receiving vehicles take the same time. In fact, during transfer from rail transit to feeder bus, even in seamless transfer rail transit hub, the transfer walking distance exists objectively. And the arriving time is random due to different individual behaviors of transfer passengers. Even if under the condition of same departure interval of feeder bus, there are big differences in waiting time of transfer passengers due to different departure time of feeder bus, so it is necessary to optimize departure time of feeder bus reasonably to promote efficiency of public transport resources and reduce waiting time of transfer passengers.

The study in this paper is based on arriving time of rail transit and optimization of departure interval of feeder bus, the average waiting time model is established according to the analysis on influenced factors of average waiting time of transfer passengers, the departure time optimal method of feeder bus is studied on minimum of average waiting time of transfer passengers from planned dispatching.

58.2 Average Waiting Time Model of Transfer Passengers

Because of the high reliability and short departure interval of rail transit operation, the problem of low transfer efficiency between rail transit and feeder bus is mainly embodied in transfer from rail transit to feeder bus. So passengers who transfer from rail transit to feeder bus are chosen as a study object. The average waiting time of passengers who transfer from rail transit to feeder bus mainly receive comprehensive functions of arriving time distribution of transfer passengers, departure interval of feeder bus, departure time of feeder bus, reliability of operation, and so on. Among them, because in morning and evening peak hours the proportion of passengers who directly transfer feeder bus after arrival of rail transit reaches more than 90 % [11], this paper takes arriving time distribution of passengers who transfer feeder bus directly as a studying object. According to the current research results, it is believed that the arriving time distribution of direct transfer passengers

obeys normal distribution [11, 12]. In coordinated dispatching conditions, the departure interval of feeder bus is usually the integral multiple (usually 1, 2, 3) [1–3] of rail transit. Based on clear departure interval of feeder bus, departure time of feeder bus is a very important factor which decides average waiting time of transfer passengers [14]. At last the reliability of feeder bus operation decides actual waiting time of transfer passengers directly. In the above-mentioned four factors, arriving time distribution of transfer passengers belongs to factors in aspect of transfer passenger behaviors, departure interval of feeder bus and departure time belong to factors in aspect of planned dispatching of feeder bus. This paper mainly studies the influence of transfer passenger behaviors and feeder bus planned dispatching on average waiting time of transfer passengers.

Firstly, the establishment conditions of model are assumed as follows: (1) The departure intervals of feeder bus and rail transit are the same, which means $H_b = H_s = H$; (2) the time is marked as zero when rail transit arrived; and (3) the vehicles of feeder bus are enough, which can satisfy requirements of departure interval. The arriving time of passengers who transfer from any metro to feeder bus conforms to normal distribution, as shown in Fig. 58.1. The average waiting time model of transfer passengers is established as follows:

$$\begin{aligned} t_w &= \int_0^{t_0} f(t)(t_0 - t)\mathrm{d}t + \int_{t_0}^{H} f(t)(H + t_0 - t)\mathrm{d}t \\ &= \int_0^{H} f(t)(t_0 - t)\mathrm{d}t + H\int_{t_0}^{H} f(t)\mathrm{d}t \\ &= t_0\int_0^{H} f(t)\mathrm{d}t - \int_0^{H} f(t)t\mathrm{d}t + H\int_{t_0}^{H} f(t)\mathrm{d}t \\ &= t_0 + H\int_{t_0}^{H} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t + \frac{\sigma}{\sqrt{2\pi}}\left(e^{-\frac{(H-\mu)^2}{2\sigma^2}} - e^{-\frac{\mu^2}{2\sigma^2}}\right) - \mu \end{aligned} \tag{58.1}$$

where $f(t)$ is normally distributed arriving time of transfer passengers from metro to feeder bus; t is maximum time spent from metro to feeder buses; μ is mean value of arriving time of the transfer passengers; σ is standard deviation of arriving time of the transfer passengers; t_w is average waiting time of transfer passengers; t_0 is departure time of feeder bus; and H is departure interval of feeder bus and rail transit.

The first item in formula (58.1) is departure time of feeder bus, the second item is shaded area in Fig. 58.1 multiplying departure interval, and the third item is a constant term which has nothing to do with departure time of feeder bus. So the change of average waiting time of transfer passengers is mainly decided by the first and second items.

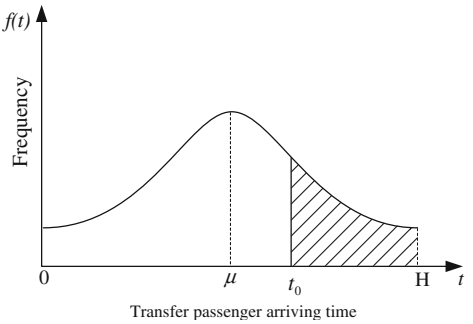


Fig. 58.1 Normal distribution of transfer passenger arriving time

58.3 Optimization of Feeder Bus Departure Time

Taking the first partial derivatives of formula (58.1), the departure time of feeder bus is determined when taking maximum or minimum value of average waiting time of transfer passengers.

$$\frac{\partial t_w}{\partial t_0} = 1 - H\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t_0-\mu)^2}{2\sigma^2}}\right) = 0 \tag{58.2}$$

$$t_0 = \mu \pm \sigma\sqrt{2\ln\frac{H}{\sigma\sqrt{2\pi}}} \tag{58.3}$$

When $\frac{\partial t_w}{\partial t_0} > 0$, which means $t_0 > \mu + \sigma\sqrt{2\ln\frac{H}{\sigma\sqrt{2\pi}}}$ or $t_0 < \mu - \sigma\sqrt{2\ln\frac{H}{\sigma\sqrt{2\pi}}}$, t_w is an increasing function. When $\frac{\partial t_w}{\partial t_0} < 0$, which means $\mu - \sigma\sqrt{2\ln\frac{H}{\sigma\sqrt{2\pi}}} < t_0 < \mu + \sigma\sqrt{2\ln\frac{H}{\sigma\sqrt{2\pi}}}$, t_w is a decreasing function.

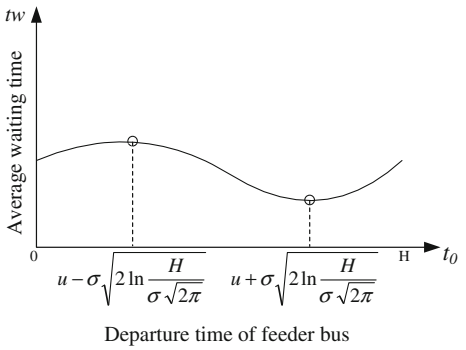


Fig. 58.2 Variation of average waiting time with the departure time of feeder buses

Table 58.1 The value of transfer passenger average waiting time with feeder bus departure time

	Arriving time of 0.5 % transfer passenger	Arriving time of 2.5 % transfer passenger	When it is the maximum waiting time	Arriving time of 16 % transfer passenger	Arriving time of 50 % transfer passenger	Arriving time of 84 % transfer passenger	When it is the minimum waiting time	Arriving time of 97.5 % transfer passenger	Arriving time of 99.5 % transfer passenger
Departure time of feeder bus	0.194	0.628	1.1	1.3	2	2.7	2.9	3.372	3.806
Average waiting time	2.191	2.345	2.73	2.675	2.009	1.34	1.2	1.472	1.826

Unit: minute

When $\frac{\partial^2 t_w}{\partial t_0^2} > 0$, which means $t_0 > \mu, t_w$ is a concave function; $\frac{\partial^2 t_w}{\partial t_0^2} < 0$, which means $t_0 < \mu, t_w$ is a convex function.

So, it can be determined that the change rule of the average waiting time with departure time is shown in Fig. 58.2, and the optimal departure time is

$$t_0 = \mu + \sigma\sqrt{2\ln\frac{H}{\sigma\sqrt{2\pi}}} \tag{58.4}$$

Taking Shenzhen City as an example, according to the investigation in metro transfer hub, the departure interval of feeder bus and the mean value of transfer passengers arriving time, respectively, equal 4 min and 2. Since the variance of arriving time of transfer passengers is not large, it is generally less than half of the mean value. This paper takes standard deviation of arriving time as 0.7. The average waiting time of transfer passengers is calculated when feeder bus is in different departure time and optimal departure time of feeder bus is determined which are shown in Table 58.1 specifically. Among them, the arriving time of 0.5, 2.5, 16, 50, 84, 97.5, and 99.5 % of transfer passengers corresponds to $\mu - 2.58\sigma$, $\mu - 1.96\sigma$, $\mu - \sigma$, μ, $\mu + \sigma$, $\mu + 1.96\sigma$, and $\mu + 2.58\sigma$, respectively.

It can be easily known from Table 58.1 that when departure interval of feeder bus is 4 min, the optimized departure time of feeder bus is 2.9 min. At this time, the average waiting time of transfer passengers is the minimum value, and it is only 1.2 min which is 40 % less than 2 min, the half of the departure interval.

58.4 Stability of Optimal Departure Time

This part uses sensitivity analysis to study whether optimal departure time of feeder bus changes with arriving behaviors of transfer passengers and departure interval and find out the degree of the changes.

58.4.1 Departure Interval of Feeder Bus

In formula (58.4), the condition of application is $H_b = H_s = H$, the change of optimal departure time of feeder bus is analyzed as follows when $H_b = 2H$ and $H_b = 3H$. Firstly, average waiting time model of transfer passengers is established. When $H_b = 2H$:

$$t_w = \frac{1}{2}\left(\int_0^{t_0}(t_0 - t)f(t)\mathrm{d}t + \int_H^{2H}(2H + t_0 - t)f(t)\mathrm{d}t + \int_{t_0}^{H}(2H + t_0 - t)f(t)\mathrm{d}t\right)$$
$$= t_0 + H\int_{t_0}^{H} f(t)\mathrm{d}t + H - \frac{1}{2}\int_0^{2H} tf(t)\mathrm{d}t$$
$$= t_0 + H\int_{t_0}^{H}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t_0-\mu)^2}{2\sigma^2}}\mathrm{d}t + H + \frac{\sigma}{2\sqrt{2\pi}}\left(e^{-\frac{(2H-\mu)^2}{2\sigma^2}} - e^{-\frac{\mu^2}{2\sigma^2}}\right) - \mu \tag{58.5}$$

When $H_\mathrm{b} = 3H$

$$t_w = \frac{1}{3}\left(\int_0^{t_0}(t_0 - t)f(t)\mathrm{d}t + \int_{2H}^{3H}(3H + t_0 - t)f(t)\mathrm{d}t + \int_H^{2H}(3H + t_0 - t)f(t)\mathrm{d}t\right.$$
$$\left. + \int_{t_0}^{H}(3H + t_0 - t)f(t)\mathrm{d}t\right)$$
$$= t_0 + H\int_{t_0}^{H} f(t)\mathrm{d}t + 2H - \frac{1}{3}\int_0^{3H} tf(t)\mathrm{d}t$$
$$= t_0 + H\int_{t_0}^{H}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t_0-\mu)^2}{2\sigma^2}}\mathrm{d}t + 2H + \frac{\sigma}{3\sqrt{2\pi}}\left(e^{-\frac{(3H-\mu)^2}{2\sigma^2}} - e^{-\frac{\mu^2}{2\sigma^2}}\right) - \mu \tag{58.6}$$

It can be found that the first and second items in the three formulas are completely the same and there is only difference in the third item of the three formulas in comparison with formula (58.1), (58.5), and (58.6). Because the third item is a constant term which has nothing to do with departure time, the results of $\frac{\partial t_w}{\partial t_0}$ in formula (58.5) and (58.6) are completely the same as results in formula (58.3). So when $H_\mathrm{b} = \mathrm{m} \cdot \mathrm{H_s}$ under the coordinated dispatching using ATPS, the optimal departure time of feeder bus keeps unchanged, which is $t_0 = \mu + \sigma\sqrt{2\ln\frac{H}{\sigma\sqrt{2\pi}}}$.

58.4.2 Arriving Time Distribution of Transfer Passengers

The optimal departure time of feeder bus is the function of mean value μ and standard deviation σ of transfer passengers arriving time. Because there is a certain

degree of difference in arriving time mean value μ and standard deviation σ transferring from each metro to feeder bus, the influence of mean value μ and standard deviation σ on optimal departure time is studied as follows.

58.4.2.1 The Influence of Mean Value

According to investigations on arriving time distribution rule of transfer passengers, there are differences in arriving time mean value during the peak period and off-peak period in integrated transfer hub due to varieties of transfer passenger behaviors. During the off-peak period, the mean value of arriving time is the same, while during the peak period, the mean value of arriving time has some difference influenced by passenger flow volumes in the hub, but the difference is not very large (less than 10 %) [13]. When only considering changes of mean value, the influence on minimum average waiting time is very small. When the change of mean value is in the range of ±20 %, the minimum average waiting time is less than ±2 % [14], which can be ignored. So the influence of mean value on optimal departure time of feeder bus could not take into consideration.

58.4.2.2 The Influence of Standard Deviation

According to the investigation, the difference of arriving time standard deviation is small during the peak period, which is generally in the range of ±10 %. While during the off-peak period, the difference of arriving time standard deviation is in the range of ±30 %, which is larger. The change of average waiting time caused by standard deviation cannot be ignored [14]. So the optimal departure time of feeder bus is varied with standard deviation.

It is assumed that the mean value of arriving time keeps unchanged and analysis is made on changes of the optimal departure time caused by standard deviation. The results are shown in Fig. 58.3. It can be found that optimal departure time of feeder bus caused by standard deviation is less than 3 % when standard deviation

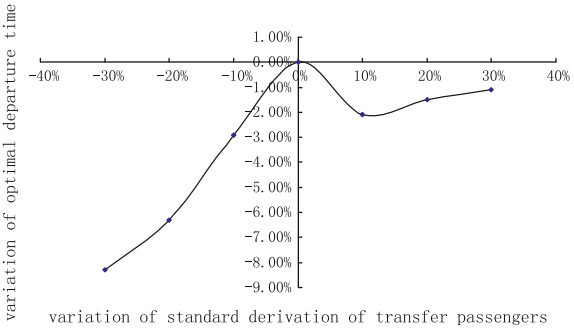


Fig. 58.3 The variation of optimal departure time with standard derivation

fluctuates in the range of ±10 % during the peak period, which is relatively small. So the change of the optimal departure time caused by standard deviation can be ignored during the peak period. While during the off-peak period, the standard deviation of arriving time fluctuates in the range of ±30 %, the change of the optimal departure time is in the range of ±9 %, which is relatively large. So the change of optimal departure time caused by the standard deviation cannot be ignored during the off-peak period.

In conclusion, the influences of arriving time mean value and standard deviation on optimal departure time can be ignored during the peak period. During the off-peak period, the influences of mean value on optimal departure time can be ignored, while standard deviation cannot.

58.5 Conclusion

Under the conditions of the coordinated dispatching, departure time of feeder bus is one of the key factors which decides waiting time of transfer passengers. On the point of the passengers, this research established optimal departure time model of feeder bus based on normal distribution of transfer passenger arriving time. Differential theory was applied to solve this model and studied the stability of optimal solution. The result shows that the optimal departure time of feeder bus has good stability during the peak period under coordinated dispatching, and it does not change with the departure interval of feeder bus and arriving time distribution of the passengers. So during the peak period, the formula $t_0 = \mu + \sqrt{2\ln\frac{H}{\sqrt{2\pi}\sigma}}\sigma$ can be used to optimize departure time of feeder bus to reduce transfer waiting time of the passengers, construct seamless transfer system, and improve continuity and comfort of transit trip, which is of great practical significance in promoting the proportion of transit trip in commuting travel.

Acknowledgments This work is supported by the National Natural Science Foundation of China (No. 51208307 and No. 51208308).

References

1. Lee KKT, Schonfeld P (1991) Optimal slack time for timed transfers at a transit terminal. J Adv Transp 25(3):281–308
2. Lee KKT, Schonfeld P (1994) Real-time dispatching control for coordinated operation in transit terminals. Transp Res Rec 1433:3–9
3. Schonfeld P (1993) Optimization of timed transfers in transit terminals. University of Maryland, Washington, DC
4. Shoaib CM, Chien SI (2001) Optimization of transfer coordination for intermodal transit networks. Transportation Research Board the 80th Annual Meeting, preprint CD-ROM, Washington DC

5. Zhang YS, Chen XM, Yu L et al (2009) Study on model of coordinated operation between urban rail and bus systems at transfer stations. J China Railway Soc 31(3):11–19 (in Chinese)
6. Chen XM, Lin GX, Yu L (2009) Modeling operation scheduling coordination between urban rail system and bus system. Syst Eng Theor Pract 29(10):165–173 (in Chinese)
7. He B (2009) Study on coordinated optimization of urban rail transit and bus service. Beijing JiaoTong University, Beijing (in Chinese)
8. Teng J, Yang XG (2004) A study on optimizing method for dispatching problem of public transit hub under the condition of APTS. Syst Eng 22(8):78–82 (in Chinese)
9. Ceder A, Golany B, Tal O (2001) Creating bus timetables with maximal synchronization. Transp Res 35A(10):913–928
10. Zhou XM, Yang XG (2004) Study of dispatching at minimum waiting time of public transportation transfer under the condition of ITS. China J Highway Transp 17(2):82–85 (in Chinese)
11. Guo SX, Chen XM, Yu L, Hu DF (2010) Average waiting time for transfer from rail transit to buses. J Transp Syst Eng Inf Technol 10(2):143–147 (in Chinese)
12. Chang D (2010) Quantified study of microscopic pedestrian behavior parameters in subway. Beijing Jiaotong University, Beijing (in Chinese)
13. Lv(Lyu) S, Tian F (2013) The means characteristics of transfer passenger arrival time at rail transit station. Transp Stand (21):21–25 (in Chinese)
14. Lu(Lyu) S, Tian F, Mo YK (2014) Average waiting time of passengers transferring from subway to feeder bus at rail transit hub. J Highway Transp Res Dev 31(12):92–97 (in Chinese)

Chapter 59
A New Approach to Attribute Reduction of Decision Information Systems

Fachao Li and Jinning Yang

Abstract Rough set theory is a theoretical tool for complete conclusions by incomplete information; its basic idea is using some partition to consider the description problem of concept on university. In this paper, we firstly introduce some usual attribute reduction methods. Secondly, we establish an effect-based roughness measure model, named as effect rough degree (ERD) by combining with a basic measure factor for roughness-lower (upper) accuracy of rough set. Finally, we propose an ERD-based attribute reduction method (abbreviated as ERD-RM) and then discuss the difference and relation between ERD-RM and the existing reduction methods. The theoretical analysis showed that the ERD-RM includes other reduction methods to a certain degree, and it also has basic difference with other existing methods.

Keywords Rough sets · Decision information systems · Attribute reduction · Effect rough degree

59.1 Introduction

Rough set theory was proposed by Pawlak [1] in 1984. Since then, rough set theories and its applications have attracted an extensive attention in the academic field. Further, many researchers have made several useful discussions and also obtained lots of important achievements. For example, Ziarko [2] proposed a variable precision rough sets model, aimed at modeling classification problems

F. Li · J. Yang (✉)
School of Economics and Management, Hebei University of Science and Technology, 050018 Shijiazhuang, Hebei, China
e-mail: sunnyyjn@163.com

F. Li
e-mail: lifachao@tsinghua.org.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_59

involving uncertain or imprecise information as the objects often have multiple ownerships. Wang et al. [3] discussed some basic properties of covering information systems and decision systems under homeomorphisms.

Although rough set theory is not exactly the same as the other data mining methods, it can integrate with others to achieve good results, which has been widely recognized. Although some scholars have discussed the characteristics of Pawlak metric, for example, Liang and Wang [4] addressed the issues of uncertainty of a set in an information system and approximation accuracy of a rough classification in a decision table. Wang [5] studied the uncertainty of information system from the information theory angle and discussed the differences between Pawlak roughness measure and roughness measure basing on information entropy. But these discussions did not give a construction principle of roughness measure.

By combining the characteristics of data mining, taking rough set theory as a tool for describing uncertain information, many scholars established data system-based attribute reduction methods. They not only can simply reflect the essential characteristics of data processing, but also has good interpretable ability. For example, Zhang et al. [6] proposed a rule confidence preserving attribute reduction approach in order to extract both the compact certain rules and the compact possible rules from a covering decision system with their confidence degree being not less than a prespecified threshold value; Zhang [7] proposed the concepts of revisionary conditional information entropy and relative fuzzy entropy in the decision information system, and using the algorithm of adding attributes for attribute reduction; Chen [8] expounded the basic conceptions of rough set theory and information entropy, and an algorithm of attribute reduction based on rough set and information entropy was put forward; Li [9], accordingly, proposed an algorithm for reduction of knowledge based on the roughness of knowledge.

Based on the above analysis, for the attribute reduction of decision information systems, we mainly do the following work: (1) We introduced several attribute reduction methods; (2) we established a roughness measure model with structural features (called effect rough degree, and abbreviated as ERD); (3) we proposed a reduction method based on ERD (abbreviated as ERD-RM) and analyzed the relationship and differences between ERD-RM and other methods through specific examples.

59.2 Several Common Attribute Reduction Methods

For convenience, in the following: (1) $|A|$ denotes the number of elements in a finite set A; (2) for the equivalence relation R on U (that is, $R \subset U \times U$ and it satisfies the following: (i) $(x, x) \in R$ for any $x \in U$; (ii) if $(x, y) \in R$, then $(y, x) \in R$; (iii) if $(x, y) \in R$, $(y, z) \in R$, then $(x, z) \in R$)); $[x]_R = \{y \mid (x, y) \in R\}$ denotes the R equivalence class of x, $U/R = \{[x]_R \mid x \in U\}$, and (U, R) is called approximation space; (3) (U, C, d, F_C) denotes a decision information system; (4) For (U, C, d, F_C) and $B \subset C$, let $R_B = \{(x, y) | (x, y) \in U \times U$ and $f_c(x) = f_c(y)$ for any $c \in B\}$ be the

equivalence relation on U determined by B, $[x]_B$ be the R_B-equivalence classes of x; let $R_d = \{(x, y)|(x, y) \in U \times U \text{ and } f_d(x) = f_d(y)\}$ be the equivalence relation on U determined by decision attribute d, $[x]_d$ be the R_d-equivalence classes of x, and we briefly sign U/R_B as U/B and U/R_d as U/d.

Definition 2.1 (see [1]) Let U be a finite universe, (U, R) be an approximation space, $X \subset U$,

$$\underline{R}(X) = \{x \,|\, x \in U, \text{ and } [x]_R \subset X\}, \tag{59.1}$$

$$\bar{R}(X) = \{x \,|\, x \in U, \text{ and } [x]_R \cap X \neq \phi\}. \tag{59.2}$$

(1) Let $\underline{R}(X)$ be the R-lower approximation set of X, and $\bar{R}(X)$ be the R-upper approximation set of X. (2) If $\underline{R}(X) = \bar{R}(X)$, then we call X is definable about (U, R); otherwise, X is not definable about (U, R). (3) Let $\underline{R}(X) \triangleq Pos(X, R)$ be the R positive region of X, $U - \bar{R}(X) \triangleq Neg(X, R)$ be the R negative region of X. (4) We call

$$\alpha_R(X) = \begin{cases} \frac{|\underline{R}(X)|}{|\bar{R}(X)|}, & X \neq \emptyset \\ 1, & X = \emptyset \end{cases} \tag{59.3}$$

as the R approximate accuracy of X, and

$$\rho_R(X) = 1 - \alpha_R(X) \tag{59.4}$$

as the R roughness of X. For convenience, and we have (4) be the Pawlak Roughness measure in the following.

As a quantitative description method of rough uncertainty, roughness measure is not only an important concept of rough set theory, but also a critical link to establish rough decision-making method. And its application is broad that many scholars had got some good conclusions, especially in the reduction aspect, we can find some common attribute reduction methods as follows:

Partition reduction: If B is a partition coordination set of C (that is, $R_B = R_C$), and any $E \subset B$, $E \neq B$ is not a partition coordination set of C, then we call B is a partition reduction of (U, C, d, F_C).
Division reduction: If B is a distribution coordination set of C (that is, $|D_i \cap [x]_B| \,/\, |[x]_B| = |D_i \cap [x]_C| \,/\, |[x]_C|$ for any $i \in \{1, 2, \cdots, m\}$ and $x \in U$), and any $E \subset B$, $E \neq B$ is not a distribution coordination set of C, then we call B is a division reduction of (U, C, d, F_C).
Max division reduction: If B is the maximum distribution coordination set of C (that is, $\max\{|D_i \cap [x]_B| \,/\, |[x]_B|\} = \max\{|D_i \cap [x]_C| \,/\, |[x]_C|\}$), and any $E \subset B$, $E \neq B$ is not the maximum distribution coordination set of C, then we call B is a max division reduction of (U, C, d, F_C).

Positive domain reduction: If B is positive region coordination set of C (that is, $\cup_{i=1}^{m}\underline{R_B}(D_i) = \cup_{i=1}^{m}\underline{R_C}(D_i)$), and any $E \subset B,\ E \neq B$ is not positive region coordination set of C, then we call B is a positive domain reduction of (U, C, d, F_C).
Lower ⊕ upper approximate reduction: If B is a lower ⊕ upper approximation coordination set of C (that is, $\underline{R_B}(D_i) = \underline{R_C}(D_i)$ and $\overline{R_B}(D_i) = \overline{R_C}(D_i)$ hold for any $i \in \{1,\ 2, \cdots, m\}$), any $E \subset B,\ E \neq B$ is not a lower ⊕ upper approximation coordination set of C, then we call B is a lower ⊕ upper approximate reduction of (U, C, d, F_C).

59.3 Effect Rough Degree of Rough Set

Because the usual measure mode (4) only has considered the differences between upper and lower approximation set of X, but has not considered the features of X itself, which often leads to inconsistent in measurement results and intuitive facts. In order to makes up the mutability of Pawlak roughness measure, in this section, we will discuss the construction mechanism of rough measurement.

I think a reasonable roughness measure mode should have the following two characteristics: (1) It can reflect the role of concept itself; (2) It can reflect the link between the concept and its lower approximation and upper approximation. For $X \neq \emptyset$, if we let

$$\alpha_{\underline{R}}(X) = \frac{|\underline{R}(X)|}{|X|},\ \alpha_{\bar{R}}(X) = \frac{|X|}{|\bar{R}(X)|}, \tag{59.5}$$

then it is easy to see that $(\alpha_{\underline{R}}(X),\ \alpha_{\bar{R}}(X))$ meet the above two characteristics, so $(\alpha_{\underline{R}}(X),\ \alpha_{\bar{R}}(X))$ can be used as a pair of basic indicts for constructing roughness measure. For convenience, we call $\alpha_{\underline{R}}(X)$ as R-lower accuracy of X and $\alpha_{\bar{R}}(X)$ as R-upper accuracy of X and suppose that $\alpha_{\underline{R}}(\phi) = \alpha_{\bar{R}}(\phi) = 1$.

For approximation space (U, R) and $X \subset U$, we can easily learn that if we abstract the synthesis problem of $\alpha_{\underline{R}}(X)$ and $\alpha_{\bar{R}}(X)$ as a function $S(u,\ v)$ (called **effect synthesis function**) of $[0, 1] \times [0, 1] \to [0, 1]$, then $S(u,\ v)$ should satisfy the following conditions: (1) ***Monotonicity***: when $u_1 \leq u_2,\ v_1 \leq v_2,\ S(u_1, v_1) \leq S(u_2, v_2)$; (2) ***Normativity***: $S(0, 0) = 0,\ S(1, 1) = 1$.

It is easy to verify that

$$S_1(u, v) = w_1 u^{\alpha} + w_2 v^{\beta}, \tag{59.6}$$

$$S_2(u, v) = \begin{cases} \frac{u+v-\varepsilon}{2-\varepsilon}, & u + v \geq \varepsilon, \\ 0, & u + v < \varepsilon, \end{cases} \tag{59.7}$$

$$S_3(u,v) = \begin{cases} 0, & \max\{u,v\} \le k_1 \\ \dfrac{\max\{u,v\} - k_1}{k_2 - k_1}, & k_1 \le \max\{u,v\} \le k_2 \\ 1, & \max\{u,v\} \ge k_2 \end{cases} \tag{59.8}$$

are all effect synthesis functions. Here, $\alpha \ge 0,\ \beta \ge 0;\ w_1, w_2 \ge 0$, and $w_1 + w_2 = 1$; $0 \le \varepsilon \le 2;\ 0 \le k_1 < k_2 \le 1$.

For convenience, in the following we call

$$ERD(X, R \oplus S) = 1 - S_R(\underline{\alpha_R}(X), \alpha_{\bar{R}}(X)) \tag{59.9}$$

effect rough degree based on (U, R) and $S(u, v)$.

It is easy to verify that (1) if $S(u,v) = uv$, then $ERD(X, R \oplus S)$ is the roughness measure model (4) in this paper [1]. It indicates that (9) is an extension of Pawlak roughness measure. (2) When X is R definable set on (U, R), $ERD(X, R \oplus S) = 0$ for any effect synthesis function. It shows that (9) is consistent with the intuitive fact "definable set is not rough." (3) For the right effect synthesis function (such as $S(u,v) = w_1 u^{\alpha} + w_2 v^{\beta}$), model (9) can reflect the continuity of roughness measure. We can see that this feature makes up the mutability of Pawlak roughness measure.

Above all, (9) is a general roughness measure method. Where $S(u, v)$ can be understood as a criteria for identifying the roughness in different preference, the roughness measure results of rough set may be different (even with a big difference) with different $S(u, v)$. We can see that the core of effect rough degree is effect synthesis function. Therefore, when building a roughness measure model, we should systematically select effect synthesis function combing with the characteristics of specific issues and the action features of roughness measure in the decision-making process.

59.4 Attribute Reduction Based on Effect Roughness and Compare with Other Methods

For a decision information system (U, C, d, F_C), $B \subset C$ and $S(u, v)$ be a effect synthesis function. If we sign $U/d = \{D_1, D_2, \ldots, D_m\}$, then

$$ERD(U/d, B \oplus S) = \sum_{i=1}^{m} \frac{|D_i|}{|U|} \cdot ERD(D_i, R_B \oplus S) \tag{59.10}$$

is a kind of weighted average effect rough degree about (U, R_B) (we call it an effect rough degree of U/d based on (U, R_B) and $S(u, v)$) where $|D_i| / |U|$ represents the share of D_i in U. If we take (U, B, d, F_B) as a kind of decision knowledge base, then $ERD(U/d, B \oplus S)$ can be understood as a kind of credibility measure for decision-making knowledge. Therefore, $\Delta(B) = ERD(U/d, B \oplus S) - ERD(U/d,$

Table 59.1 Satisfactory sheet on scientific research teaching quality from 9 institutes

U	C_1	C_2	C_3	C_4	d
x_1	2	1	2	3	1
x_2	2	2	2	2	2
x_3	1	2	1	1	2
x_4	1	2	1	1	1
x_5	1	2	1	1	2
x_6	1	2	3	2	2
x_7	1	2	1	1	1
x_8	1	2	3	2	2
x_9	2	2	2	2	2

$C \oplus S$) reflects the knowledge difference between (U, B, d, F_B) and (U, C, d, F_C). We can establish an attribute reduction method (effect rough degree- based reduction methods for short, abbreviated as ERD-RM) through giving $\Delta(B)$ a value smaller than the given threshold value δ (that is, $\Delta(B) \leq \delta$). For $\delta = 0$, we will analyze the feature of ERD-RM and the relation between ERD-RM and the existing reduction methods in the following.

We can find that when the effect synthesis functions are not strictly monotonic increasing, ERD-RM may be different from any other reduction methods in Sect. 59.2. We will further analyze the characteristics of ERD-RM, combining with Table 59.1 which gives us some satisfaction situations of six scientific institute researches on teaching quality of a university where (1) C_1, C_2, C_3, C_4 is conditional attribute, and d is decision attribute. (2) C_1 presents the school's hardware facilities, values for {common (1), good (2)}, C_2 presents the school's software facilities, values for {good (1), common (2)}, C_3 presents the school's teacher quality, values for {low (1), common (2), good (3)}, C_4 presents the school's students quality, values for {good (1), common (2), low(3)}, satisfaction (d) values for {displeasure (1), satisfy (2)}; (3) x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9 presents six scientific institutes.

Let $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$ and $C = \{C_1, C_2, C_3, C_4\}$. Then, (1) $R_C = \{\{x_1\}, \{x_2, x_9\}, \{x_3, x_4, x_5, x_7\}, \{x_6, x_8\}\}$, $R_d = \{D_1, D_2\} = \{\{x_1, x_4, x_7\}, \{x_2, x_3, x_5, x_6, x_8, x_9\}\}$; and (2) ($U$, C, d, F_C) is a inconsistent decision information system that if $[x]_A \subset [x]_d$ do not hold for any $x \in U$; and (3) reduction results of C concerning several reduction methods and ERD-RM are shown in Table 59.2.

Synthesizing the above analysis, we can see that (11) is a generalized roughness measure methods where $S(u, v)$ can be understood as the roughness-identified criteria of decision-makers. Therefore, the roughness measure results of rough set may be different with different $S(u, v)$; then, the corresponding EDR-RM is different. For example, in terms of the information systems in Table 59.1, we can learn the following from Tables 59.2 and 59.3: (1) For $S_1(u, v) = u^{\alpha}v^{\beta}$, the reduction results of EDR-RM are same with the results of partition reduction, lower $\oplus$ upper approximate reduction and partition reduction; (2) for $S_2(u, v)$, we can know that the reduction results vary with ε; and (3) For $S_3(u, v)$, we can know that the reduction results vary with different k_1 and k_2.

Table 59.2 Comparison of reduction results of several reduction methods for Table 59.1

Reduction method	Reduction results
Partition reduction method	$\{C_2, C_3\}, \{C_1, C_4\}, \{C_3, C_4\}$
Distribution reduction method	$\{C_2, C_3\}, \{C_1, C_4\}, \{C_3, C_4\}$
Max distribution reduction method	$\{C_2, C_3\}, \{C_4\}$
Lower ⊕ upper approximate reduction method	$\{C_2, C_3\}, \{C_4\}$
EDR-RM	Table 59.3

Table 59.3 Reduction results of EDR-RM for Table 59.1

Effect synthesis function $S(u, v)$			Reduction results
$S_1(u, v) = uv$			$\{C_2, C_3\},\{C_4\}$
$S_2(u,v) = \begin{cases} \frac{u+v-\varepsilon}{2-\varepsilon}, & u+v \geq \varepsilon \\ 0, & u+v<\varepsilon \end{cases}$	$\varepsilon = 1.5$		$\{C_1\}, \{C_2\}$ or $\{C_3\},\{C_4\}$
	$\varepsilon = 0.5$		$\{C_2, C_3\},\{C_4\}$
$S_3(u,v) = \begin{cases} 0, & \max\{u,v\} \leq k_1 \\ \frac{\max\{u,v\}-k_1}{k_2-k_1}, & k_1 \leq \max\{u,v\} \leq k_2 \\ 1, & \max\{u,v\} \geq k_2 \end{cases}$	$k_1 = 0.1$	$k_2 = 0.9$	$\{C_2, C_3\},\{C_4\}$
	$k_1 = 0.6$	$k_2 = 0.7$	$\{C_2\},\{C_4\}$
	$k_1 = 0.2$	$k_2 = 0.3$	$\{C_1\}, \{C_2\}$ or $\{C_3\},\{C_4\}$

59.5 Conclusions

In this paper, for the attribute reduction of decision information system, we mainly do the following work: (1) We introduce several common attribute reduction methods; (2) we propose the roughness measure model based on the synthesis effect; (3) we establish an ERD-based attribute reduction method; and (4) we discuss the difference and relation between ERD-RM and the existing reduction methods. Theoretical analysis indicates that the ERD-RM includes other reduction methods to a certain degree, and it also has basic difference with other existing methods. It is worth noting that using ERD-RM to find minimum reduction is still a NP difficult problem. Our discussions cannot only enrich the existing theories to a certain degree, but also has wide application value in many fields such as artificial intelligence, resource management, and optimization of complex systems.

Acknowledgment This work is supported by the National Natural Science Foundation of China (71371064) and the Natural Science Foundation of Hebei Province (F2015208100).

References

1. Pawlak Z (1982) Rough sets. Int J Inf Comput Sci 11(5):341–356
2. Ziarko W (1993) Variable precision rough set model. J Comput Syst Sci 46(1):39–59

3. Wang C, Chen D, Sun B, Qinghua H (2012) Communication between information systems with covering based rough sets. Inf Sci 216(20):17–33
4. Liang JY, Wang JH (2009) A new measure of uncertainty based on knowledge granulation for rough sets. Inf Sci 179:458-470
5. Guoyin W (2001) Rough set theory and knowledge acquisition. Xi'an Jiao Tong University press, Shaanxi
6. Zhang X, Mei C, Chen D, Li J (2013) Multi-confidence rule acquisition oriented attribute reduction of covering decision systems via combinatorial optimization. Knowl Based Syst (50):187–197
7. Zhang Q, Xiao Y (2013) New attribute reduction on information entropy. J Front Comput Sci Technol 7(4):359–367 (in Chinese)
8. Chen Y, Yang D (2013) Attribute reduction algorithm based on information entropy and its application. J Chongqing Univ Technol (Nat Sci) 27(1):42–46 (in Chinese)
9. Li X, Wang X (2009) A roughness of knowledge based algorithm for reduction of knowledge. J Xi'an Univ Arts Sci 12(2):40–43 (in Chinese)

Chapter 60
Analysis of Urban Rail Vehicle Bogie System Reliability Based on the Theory of Survival

Huaixian Yin, Kai Wang, Yong Qin, Limin Jia and Qingsong Hua

Abstract The bogie failure time has delete loss in urban rail train; as a result, it is of great significance for the running stability and security to evaluate the reliability of bogie system and predict bogie system failure rate accurately. Based on the field failure data of urban rail train bogie system as an example, the maximum likelihood estimation method in survival analysis theory is adopted to analyze reliability. Minitab software is used to determine the best failure distribution model of the subsystems and through the study of the comparison and analysis of the key components to determine the weak link of bogie system. Simulation and experimental results show that the survival analysis theory can effectively solve the delete data loss under the mechanism of failure time interval uncertain problems, and we can get more reasonable and reliable result.

Keywords Bogie · Reliability · Survival analysis · Urban rail vehicles

60.1 Introduction

The bogie system is one of the most important components in urban rail train. It is used to transfer various loads and stick between the rails to ensure the generation of traction [1]. Its performance largely determines urban rail train running stability and safety [2, 3]. Therefore, accurate measurement of the reliability and prediction of

H. Yin · Y. Qin (✉) · L. Jia
School of Traffic and Transportation, Beijing Jiaotong University,
No. 3 Shang Yuan Cun, Hai Dian District, 100044 Beijing, China
e-mail: qinyong2146@126.com

H. Yin
e-mail: yhx6279010@163.com

H. Yin · K. Wang · Q. Hua
Dynamic Integration and Energy Storage Systems Engineering
Technology Research Center, Qingdao University, Qingdao 266071, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_60

bogie system failure rate is a great significance to the stability and safety of urban rail train [4–6].

Survival analysis theory has developed for nearly thirty years, which is a new branch of mathematical statistics [7–9]. It is widely used in medicine, biology, economics, demography, public health, engineering, and etc. Survival analysis is an analyzing and statistical method to survival time in biology. The data are not complete, which is the important characteristic to survival data in survival analysis. In other words, at the end of the study period, some individuals do not appear concerned event, or the specific time is unknown, so the data are delete loss or truncated. For example, in the reliability study, life data can be obtained in the field investigation and deliberately arrange test [10]. Therefore, in all sorts of truncated data (destiny truncation, timing truncation random truncation, etc.), there may be only a portion of the sample individuals who know the exact life value, and rest only know more than a particular life value, or unknown expiry life because of failure about the sample at a certain time, or can only determine less than a particular life value because of unknown failure time of sample, and so on.

The research method in the reliability of the system has mainly statistical model and mechanism model. Statistical model method has obtained the related reliability indexes of the system by mainly analyzing the probability distribution function of system failure time, such as the average time between failures, and failure rate. Mechanism model method has analyzed reliability mainly based on the structure principles of system. These research methods are based on monitoring or inspection data, sometimes which cannot accurately grasp the life of the bogie and be lack of consideration on the uncertainty problem. This article uses the theory of survival analysis to delete process of data, so as to get more accurate life time fitting function and achieve better effect about reliability analysis of the bogie.

60.2 Reliability Analysis

This article selects wheel on the axle box system as the object. There is 589 valid data of the failure data about a metro line 2 between November 2010 and April 2013 as an example. This article concludes the failure distribution model fitting inspection figure and distribution fitting figure of wheel on the axle box system, so as to predict the rest of the subsystem.

60.2.1 Life Data Types

The research object of survival analysis is non-negative random variable T, the statistical inference accords to the observed objects. There are four types of values to observe the life of each individual, as shown in Table 60.1.

Table 60.1 Life data types

Name	Define	Symbol
Exactly life data	Know the individual's life exactly t	t
Right delete loss data	Only know the individual's life greater than t	$t+$
Left delete loss data	Only know the individual's life less than t	$t-$
Interval delete loss data	Only know the individual's life between $t1$ and $t2$	[$t1$, t2]

(1) Exactly life data: detect the exact value of the life data, extreme circumstances of interval delete data (interval length is zero).
(2) Right delete loss data: not observed the exact value of individual life, only known exact value is greater than a certain number of t, the observed value in t right delete is marked as $t+$.
(3) Left delete loss data: not observed the exact value of individual life, only known exact value is less than a certain number of t, the observed value in t left delete is marked as $t-$.
(4) Interval delete loss data: only known is the individual's life between two numbers. Common statistics only discuss data which are completely life situation and do not discuss delete data, so the survival analysis theory can be more reasonable to assess reliability of bogie system.

60.2.2 Determination of Life Distribution Model

The article uses survival analysis method with parameters to determine life distribution model of the subsystems of the bogie system through the example.

A-D inspection inspects data to obey a certain distribution, the better data and distribution of fitting, the less A-D statistics. We can be compared with several different distribution fitting by A-D statistics, so as to find out the best distribution data. A-D statistics measure the distance between figure point in probability diagram and the fitting line, which the statistic is the weighted square distance between figure point and the fitting line, and the more close to the end of the distribution, the greater the weight. The expression of A-D statistic is

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\mathrm{In}F(x_i) + \mathrm{In}(1 - F(x_{n+1-i}))] \tag{60.1}$$

Therefore, $F(x_i) = \Phi(\frac{x_i - \bar{x}}{\sigma})$ is the empirical cumulative distribution function obeys to the normal distribution.

Minitab software calculates adjust A-D statistics, the smaller the A-D statistic, the better the distribution and data fitting. If the difference is large enough between the empirical cumulative distribution function of the sample data and the

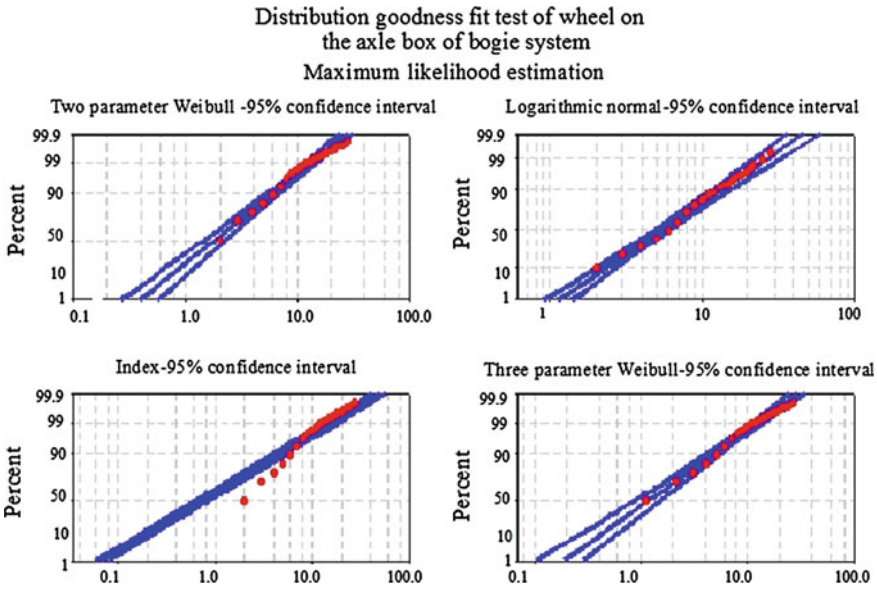


Fig. 60.1 Bogie wheel system life distribution goodness of fit test

assumption of normal distribution expected, the inspection will be denied the null hypothesis of overall normally distributed.

The fault data of wheel on the axle box is imported Minitab worksheet. This article chooses reliability/survival statistics tool and uses the maximum likelihood estimation method to carry out the parameter estimation and fitting inspection for index distribution, logarithmic normal distribution, the two-parameter Weibull distribution, and three-parameter Weibull distribution. The fitting inspection test results of the distribution model are shown as Fig. 60.1.

The calculated value of the parameters is shown in Table 60.2.

In Fig. 60.1, the life data are in the confidence interval of two-parameter Weibull distribution, logarithmic normal distribution, and three-parameter Weibull distribution; the worst fitting degree is the exponential distribution. In Table 60.2, A-D statistics of two parameters Weibull distribution is 0.436, which is the minimum value in all the distribution model ,so two parameter Weibull distribution has the best fitting effect. As a result,the optimal distribution of the wheel on the axle box system is two parameter Weibull distribution.

Table 60.2 Bogie wheel system life distribution model

Distribution model	A-D statistics	Selected distribution
Index distribution	2.169	Two-parameter Weibull distribution
Logarithmic normal distribution	0.541	
Two-parameter Weibull distribution	0.436	
Three-parameter Weibull distribution	0.464	

Similarly, the life distribution model of frame, spring device, connection device, driving mechanism, and basic of braking device can be determined. Each subsystem shown in Table 60.3 gives A-D statistics and obeys the best distribution and parameter estimation of maximum likelihood method determining in the process of life; each subsystem essentially obeys Weibull distribution.

60.2.3 Reliability Characteristic Value

Reliability can be integrated with various quantitative indicators to reflect a trouble-free, serviceability and durability of the products and efficiency, etc. Reliability, in no-reliability, failure rate, and failure density function, as long as knowing one index can calculate other three indicators and average values of life. The relationship between the indexes is shown in Fig. 60.2 (N samples).

Reliability characteristic value is used for evaluating the reliability indicators of equipments, mainly including the cumulative distribution function, reliability function, probability density function, failure rate function, and mean time between failure. Table 60.4 gives the mean time between failure of key subsystem bogie (MTBF).

The mean time between failures of basic of braking device is the highest, about 41 days, and the spring device and wheel on axle box are lower, about 6.79 days and 7.05 days, respectively.

Wheel on axle box obeys the model of two-parameter Weibull distribution, the shape parameter (β) is 1.54270, the scale parameter (η) is 7.83827, so the fault time probability density function $(f(t))$, cumulative distribution function $(F(t))$, reliability function $(R(t))$, and failure rate function $(\lambda(t))$ are as follows.

The expression for the fault time probability density function is

$$f(t) = \frac{\beta}{\eta}\left[\frac{t}{\eta}\right]^{\beta-1}\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] = \frac{1.54270}{7.83827^{1.54270}}t^{1.54270}\mathrm{e}^{-\left(\frac{t}{7.83827}\right)^{0.54270}} \tag{60.2}$$

The expression for cumulative distribution function is

$$F(t) = 1 - \exp\left\{-\left[\frac{t}{\eta}\right]^{\beta}\right\} = 1 - \mathrm{e}^{-\left(\frac{t}{7.83827}\right)^{1.54270}} \tag{60.3}$$

The expression for reliability function

$$R(t) = \exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\} = \mathrm{e}^{-\left(\frac{t}{7.83827}\right)^{1.54270}} \tag{60.4}$$

Table 60.3 The goodness-of-fit test table of each subsystem life distribution

System	Candidate distribution model	A-D statistics	Best distribution	Parameter (maximum likelihood method)
Frame	Index distribution	0.530	Three-parameter Weibull distribution	Shape: 1.11034
	Logarithmic normal distribution	0.500		Scale: 17.3206
	Two-parameter Weibull distribution	0.820		The threshold value: 0.576497
	Three-parameter Weibull distribution	0.484		
Spring device	Index distribution	2.226	Two-parameter Weibull distribution	Shape: 1.76362
	Logarithmic normal distribution	0.449		Scale: 7.62358
	Two-parameter Weibull distribution	0.328		
	Three-parameter Weibull distribution	0.346		
Connection device	Index distribution	1.808	Two-parameter Weibull distribution	Shape: 1.45032
	Logarithmic normal distribution	1.166		Scale: 28.0343
	Two-parameter Weibull distribution	0.694		
	Three-parameter Weibull distribution	0.734		
Wheel on the axle box system	Index distribution	2.169	Two-parameter Weibull distribution	Shape: 1.54270
	Logarithmic normal distribution	0.541		Scale: 7.83827
	Two-parameter Weibull distribution	0.436		
	Three-parameter Weibull distribution	0.464		
Driving mechanism	Index distribution	1.804	Three-parameter Weibull distribution	Shape: 1.37217
	Logarithmic normal distribution	0.330		Scale: 11.6082
	Two-parameter Weibull distribution	0.358		The threshold value: 0.627144
	Three-parameter Weibull distribution	0.321		

(continued)

Table 60.3 (continued)

System	Candidate distribution model	A-D statistics	Best distribution	Parameter (maximum likelihood method)
Basic of braking device	Index distribution	2.033	Three-parameter Weibull distribution	Shape: 1.41254
	Logarithmic normal distribution	1.180		Scale: 42.7584
	Two-parameter Weibull distribution	1.029		The threshold value: 2.12370
	Three-parameter Weibull distribution	1.025		

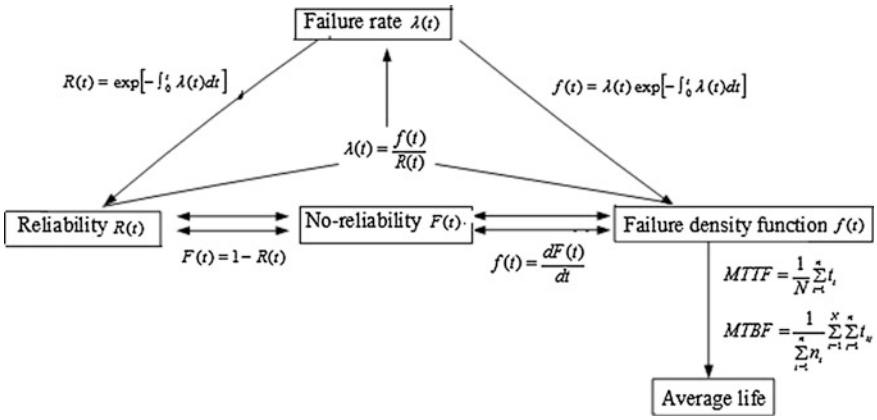


Fig. 60.2 The relationship between each reliability index

Table 60.4 MTBF of each subsystem

System	Frame	Spring device	Connection device	Wheel on the axle box	Driving mechanism	Basic of braking device
MTBF (Day)	17.2386	6.78672	25.4185	7.05325	11.2418	41.0411

The expression for failure rate function is

$$\lambda(t)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}=\frac{1.54270}{7.83827}\left(\frac{t}{7.83827}\right)^{0.54270} \tag{60.5}$$

In the formula, T represents the current no-trouble working time; the corresponding function graph of the characteristic value of reliability of wheel on axle box system is as shown in Fig. 60.3.

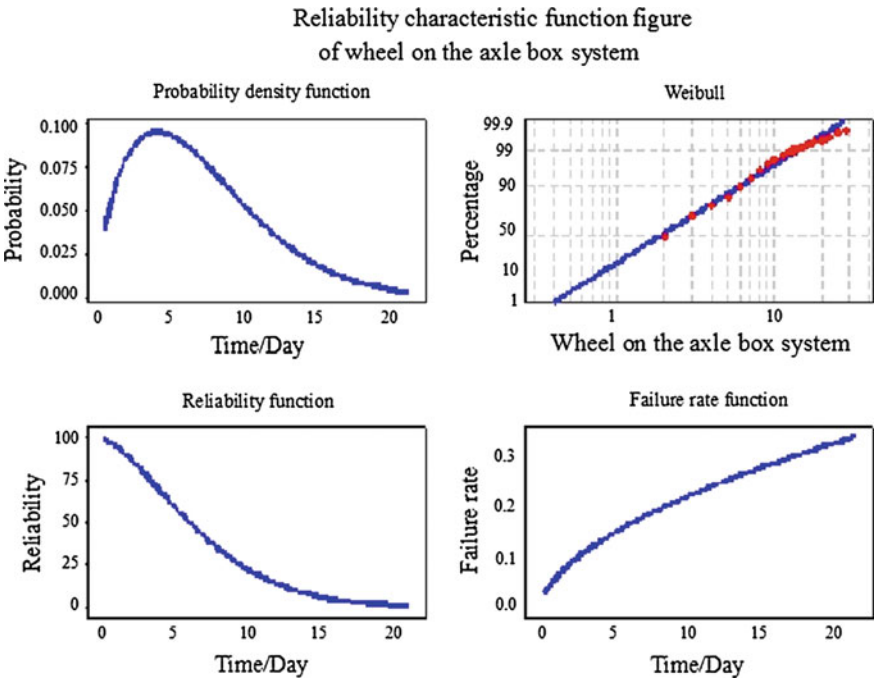


Fig. 60.3 Bogie wheel system reliability characteristic function figure

And so on, reliability characteristic value can be drawn from basic of braking device, connection device, frame, wheel on the axle box, spring mechanism of driving mechanism; thus, the system reliability contrast diagram is as shown in Fig. 60.4.

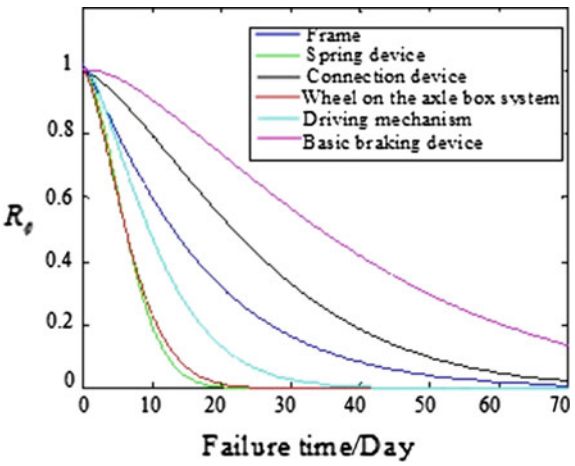


Fig. 60.4 Reliability comparison charts of bogie subsystems

According to the reliability, From high to low order is the basis braking device, connection device, frame, wheel on axle box of driving mechanism, and the spring mechanism, So the weakest links of bogie system are wheel on axle box and the spring mechanism, which should be focused on. The results show that the reliability of wheel on the axle box and spring device is relatively low.

60.3 Conclusions

The article bases on a field failure data of urban rail train bogie system as an example, the maximum likelihood estimation method in survival analysis theory is adopted to analyze reliability. Bogie Minitab software is used to determine the best failure distribution bogie model of the subsystems, and through the study of the comparison and analysis of the key components to determine the weak link of bogie system.

The research method of the reliability based on survival model analyzes the reliability of the parts only according to the using data of parts, which ignores the specific reason of the engine parts fault, at the same time the analysis results are completely driven by data,which aren't involved for hypothesis distribution model, so the analysis result is quite good. According to the determined distribution model, the failure distribution function, reliability function, and failure rate function can be calculated, and characteristic parameters such as reliability can be forecast through these functions; it is very important practical significance.

Simulation and experimental results show that the survival analysis theory can effectively solve the delete data loss under the mechanism of failure time interval uncertain problems, and we can get more reasonable and reliable result.

Acknowledgments This work has been supported by National High Technology Research and Development Plan (2011AA11A102), National Key Technology Research and Development Program (2015BAG01A01).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

1. Baek SH, Cho SS, Joo WS (2008) Fatigue life prediction based on the rainflow cycle counting method for the end beam of a freight car bogie. Int J Automot Technol 9(1):95–101
2. Ye ZS, Tang LC, Xu HY (2011) A distribution-based systems reliability model under extreme shocks and natural degradation. IEEE Trans Reliab 60(1):246–256
3. Barabadi A, Barabady J, Markeset T (2011) Maintainability analysis considering time-dependent and time-independent covariates. Reliab Eng Syst Safety 96(1):210–217

4. Zhou ZJ, Hu CH (2008) An effective hybrid approach based on grey and ARMA for forecasting gyro drift. Chaos, Solitons Fractals 35(3):525–529
5. Pham HT, Yang BS (2010) Estimation and forecasting of machine health condition using ARMA/GARCH model. Mech Syst Signal Process 24(2):546–558
6. Chen ZS, Yang YM, Hu Z et al (2006) Detecting and predicting early faults of complex rotating machinery based on cyclostationary time series model. J Vib Acoust 128(5):666–671
7. Huang R, Xi L, Li X et al (2007) Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mech Syst Signal Process 21(1):193–207
8. Barabadi A, Barabady J, Markeset T (2011) Maintainability analysis considering time-dependent and time-independent covariates. Reliab Eng Syst Safety 96(1):210–217
9. Volf P (2004) An application of nonparametric cox regression model in reliability analysis: a case study. Kybernetika 40(5):639–648
10. Jiao B, Lian Z, Gu X (2008) A dynamic inertia weight particle swarm optimization algorithm. Chaos, Solitons Fractals 37(3):698–705

Chapter 61
Simulation of Emergency Evacuation in Civil Aircraft Based on Multi-agent and Cellular Automata Technology

Tianchun Zou, Yuejuan Du, Dawei Chen, Haolei Mou and Kun Chen

Abstract Emergency evacuation has been and will continue to be an important research field of civil aircraft cabin safety. A typical civil aircraft cabin model and occupant model which meets the airworthiness regulation requirements are developed to simulate the emergency evacuation by using multi-agent and cellular automata theory. The situations of four different exit configurations are simulated for emergency evacuation. The emergency evacuation time and optimal performance statistic (OPS) are analyzed by simulating emergency evacuation to determine the optimal configuration. The simulation results show that the evacuation times are less than 90 s for the four different situations, and the evacuation time is shortest when the same side exits of fuselage section are opened, which has provided recommended principles for full-scale evacuation demonstration test, and the evacuation efficiency can be improved by designing the reasonable emergency evacuation procedures.

Keywords Multi-agent · Cellular automata · Emergency evacuation

T. Zou · Y. Du · H. Mou (✉) · K. Chen
Tianjin Key Laboratory of Civil Aircraft Airworthiness and Maintenance, Civil Aviation University of China, 2898, JinBei Rd, Dongli District, TianJin, China
e-mail: mhl589@163.com

T. Zou
e-mail: zoutianchun@126.com

K. Chen
e-mail: cknuaa@gmail.com

D. Chen
College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_61

61.1 Introduction

Currently, the safety issues of civil aircrafts are increasingly significant with the rapid development of civil aviation industry around the world. How to quickly and safely evacuate in an emergency from the aircraft has been and will continue to be the main research issue for Federal Aviation Administration (FAA), Civil Aviation Authority (CAA), airlines, and aircraft manufacturers. Therefore, the certification criteria and test methods to evaluate the performance of new aircrafts are given in Federal Aviation Regulation 25 (FAR25).

FAA and CAA have conducted full-scale evacuation demonstration tests within 90 s for each new type aircrafts. The A380 also must meet the emergency evacuation requirements within 90 s. In March 2006, the Airbus conducted a full-scale A380 emergency evacuation demonstration test with 873 participants. The full-scale demonstration was completed in 78 s with the supervision of airworthiness inspectors from FAA and European Aviation Safety Agency (EASA). The success of demonstration has played a vital role in getting A380-type certification from EASA and FAA [1]. However, there are several limitations related to emergency evacuation demonstration. Firstly, full-scale demonstrations are costly and expose participants to significant hazards. Secondly, certification results may provide little useful suggestions for evacuation procedures, since evacuation scenario in real emergency may be very different as in the certification trials. What's more, only one single full-scale certification trial is conducted by certification requirements and it is impossible to represent the real capacity of evacuation and provides enough information to design the effective cabin configuration.

Computer evacuation simulation models can be performed repeatedly and easily and reproduce evacuation performance of aircraft in full-scale certification trials. Evacuation simulations using computer models has been developed for at least 40 years [2]. The first simulation model to address aircraft emergency evacuation issues was that of FAA in the early 1970s known as the GPSS model [3]. What's more, numbers of aircraft evacuation models have been reported in the literature, such as Gourary Associates model (GA model) [4], discrete element method (DEM) [5], ARCEVAC [6], STRATVAC [7], VacateAir [8], and airEXODUS [9]. In these models, GPSS, GA, and DEM models are mainly concentrated on simulating passengers' movement. ARCEVAC, STRATVAC, VacateAir, and airEXODUS are considered to incorporate human factors of passengers and capable to predict different human behaviors adding to their movement.

The emergency evacuation simulation model is established based on multi-agents and cellular automata theory. In this paper, the situations of four different exit configurations are simulated for the emergency evacuation to determine the optimal configuration of the available exit arrangement and provide suggestions for full-scale evacuation demonstration tests of airplanes.

61.2 Emergency Evacuation Model

61.2.1 Occupant Properties

FAR25 Appendix J "Emergency Evacuation" regulated that a representative passenger load of persons in normal health must be used as follows:

(1) At least 40 % of the passenger load must be female.
(2) At least 35 % of the passenger load must be over 50 years of age.
(3) At least 15 % of the passenger load must be female and over 50 years of age.
(4) Three life-size dolls, not included as part of the total passenger load, must be carried by passengers to simulate live infants 2 years old or younger.

According to the above requirements, Table 61.1 shows the typical population distribution of evacuation demonstrations.

61.2.2 Civil Aircraft Model

The model uses meshing mechanism, whose size is 0.5 m × 0.5 m. The cabin seats, partitions, kitchen, bathroom, etc., are simplified into 0.5 m × 0.5 m square grids. In order to carry out an emergency evacuation simulation analysis, the model must be simplified, and the physical scene discrete model can be used in the scene. Considering the impact of the evacuation slide, boarding doors, and service doors, they should reduce to a grid, where there can be through only one person at the same time; what's more, emergency exits and aisle widths are reduced to a grid width consistent with passengers' grid size. Emergency exits at the front and rear wings that have accessible aisle width are simplified into two square grids. The front–rear direction of seat is simplified into two grids and lateral direction into one grid. Figure 61.1 shows the cabin configuration of aircraft.

The traditional four discrete grid accuracies cannot meet the requirements because the aircraft cabin is a small confined space. Therefore, the traditional discrete grid needs add a factor to enlarge or narrow the cabin configuration.

Table 61.1 Population distribution of evacuation demonstrations

Gender	Proportion	Ages	Proportion
Female	≥40	18–49	$(N_{\mathrm{F}} - N_{\mathrm{F50–60}})/N$
		50–60	≥15
Male	≤60	18–49	$(N_{\mathrm{M}} - N_{\mathrm{M50–60}})/N$
		50–60	≥20
Total	100		100

Note N_F Female number; $N_{F50–60}$ Female must be over 50 years of old; N_M Male number; $N_{M50–60}$ Male must be over 50 years of old; N Total number

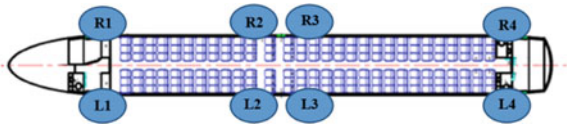


Fig. 61.1 Cabin configuration of aircraft

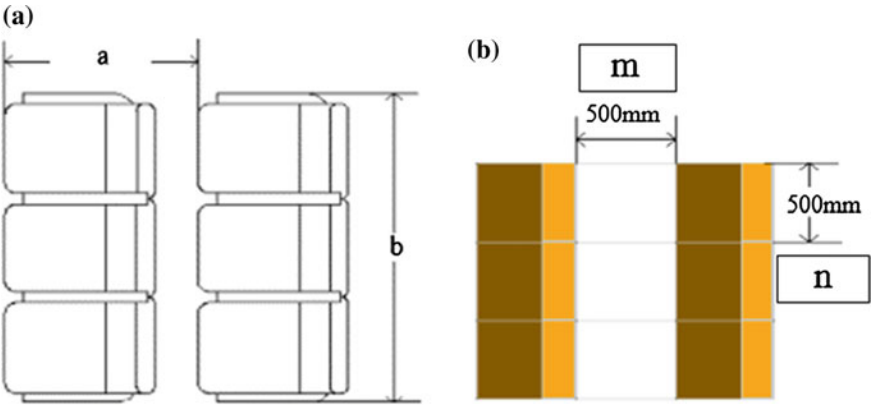


Fig. 61.2 Discrete element method. **a** Seat arrangement. **b** Discrete cellular

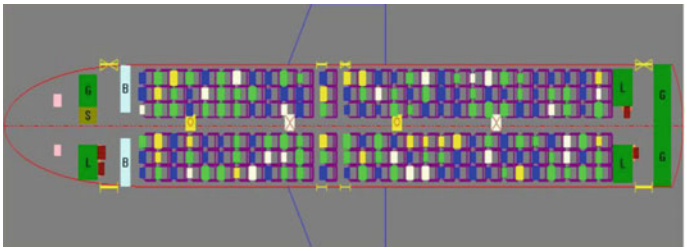


Fig. 61.3 The simplify scene of cabin layout

Various factors (aisle, seats, etc.) can be discrete into 0.5 m × 0.5 m in aircraft cabin and add the scaling factor in both directions.

For example, as shown in Fig. 61.2, the passenger seat pitch is a mm; seat rows are in a 3-3 configuration which width is *b* mm, each row of seats can be discrete as 2 × 3 in cellular space, and each grid scaling coefficient is $a/(2 \times 500) = m$ and $b/(3 \times 500) = n$. Figure 61.3 shows a simplified scenario cabin configuration, a total of 174 people.

61.3 Emergency Evacuation Simulation Analysis

61.3.1 Emergency Evacuation Performance Key Parameters

This study uses total evacuation time (TET) and optimal performance statistic (OPS) to compare and analyze the evacuation performance for the four different conditions.

The OPS can be calculated for each evacuation, providing a measure of the degree of performance. It is developed by the Fire Safety Engineering Group, University of Greenwich (FSEG), and the OPS is defined as follows [10]:

$$\mathrm{OPS} = \frac{\sum_{i=1}^{n} (\mathrm{TET} - \mathrm{EET}_i)}{(n-1) \times \mathrm{TET}} \tag{61.1}$$

Note:

i	evacuation exit,
n	number of exits used in the evacuation,
EET_i	evacuation time (last passenger) of exit i (s),
TET	total evacuation time (s), i.e., MAX [EET_i].

Quadrant1: optimal evacuation with TET below 90 s;
Quadrant 2: optimal evacuation with TET above 90 s;
Quadrant 3: Suboptimal evacuation with TET below 90 s;
Quadrant 4: Suboptimal evacuation with TET above 90 s.

An aircraft will be marked by very low values of OPS. However, selecting an acceptable value for OPS is somewhat arbitrary. The FSEG considers OPS values of 0.1 or less as being optimal. It is informative to plot the TET for an evacuation simulation against the OPS, producing the plot of aircraft evacuation performance similar to Fig. 61.4 [11].

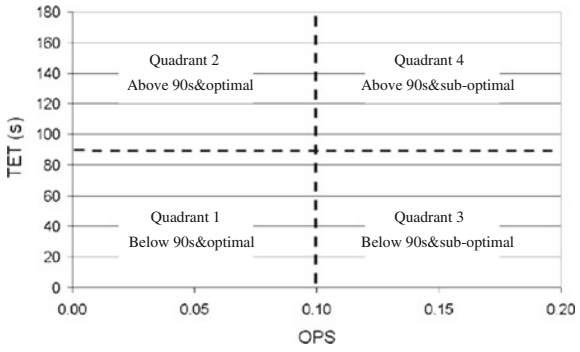


Fig. 61.4 The four quadrants of evacuation time and performance level

61.3.2 *Influence of Four Different Exit Configurations*

Exits used in the demonstration must consist of one exit from each exit pair. What's more, the demonstration utilizes a population which satisfies an age and gender mix specified in the FAR rules, selected by the manufacturer and approved by the regulatory authority. Thus, four different exit configurations are available to study the effects on emergency evacuation.

1. Case1: open the exits of L1, L2, L3, and L4

As shown in Fig. 61.5, this condition simultaneously opens the side exits. Figure 61.6a and Table 61.2 show that the mean TET is 84.42 s (1000 simulations) under this condition. The result is consistent with emergency evacuation requirements of FAR25.

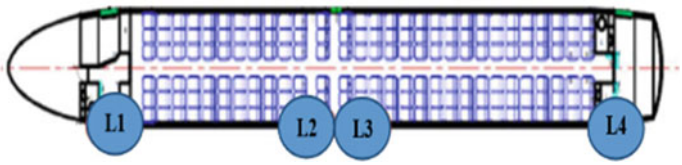


Fig. 61.5 Exits of L1, L2, L4, and R3

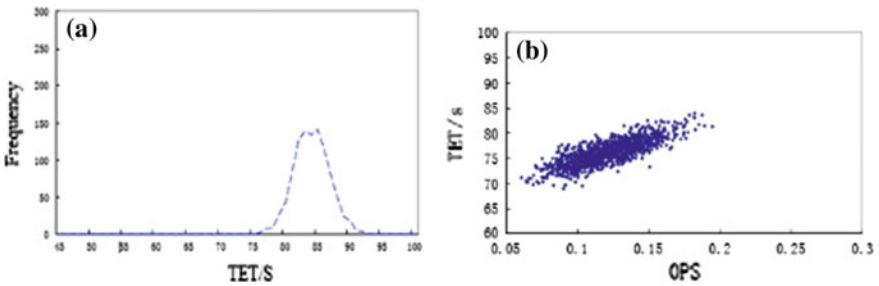


Fig. 61.6 Simulation results (1000 simulations). **a** Frequency distribution of TET. **b** Scatter diagram of TET and OPS

Table 61.2 Simulation results

	T_{Total} (s)	$T_{\text{Exit Prep}}$ (s)	$T_{\text{Exit Flow}}$ (s)
Min	76.64	10.42	66.22
Mean	84.42	13.87	70.55
Max	92.64	16.13	76.51
STD	2.66	0.95	

Note $T_{\text{Total}} = T_{\text{Exit Prep}} + T_{\text{Exit Flow}}$; T_{Total} Total evacuation time for the exit; $T_{\text{Exit Prep}}$ Time for exit preparation; $T_{\text{Exit Flow}}$ Time of exit flow

Figure 61.6b shows that the evacuation points fall in Quadrant 1 and Quadrant 3. Evacuation points falling within the Quadrant 1 indicate that the evacuation is optimal and the TET is less than 90 s. These aircrafts have performed well and represented a good evacuation procedure. Quadrant 3 shows that the evacuation is suboptimal yet the TET is less than 90 s. These aircrafts can generate acceptable evacuation times even when the evacuation is suboptimal. Figure 61.7 shows evacuation envelope, and the evacuation efficiency can be further improved by increasing the ability to command the crew.

2. Case 2: open the exits of L1, L2, L4, and R3

As shown in Fig. 61.8, these simulation scenarios simultaneously open the exits L1, L2, L4, and R3. Figure 61.9a shows that the mean TET is 88.33 s (1000 simulations). The result is consistent with emergency evacuation airworthiness requirements of FAR25. Figure 61.9b shows the data points fall in Quadrant 3. The results are suboptimal, and total evacuation time is less than 90 s. Even if the evacuation process is not very satisfactory, the evacuation time was less than 90 s and met the evacuation airworthiness requirements. If evacuation procedures were reasonably designed, total evacuation time can also be reduced.

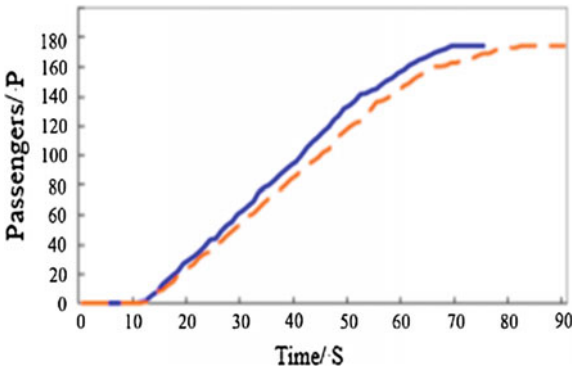


Fig. 61.7 Evacuation envelope (1000 simulations)

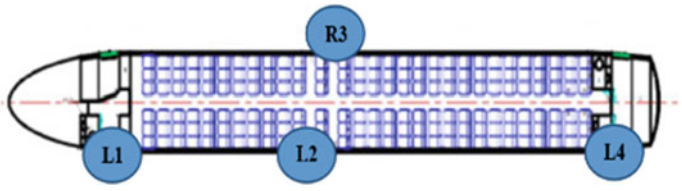


Fig. 61.8 Exits of L1, L2, L4, and R3

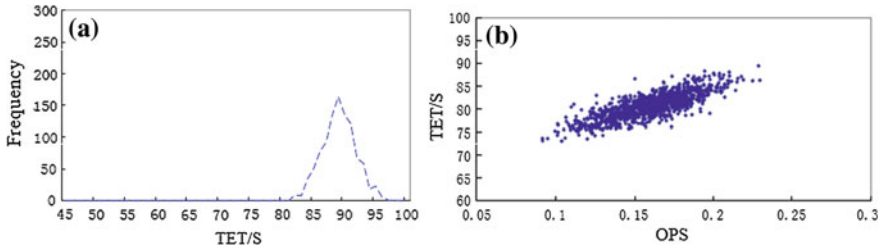


Fig. 61.9 Simulation results (1000 simulations). **a** Frequency distribution of TET. **b** Scatter diagram of TET and OPS

3. Case 3: open the exits of L1, L2, L3, and R4

As shown in Fig. 61.10, these simulation scenarios simultaneously open the same side of emergency exits L1, L2, and L3, and the rear service door R4. Figure 61.11a shows that the mean TET is 88.33 s (1000 simulations). The result is consistent with emergency evacuation airworthiness requirements of FAR25. Figure 61.11b shows the data points fall in Quadrant 3. The results are suboptimal, and the total evacuation time is less than 90 s. Even if the evacuation process is not very satisfactory, the evacuation time was less than 90 s and meets the evacuation airworthiness requirements. If evacuation procedures are reasonably designed, total evacuation time can also be reduced.

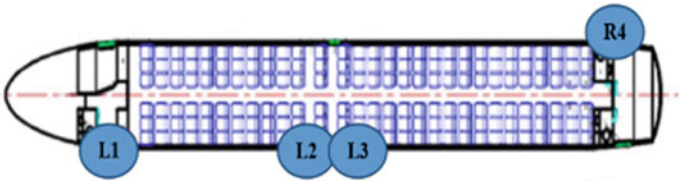


Fig. 61.10 Exits of L1, L2, L3, and R4

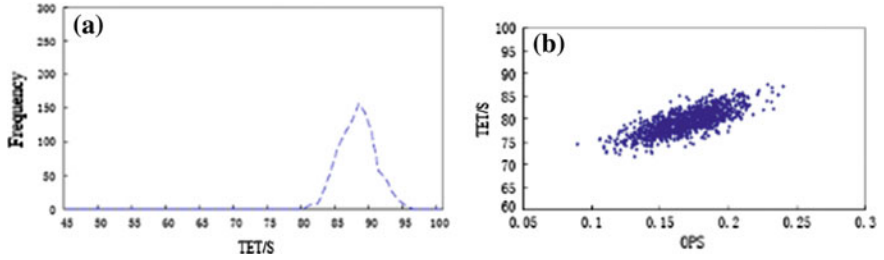


Fig. 61.11 Simulation results (1000 simulations). **a** Frequency distribution of TET. **b** Scatter diagram of TET and OPS

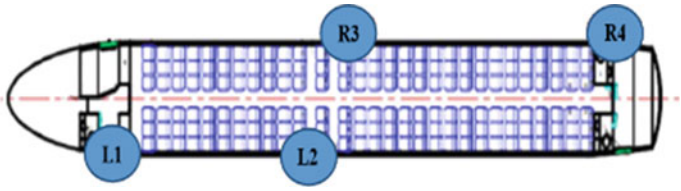


Fig. 61.12 Exits of L1, L2, R3, and R4

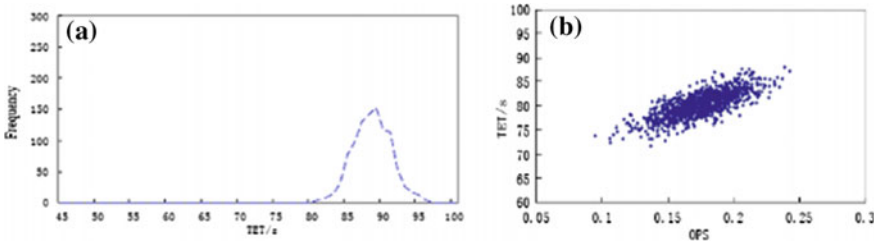


Fig. 61.13 Simulation results (1000 simulations). **a** Frequency distribution of TET. **b** Scatter diagram of TET and OPS

4. Case 4: open the exits of L1, L2, R3, and R4

As shown in Fig. 61.12, under this exit configuration, Fig. 61.13a shows that the mean TET is 88.97 s (1000 simulations). The result is consistent with emergency evacuation requirements of FAR25.

As shown in Fig. 61.13b, the data points fall in Quadrant 3. The results are suboptimal, and the total evacuation time is less than 90 s. Even if the evacuation process is not very satisfactory, the evacuation time is less than 90 s and meets the evacuation airworthiness requirements. If evacuation procedures are reasonably designed, total evacuation time can also be reduced.

61.4 Conclusion

According to FAR25 airworthiness requirements relating to emergency evacuation, the civil aircraft cabin emergency evacuation models are built based on multi-agent and cellular automata theory, and four possible evacuation exit configurations were simulated for full-scale demonstration test. The results show that:

(1) There are four different ways to open the emergency evacuation exits, which significantly affect the evacuation time. However, the emergency evacuation simulation of average total evacuation time is less than 90 s, which satisfy the

requirements of airworthiness regulations relating to emergency evacuation in theory.

(2) There are significant differences among the four exit configurations. Most of the evacuation simulation data points fall in Quadrant 1 when selected to the same side exits. The total evacuation time is shortest. Therefore, it is recommended that manufacturers may conduct full-scale evacuation demonstrations using this exit configuration.

(3) The evacuation efficiency can be improved through reasonably designing emergency evacuation procedures and enhancing the capabilities of crew command and passenger performance.

Acknowledgment The authors acknowledge the supports from the Fundamental Research Funds for Central Universities (NO. 3122014B006) and the Fund of Tianjin Key Laboratory of Civil Aircraft Airworthiness and Maintenance in CAUC (Civil Aviation University of China).

References

1. Daly K (2006) A380: everyone off in time. Flight Int 4:4–10
2. Gwynne S, Galea ER, Owen M, Lawrence PJ, Filippidis L (1999) A review of methodologies used in evacuation modeling. Fire Mater 23:383–388
3. Garner JD (1978) GPSS computer simulation of aircraft passenger emergency evacuations, U. S. Department of Transport, Federal Aviation Administration, DOT/FAA/AM-78-23
4. Gourary Associates (1990) Demonstration manual for microcomputer model, User documentation for Gourary associates evacuation model version 6.1
5. Robbins CR, Mckee S (2001) Simulating the evacuation of a commercial airliner. Aeronaut J 105:323–328
6. Aviation Research Corporation (1994) The development of ARCEVAC an emergency evacuation simulation model: Beta version 1.0, Draft Report by Aviation Research Corporation
7. Cagliostro DE (1984) A user-operated model to study strategy in aircraft evacuation. J Aircraft 21:962–965
8. Xue Z, Bloebaum CL (2008) Experimental design based aircraft egress analysis using VacateAir an aircraft evacuation simulation model. In: 12th AIAA/ISSMO multidisciplinary analysis and optimization conference, pp 1–23
9. Galea ER, Blake SJ, Lawrence PJ (2002) The airEXODUS evacuation model and its application to aircraft safety. In: Fire safety and survivability, applied vehicle technology (AVT) panel symposium, NATO Research and Technology Organization, Aalborg, Denmark, pp 23–26
10. Owen M, Galea ER, Lawrence PJ, Filippidis L (1998) The numerical simulation of aircraft evacuation and its application to aircraft design and certification. Aeronaut J Royal Aeronaut Soc 102:301–312
11. Galea ER, Blake SJ, Lawrence PJ (2005) Report on the testing and systematic evaluation of the airEXODUS aircraft evacuation model, CAA Paper 2004/05

Chapter 62
Research and Development of Efficient Refrigerant Transducer Module

Lin Yang, Dingfeng Wu, Yanyong Li, Xingping Liu and Siyu Liu

Abstract Main technical parameters, power circuit, general structure and characteristics, refrigerant cooling methods, and operating principles of converter module were introduced. United tests and operation tests with commercial air-conditioning system show that the design of the converter module satisfied the application requirements.

Keywords Power circuit · Refrigerant cooling · Converter module control

62.1 Introduction

With the national requirements of energy conservation, emission reduction, and low carbon as well as the development of converting technologies, at present, most industrial drive converting technologies are switched into AC drive from DC drive to improve the efficiency, reliability, and maintainability of equipment and reduce the later maintenance cost. As one of the core components of the converter, the designed performances of the converter module directly determine the operating efficiency and reliability of the converter and its system performances as well as effective utilization rate of parameters of various devices, and thus, the requirements

L. Yang (✉) · D. Wu (✉) · Y. Li (✉) · X. Liu (✉) · S. Liu (✉)
Beijing, China
e-mail: yangling@teg.cn

D. Wu
e-mail: wudf@teg.cn

Y. Li
e-mail: liyy@teg.cn

X. Liu
e-mail: liuxp2@teg.cn

S. Liu
e-mail: liusy6@teg.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_62

for designing the converter module with excellent comprehensive performances are very strict.

Aiming at the comprehensive conditions, in this article, how to design a refrigerant-based converter module is proposed. The main circuit of the module is a AC–DC–AC topology structure; a rectifier and an inverter adopt insulated gate bipolar translator (IGBT) as the switching elements and use the four-quadrant rectifying mode to improve the intermediate voltage, thus improving the output voltage; the advantage is that under the same output power, the higher the output voltage, the smaller the output current, so the losses of motors and lines are easy to reduce, the motor volume is reduced, and the sectional areas of conductors are reduced to save the cost; in addition, refrigerant is adopted as the cooling medium, the heat produced when elements work is dissipated through evaporated latent heat, and the temperatures of radiators are controlled by controlling the refrigerant flow, so that the radiators can be controlled with in a lower temperature range, and then the rated parameters of the elements are used highly, and the output capacities of the devices are further improved to further improve the operational reliability while achieving high cost performance.

62.2 Main Circuit and Main Parameters of Transducer Module

See Table 62.1.

62.2.1 Main Circuit of Module

The main circuit of the module uses a typical voltage type AC–DC–AC circuit topology structure (see Fig. 62.1). For the first part, namely the AC–DC part, 6

Table 62.1 The main parameters of module

Parameters	Value
Rated input voltage	3AC 380 V (±10 %)
Input frequency	50 ± 3HZ
Rated input current	3AC 404 A
Maximum input current	3AC 450 A
Intermediate voltage	650 V
Output voltage	0–380 V
Rated output current	3AC 395 A
Rated output power	260 kW
Switching frequency	3000 HZ
Cooling method	Cooling through refrigerant
Cooling medium	R134a

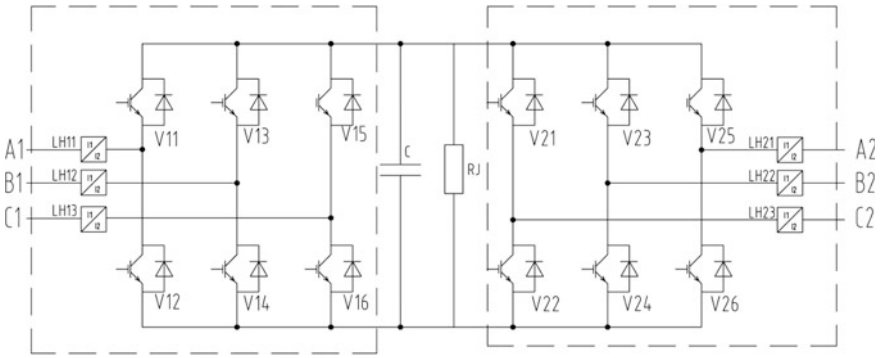


Fig. 62.1 Main circuit diagram of module

IGBTs form a four-quadrant rectifying bridge; because the IGBT power module can realize the bidirectional flow function of energy, IGBTs are adopted as the rectifying devices; through the proper pulse-width modulation (PWM) mode, not only the output DC voltage of the PWM converter can be controlled, but also the size and phase of the network-side AC current of the converter can be controlled so as to approximate the sine wave and be in same or inverse with the phase of the grid voltage; thus, the power factor of the system approximates ±1 [1]. By using the bidirectional flow function, current is forcibly changed, the phases of the input current and the input voltage are the same, and the power factors are improved; by regulating the PWM control pulse, on the one hand, the input power factors can be regulated, and pollution of power harmonic waves can be reduced; on the other hand, the intermediate DC voltage can be increased through a front-end input electric reactor, and the intermediate voltage is more stable through the constant-voltage regulation; because the intermediate voltage is improved, the output voltage can be improved and high-quality output current can be obtained; for the latter part, namely the DC–AC part, 6 IGBTs form a three-phase inverter circuit; the voltage, current, and frequency of the output power supply are regulated by using a PWM control pulse method, so as to control the output properties of load. LH11–LH23 is a current sensor, and it is used for collecting the input and output current signals; the current signal is transmitted to the main control unit to control the PWM pulse by treating the data; in addition, this signal is further used for judging overcurrent and short circuit. RJ stands for discharge resistance and is used for discharging the energy stored in the capacitor on the module to prevent electric shocks produced when operators maintain the converter after power off and to protect personal safety.

The main circuit is shown in Fig. 62.1.

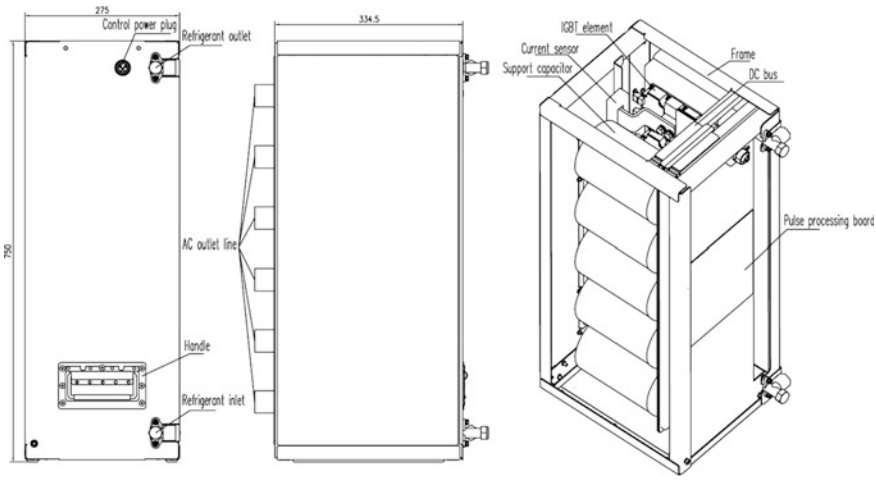


Fig. 62.2 Overall structure of module

62.2.2 General Structure of Transducer Module

As shown in Fig. 62.2, the designed vertical book-shaped structure of the transducer module is installed in a cabinet, the outline dimension of which is 275 mm × 334.5 mm × 750 mm (width × depth × height); the electrical port of the main circuit is located behind the module, and an auxiliary power supply plug is located on the middle part of the front panel of the module; handles are arranged at the upper and lower sides of the panel of the module to conveniently install. The module is cooled through refrigerant, the cooling medium R134a of which enters from the right lower side of the module and flows outside from the right upper side of the module.

62.3 Model Selection of Key Devices

62.3.1 Model Selection of Power Devices (IGBT)

The IGBTs are adopted as the core switches of the converter, whose operational reliability must be guaranteed; meanwhile, the cost ratio of the devices inside the converter is largest, so that rational selection of the IGBT parameters is critical.

Selection for voltage rating calculation of element is:

$$V_{CE} = (U_{dcmax} \times K_1 + \Delta U) \times K_2 = (650 \times 1.2 + 200) \times 1.1 = 1078\,(\mathrm{V}) \quad (62.1)$$

In the equation, $U_{\mathrm{dc\,max}}$ refers to the highest voltage of DC link; K_1 refers to the voltage margin coefficient; K_2 refers to the safety coefficient; and ΔU refers to the switch overvoltage.

Selection for current rating calculation of element is:

$$I_{\mathrm{C}} = I_{\mathrm{AC}} \times \sqrt{2} \times K_3 = 395 \times 1.414 \times 1.3 = 726\,(\mathrm{A}) \tag{62.2}$$

In the equation, I_{AC} refers to the output current of the inverter (effective value); K_3 refers to the current peak coefficient.

The IGBT elements of 1200 V/900 A are selected through the rating calculations of voltage and current of the IGBT element.

62.3.2 Selection and Usage of Support Capacitor

When the inverter works, ripple current will be produced at the switching frequency. The size of the ripple current is related to the total power of the inverter, and $\mathrm{d}i/\mathrm{d}t$ of IGBTs produced in the switching process can sense the high peak voltage and inductance of the DC circuit needed to be reduced. Therefore, the support capacitor shall have enough capacity and lower stray inductance.

The capacity of the support capacitor is calculated according to Eq. (62.3):

$$C = \frac{10P}{\omega U_{\mathrm{d}}^2} = \frac{10 \times 240 \times 10^3}{3 \times 3.14 \times 50 \times 650^2} = 12\,(\mathrm{mF}) \tag{62.3}$$

In the equation, P refers to the rated output power; ω refers to the frequency coefficient; U_{d} refers to the DC voltage.

62.3.3 Fixed Discharging Resistor

The main circuit of the inverter is provided with a fixed discharging circuit. When the main circuit does not quickly discharge because of faults, the energy of the capacitor can be consumed through the fixed discharging circuit. To guarantee the safety of operation personnel, the DC-link voltage must be reduced below the safe working voltage (36 V) within 15 min after the inverter is shut down, and see Eq. (62.4) for calculation method.

$$t = \tau \times \ln\frac{U_{\mathrm{d}}}{U_{\mathrm{c}}} = R \times C \times \ln\frac{U_{\mathrm{d}}}{U_{\mathrm{c}}} = R \times 12 \times 10^{-3} \times \ln\frac{650}{36} \leq 900\,(\mathrm{s}) \tag{62.4}$$

By calculating according to Eq. (62.4), we can see $R \leq 26\ \mathrm{K\Omega}$.

62.4 Cooling Technology Characteristics of Transducer Module

62.4.1 Cooling of Converter Module

The converter module is cooled by using a new refrigerant cooling method, and the cooling medium is R134a. The realization method of the radiators comprises the following steps: A copper tube is inserted in an aluminum substrate, then the IGBT elements are installed on the surface of the aluminum substrate, and the cooling medium R134a flows inside the copper tube which is arranged in the aluminum substrate, then the cooling medium is evaporated on the heating parts of the radiators, and the heat produced when the IGBTs work is dissipated through the evaporated latent heat.

62.4.1.1 Property Introduction to Refrigerant R314a

R134a: Freon R134a is a new refrigerant and belongs to hydrofluorocarbons ((hereafter referred to as HFC). The boiling point of R134a is −26.5 °C. The ozone depletion potential (ODP) value is 0, so that R134a does not destroy the ozone layer in air, and it is an environment-friendly refrigerant promoted in recent years; at present, R134a is applied to refrigerators, freezers, and car air-conditioning systems (Table 62.2).

62.4.1.2 Cooling Principle of Refrigerant

This cooling scheme uses the common cooling principle of an air-conditioning unit as shown in Fig. 62.3.

Table 62.2 Main physicochemical properties of HFC-134a

Parameters	Value
Molecular formula	CH_2FCF_3
Boiling point (101.3 kpa)	−26.1 °C
Critical temperature	101.1 °C
Critical pressure	4066.6 kPa
Liquid density	1188.1 kg/m3
Saturated vapor pressure (25 °C)	661.9 kPa
Evaporation heat/evaporation latent heat (1 atm)	216 kJ/kg
Ozone depletion potential value	0 (ODP)
ASHRAE security level	A1 (non-toxic incombustible)
Saturated liquid density 25 °C	1.207 g/m^3
Specific heat of liquid 25 °C	1.51 [KJ/(Kg °C)]
Solubility (in water, 25 °C)	0.15 %

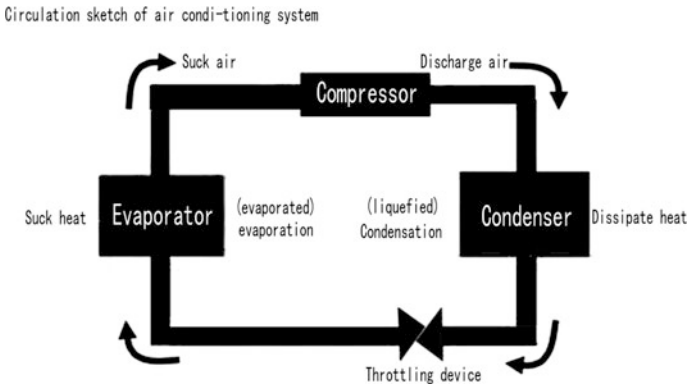


Fig. 62.3 Cooling principle of air-conditioning unit

The radiators of the transducer are equal to an evaporator in Fig. 62.3, and the operating process is as follows: A compressor presses the air flowing from the evaporator (the radiators of the converter module) into a high-pressure condenser, and then the air is switched to liquid state through the heat dissipated from the condenser (actually it is gas–liquid state), the liquid cooling medium enters the evaporator (the radiators of the transducer module) through a throttling device, then the liquid cooling medium is evaporated on the heating part, and then the heat produced when the transducer module works is dissipated through the evaporated latent heat.

The specific realization and calculation method for refrigerant cooling of the transducer module is detailed below.

Figure 62.4 shows the pressure–enthalpy diagram of R134a. The vertical coordinate of the diagram refers to the value corresponding to absolute pressure, and the horizontal ordinate refers to the specific enthalpy value h.

The cooling and phase-changing process is mainly carried out in the wet steam area in Fig. 62.5, and the wet steam area means the coexistence area of gas and liquid and a mixed area of liquid and gaseous R134a; the R134a in this area is in saturation state; the relationship between pressure and temperature is a one-to-one correspondence; and the higher the pressure, the higher the temperature.

When the transducer module works in rated working condition, the maximum total loss is 6 kW; according to the parameters of the transducer system, the pressure of the refrigerant entering the transducer module is 900 kPa; from Fig. 62.4, the properties of R134a liquid are shown in Table 62.3.

When the transducer module works in rated working condition, the refrigerant V required is:

$$V = \frac{P}{\Delta H} = \frac{6\,\text{kW}}{(408 - 250)\,\text{kJ/kg}} = 37.98\,\text{g/s} = 1.8\,\text{L/min} \tag{62.5}$$

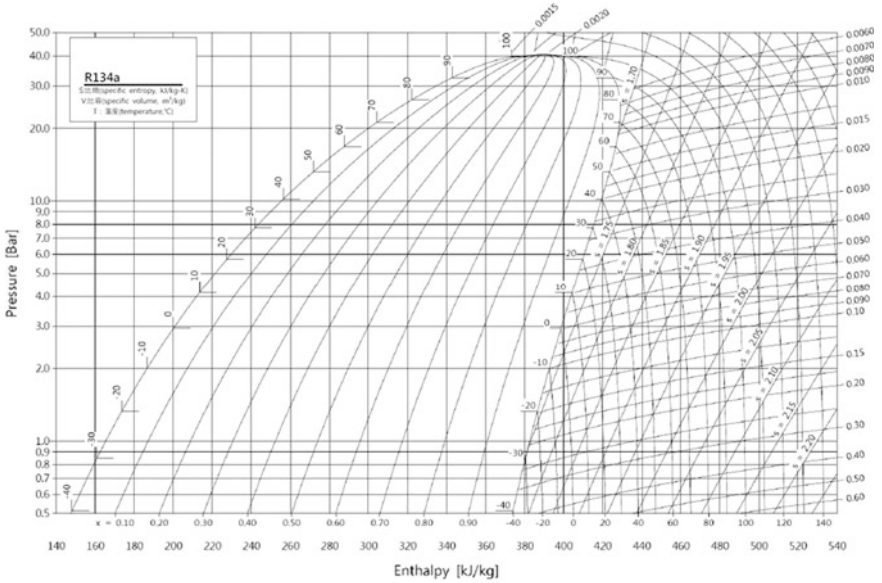


Fig. 62.4 R134a pressure–enthalpy diagram

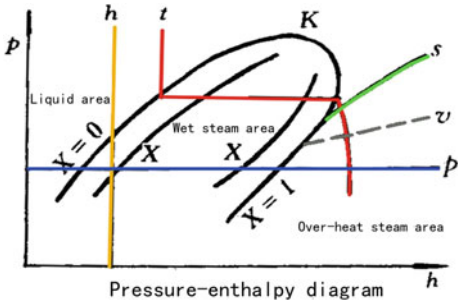


Fig. 62.5 Breakdown drawing of pressure–enthalpy diagram

Table 62.3 The properties of R134a liquid

Temperature (°C)	Enthalpy value (kJ/kg)	Pressure (kPa)
36 (gas state)	408	900
36 (liquid state)	250	900

In the Eq. (62.5), P refers to the radiating power; ΔH refers to the enthalpy difference between the former and later refrigerants.

Namely, when the transducer module works in rated working condition, the cooling requirements of the transducer module can be met when the refrigerant flow required is 1.8 L/min. In addition, because load is not a fixed value, the input quantity of the refrigerant can be controlled by a manner of regulating the openness

of the throttling device through detecting the current temperature of the radiator via the temperature sensor arranged on the radiator, so as to control the temperature of the radiator.

62.4.1.3 Cooling Test Results of Refrigerant

When the transducer module works in rated working condition, 6 thermocouples are arranged on the module, and the temperature result of the module is shown in Fig. 62.6.

From Fig. 62.7, the highest temperature of the transducer module is 40°, the temperature difference among the 6 temperature points is within 10 K, namely the

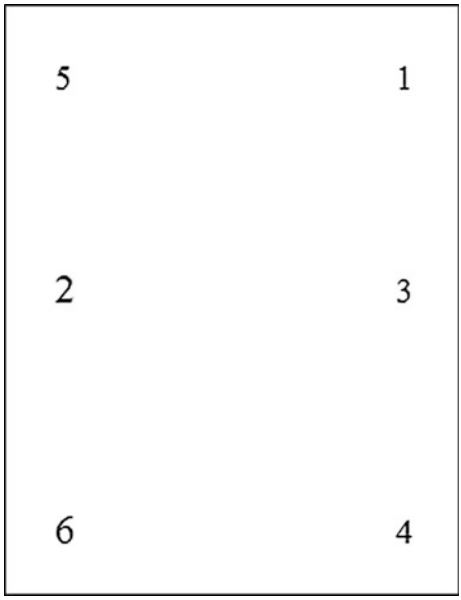


Fig. 62.6 Temperature measuring points of module

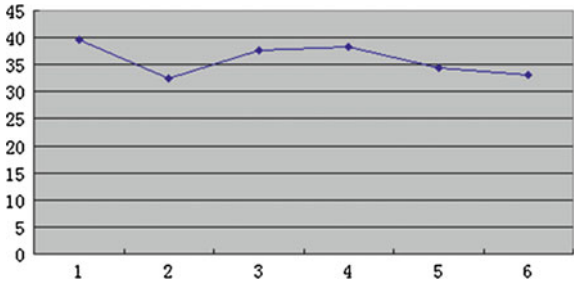


Fig. 62.7 The temperature distribution of module under rated output power

temperature uniformity of the module is very good, and experimental verification shows that the refrigerant cooling method completely meets the cooling requirements of the transducer module.

62.5 Electrical Structure of Module

The module has a compact structure. 12 IGBT elements that are arranged in one line along the up–down direction, a refrigerant radiator, a gate driving unit, a pulse distributing unit, a low-sensitive bus bar, and a support capacitor are integrated inside the module (Fig. 62.8).

The converter module totally has 12 IGBT elements to form a group of four-quadrant rectifying units and a group of inverting units. From the structure, the module has 4 layers, the lowest layer is a radiator layer, the second layer is a the IGBT element layer, crimped with the radiators through bolts; the third layer is a bus layer, mainly used for connecting the main circuit, and the fourth layer, namely the top layer, is a support capacitor layer.

Because of rational layout, the structure of the module is very compact and the power density is large. The IGBTs are connected with the capacitor through a shortest low-sensitive bus bar, so the IGBTs can produce high peak voltage in the

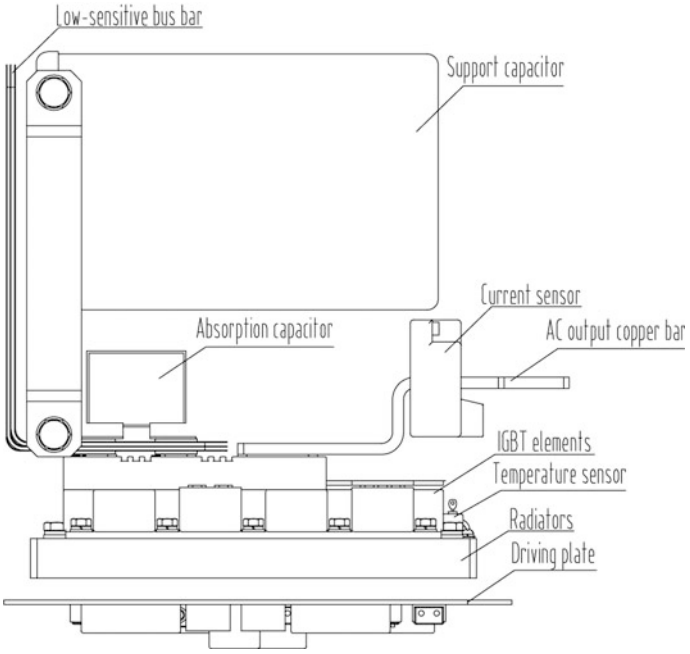


Fig. 62.8 Module IGBT distribution

fast switching process. The size of the peak voltage is related to di/dt and stray inductance in the circuit: If the turnoff of the element is reduced, the turnoff loss of the element is increased through di/dt, the efficiency of the module is reduced, and radiating load is increased. Therefore, reduction of the stray inductance in the circuit becomes the best method for overcoming the peak voltage. In the module, the connection between the IGBT elements and the support capacitor and the connection among the support capacitors are realized through the low-sensitive bus bar, so that the stray inductances on lines are greatly reduced, absorption circuits are saved, and the circuit is simpler and more reliable.

The signal transmission between the pulse distributing unit and the gate driving unit is realized through optical fibers, so that not only the high-voltage insulation problem is solved, but also the capacity of resisting disturbance of the transducer module is improved. Meanwhile, the control unit that is adopted as the lowest layer has function of protecting overcurrent and short circuit of element, judging and timely treating the abnormal situations of IGBT within the shortest time, so that the safety of the module can be guaranteed in the extreme case of load abnormity.

62.6 Installation Test of Transducer Module

This module is required to be installed in the transducer; various performance tests are completed according to the basic requirements of the transducer and the unit; temperature rise test, fast frequency rise test, fast frequency reduction test, rated working condition test, and the maximum working condition test are carried out in turn, the 10 % rise and falling fluctuation test are carried out for the grid voltage in various working conditions, and the test results meet the requirements. These tests verify that the module can stabilize the 450 A input current in the condition of AC input 380 V. The overall radiating performances and temperature uniformity of the module are good. IGBTs can be turned off through short circuit and overcurrent protection within the safety time, fault information can be reported to the main controller, and the main controller executes other protections after receiving the fault signals, including disconnecting the power supply of the main circuit.

62.7 Conclusion

IGBT converter module is one of the core components of the converter, and whether the convener module is rationally designed directly influences the reliability, expandability, maintainability, and price of the converter. The module greatly improves the utilization rate of the rated parameters of the element by using the refrigerant cooling methods; thus, it controls the temperature of the radiator within some temperature range and further controls the junction temperature of IGBT so as to improve the operational reliability of the module; because of lower temperature

control, the power density of the module is larger, and the volume is only 1/4 of an air-cooled module. We have held the fully independent intellectual property of the cooling method in power and electronic radiator series. Compared with the traditional transducer, the weight of the transducer provided with this converter module is reduced by 30 %, the volume is reduced by 50 %, and power is saved by 40 % [2], so that applications of the refrigerant cooling method on the power and electronic technologies in China are filled. At present, this transducer module has successfully operated in the transducers on central air-conditioning units, and better social and economic benefits are achieved.

References

1. Liu C (2003) Design of 4-quardant converter for electric locomotive auxiliary system. Electric Locomot Mass Transit Veh 26(6):18–20. (in Chinese)
2. Wang T (2015) Research and development of commercial central air-conditioning inverter based on PWM rectifier. High Power Conv Technol 1:56–65 (in Chinese)

Chapter 63
Minimizing Distribution Time for One-to-Many File Distribution in P2P Networks

Hongyun Zheng and Hongfang Yu

Abstract This paper addresses the problem of minimizing one-to-many peer-to-peer (P2P) file distribution time. We formulate the problem as a 0–1 integer linear program (ILP), with the bottleneck being the process capacity of node, propose a heuristic solution, and evaluate its performance. Simulation results show that the solution approaches the optimal theoretical bound in small-scale case and outperforms BitTorrent-like (BT-like) scheme.

Keywords P2P · One-to-many file distribution · Minimum distribution time

63.1 Introduction

In peer-to-peer (P2P) networks, each node can act as both a server and a client, which makes it scalable to distribute large files [1]. Many popular applications, such as PPlive [2], PPstream [3], and BitTorrent [4], distribute content overlay on P2P networks.

In this paper, we consider one-to-many file distribution in P2P networks. One-to-many file distribution refers to delivering file from a node (seeder) to multiple other nodes (leechers). We consider such a classical situation where leechers require finishing downloading a whole file from the seeder before any of them can start using this file. Under Bulk Synchronous Model (BSP) [5], such a download can be seen as a barrier synchronization. Some examples of barrier synchronization include publish–subscribe networks, distributed backup, real-time upgrade of software, and real-time distribution of network configuration data. This communication modern can not only be used as a basic model by P2P applications

H. Zheng (✉)
School of Electronic and Information Engineering, Beijing Jiaotong University, No.3 Shang Yuan Cun, Hai Dian, Beijing, China
e-mail: hyzheng@bjtu.edu.cn

H. Yu
University of Electronic Science and Technology of China, Chengdu, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_63

which require barrier synchronization, but also useful for studying the performance of P2P technology in data center network (DCN) [6]. In DCN, one-to-many traffic pattern is common for many important applications, like software update [7] and data shuffle of massively parallel computing such as MapReduce [8].

Due to the nature of barrier synchronization applications, we are concerned with minimizing the file distribution time. File distribution time is the time it takes for the slowest receiver to get the file.

Minimum distribution time (MDT), as one of the important performance metrics, has been studied. Paper [9] derived theoretical lower bound on MDT and provided a continuous fluid model to achieve this bound. The authors suggest that the upload capacity of the seeder, the download capacity of the slowest leecher, and the aggregate upload and download bandwidths of all the participating nodes are bottlenecks, and they govern MDT. This model has a certain error in approximating the real, discrete piece-based model. Sweha [10] extended work in [9], proposing aid peers to alleviate the effect of network upload capacity bottleneck.

Some works studied the fast data distribution for P2P real-time applications. Liu [11, 12] analyzed the minimum distribution delay in a tree-based P2P live streaming system and proposed a method to finish the dissemination. Zhang et al. [13] present a piece scheduling algorithm for mesh-based P2P live streaming system to achieve optimal distribution delay. Kim and Srikanty [14, 15] construct random Hamiltonian cycles and transfer file over the superposed graph of the cycles so that P2P streaming applications can achieve maximum throughput and minimum delay.

The problem of MDT is also studied in DCN [6, 16]. The authors concern if P2P technology successfully used in Internet can be efficient in DCNs. So these works focus on tree-based DCN rather than complete graph Internet networks.

Papers mentioned above do not give optimal model for one-to-many file distribution. Some existing works addressed optimal model [17, 18]. However, there are fundamental differences between these works and ours. Firstly, [17] works under a pre-assumed node-to-node communication model, i.e., an uplink-sharing model, in which at each time each node can only upload one packet to its receivers and the upload capacity has to be equally distributed to all of its receivers. Secondly, these works assumed that it is network bandwidth that is bottleneck of file distribution. However, in our paper, it is assumed that nodes' storage and computing capacity (together called 'process capacities') are the bottleneck.

This assumption is reasonable. It captures a practical scenario where file is needed to be distributed to mobile devices with limited process capacities. Even if the file is distributed to high-performance computers, it is impossible for a computer to use all of its computing and storage capacities to download and/or upload file. In high-speed network, node's process capacities rather than network bandwidth becomes bottleneck of file distribution with higher probability.

As far as we know, the problem of MDT is studied under this assumption for the first time in this paper. Moreover, the problem is formulated as a 0–1 integer linear program (ILP) problem. A distributed solution is proposed, which takes peer selection and piece selection into consideration together in order to obtain MDT.

The rest of this paper is organized as follows. We formulate the problem in Sect. 63.2. In Sect. 63.3, we describe the proposed distributed file distribution solution. The simulation results are presented in Sect. 63.4, and the Sect. 63.5 contains conclusions.

63.2 Problem Formulation

63.2.1 Assumptions

We assume that network is fully connected. The network bandwidth is not the bottleneck of file transfer; it can be assumed that the network is homogeneous. That is, all of links' bandwidths and download/upload capacities of all nodes can be same.

Files are divided into equal length pieces. Time is divided into time slots, which begin with the transfer of the first bit of a piece and end in completion of the transfer of the piece. The beginning time is zero. For simplicity, assume that the time it takes to transmit a whole file is unit time.

Once a node has received one file piece, it can transmit to any other node. At any time slot, a node can use multiple connections simultaneously for download from or upload to other nodes as long as the number of connections does not exceed the maximum of its process capacity. A node's process capacity involves its storage and computing capacity, which limits the number of concurrent connections. A pair of nodes can only use one connection to transfer data. We assume that for all of nodes, their process capacities are same.

63.2.2 Formulation of the Problem

zThere are N nodes, including seeder and leechers in the network. Denote the set of nodes by N^*. The file is broken into M pieces. Denote the set of the file pieces by F^*. The maximum capacity of node is C. Table 63.1 shows variables and definitions.

The problem of MDT is formulated as an ILP:

$$\min T$$

Subjective to:

$$\sum_j \sum_p \beta_{ij}^p(t) \le C \quad \forall i \in N^*, \quad \forall p \in F^*, \quad \forall t \tag{63.1}$$

$$\sum_i \sum_p \beta_{ij}^p(t) \le C \quad \forall j \in N^*, \quad \forall p \in F^*, \quad \forall t \tag{63.2}$$

Table 63.1 Definitions of variables

Variable	Definition	Value
T	The time it takes for all nodes to finish downloading the file	Nonnegative integer, the initial value is 0
$\beta_{ij}^{p}(t)$	At time t if node i will transmit piece p to node j	Binary is 1 if it will
$\alpha_{i}^{p}(t)$	At time t if node i has received piece p	Binary is 1 if it has
$\delta_{i}(t)$	At time t if node i has received all pieces of the file	Binary is 1 if it has
γ_{i}	The time when node i has received all pieces of the file	Non-negative integer
B	No definition	A big positive constant

$$\sum_{p} \beta_{ij}^{p}(t) \leq 1 \quad \forall i, j \in N^{*}, \ \ \forall t \tag{63.3}$$

$$\beta_{ij}^{p}(t) = 0 \quad i = j, \ \ \forall i, j \in N^{*}, \ \ \forall p \in F^{*}, \ \ \forall t \tag{63.4}$$

$$\beta_{ij}^{p}(t) \leq \frac{1 + \alpha_{i}^{p}(t) - \alpha_{j}^{p}(t)}{2} \quad \forall t, j \neq s, \ \ \forall i \in N^{*}, \ \ j \in N^{*}, \ \ \forall p \in F^{*}\beta \tag{63.5}$$

$$\frac{\sum_{i} \sum_{t} \beta_{ij}^{p}(t)}{B} \leq \alpha_{j}^{p}(t+1) \leq \sum_{i} \sum_{t} \beta_{ij}^{p}(t) \quad \forall t, j \neq s, \ \ \forall i \in N^{*}, \ \ j \in N^{*}, \ \ \forall p \in F \tag{63.6}$$

$$\frac{\sum_{p} \alpha_{i}^{p}(t) - M + 1}{M} \leq \delta_{i}(t) \leq \frac{\sum_{p} \alpha_{i}^{p}(t)}{M}, \quad \forall t, \forall i \in N^{*}, \ \ i \neq s \tag{63.7}$$

$$\gamma_{i} = \sum_{t} \delta_{i}(t) \quad \forall i \in N^{*}, \ \ i \neq s \tag{63.8}$$

$$\gamma_{i} \leq T \quad \forall i \in N^{*} \tag{63.9}$$

$$1 \leq \sum_{p} \sum_{j} \beta_{sj}^{p}(0) \leq C, \quad \forall j \in N^{*}, \ \ j \neq s, \ \ t = 0 \tag{63.10}$$

$$\alpha_{s}^{p}(t) = 1 \quad \forall t, \forall p \in F^{*} \tag{63.11}$$

$$\delta_{s}(t) = 1 \tag{63.12}$$

$$\gamma_{s} = 0 \tag{63.13}$$

$$\alpha_i^p(0) = 0 \quad \forall i \in N^*, \ \ i \neq s, \ \ \forall p \in F^*, \ \ t = 0 \tag{63.14}$$

$$\delta_i(0) = 0 \quad \forall i \in N^*, \ \ i \neq s, \ \ t = 0 \tag{63.15}$$

$$\beta_{ij}^p(0) = 0, \quad \forall i, j \in N^*, \ \ i, j \neq s, \ \ \forall p \in F^*, \ \ t = 0 \tag{63.16}$$

where (63.1)–(63.3) are *capacity constraints*. Equations (63.1) and (63.2) ensure that at time *t,* the number of concurrent upload or download connections cannot exceed node's capacity. Equation (63.3) ensures that at time *t,* a pair of peers only carries out a single connection to download–upload at the same time. Equations (63.4)–(63.6) are *transmission constraints*. Equation (63.4) requires at time *t,* any node cannot transmit to itself. Equation (63.5) requires that a node can get a piece if and only if it does not own the piece and any other has received the piece. Equation (63.6) ensures that a node can get a piece at time *t* if any other node has uploaded the piece to it at any previous time slot. Equation (63.7) ensures that a node has the whole file at time *t* if it has received all pieces of the file. Equation (63.8) calculates the time that a node has received all of file pieces. Equation (63.9) relates the minimum file dissemination time to the time when all nodes have received the file. Equations (63.10)–(63.16) are *initial conditions.* They show that at the beginning, the file is only available at the seeder, and (63.10) ensures that file distribution is started up by the seeder.

63.3 Proposed Distributed Solution

Since finding an optimal solution of the ILP modeling is computationally intractable, we propose a distributed heuristic algorithm to solve the problem. The basic idea is that each node evenly gets data; meanwhile, each piece is evenly distributed into the network, such that all of nodes can get the whole file as quick as possible.

In our solution, the selection of peers and pieces is considered together, as opposed to being done separately. Our selection algorithm is pull based; i.e., the receiver does the selection. In P2P networks where bandwidth is bottleneck, the choice of pull based or push based is related to the location of bottleneck [19]. In this paper, the bottleneck is node's capacity, and the conflicts in the selection are created by nodes (the sender or receiver, depending on the choice of pull based or push based), rather than by bandwidth. We will discuss this in Sect. 63.3.3.

63.3.1 Peer and Pieces Selections

In peer selection algorithm, firstly the possible pair-wise sender–receiver peers are found. A node that can be selected as a sender peer should own pieces the receiver peers want to get. Meanwhile, it should not exceed its maximum capacity to upload.

Otherwise, it will not become a sender. It is possible that although a node owns pieces which other nodes want to get, it cannot upload the pieces due to the constraint of its capacity. This is different from other file distribution solutions, where the receiver can absolutely get pieces from any peer owning the piece.

Then, the algorithm matches the pair-wise sender–receiver peers found above, making sure that pieces will be distributed as evenly as possible.

Once the pair-wise relationship of sender–receiver is set up, the receiver uses piece selection algorithm to select pieces. Our piece selection algorithm is the rarest first. A node first selects the rarest pieces to download. If there are multiple rarest pieces, it randomly selects one to download.

63.3.2 *Pority Queues*

We use two types of priority queues to guarantee even distribution of pieces. One is the receiver queue where the receivers are in a line on descending order of their priorities. The other type of queue is sender queues. For each receiver, there is a sender queue where all of possible senders are in a line on descending order of their priorities. For a pair of nodes, only when they are favored by each other, they can be matched for file uploading and downloading.

A piece that a node does not own is called by 'lack piece' of the node. A receiver node with more lack pieces will be assigned higher priority. When the number of lack pieces is same, the node of which lack pieces are rarer has higher priority. The sender favors to upload content to those nodes which own less pieces and where the lack pieces are rarer. A sender node that owns more lack pieces of the receiver is assigned higher priority. The receiver favors to select nodes that own more its lack pieces as its sender peers.

We propose two methods to assign the priority of receiver nodes. The pseudocode of priority assign algorithms is listed in Table 63.2.

Table 63.2 The receiver priority assignment algorithm(s)

Assignment algorithm: at time slot t
$n_i^r(t) \leftarrow$ the number of lack pieces for receiver node i $P_b(t) \leftarrow$ the popularity of piece b, the number of its copies has been distributed into network $P_x^i(t) \leftarrow$ the popularity of piece x which is a lack piece of receiver node i $w_i(t) = \sum_x P_x^i(t) \leftarrow$ the weight of lack pieces for receiver node i
Assign higher priority to receiver i with bigger value of $n_i^r(t)$ If $n_i^r(t) = n_j^r(t)$ then assign higher priority to receiver i with smaller value of $w_i(t)$
Endif

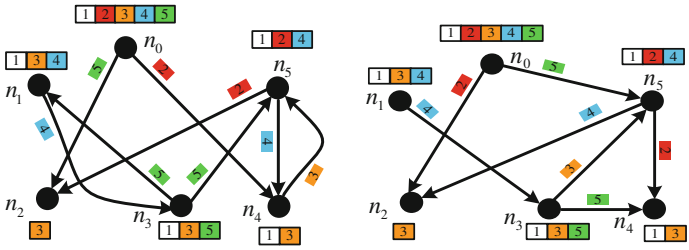


Fig. 63.1 File distribution with method 1 (*left*) and method 2 (*right*)

The second assignment method directly uses $w_i(t)$ to prioritize receiver nodes, giving receiver with bigger value of w_i to higher priority.

Consider an example shown in Fig. 63.1. There are 6 nodes, the seeder is n_0, the file is divided into 5 pieces, and maximum capacity of nodes is 2. With method 1, the receivers' queue is $\{n_2, n_4, n_1, n_3, n_5\}$. Because node n_2 owns the least pieces, it has the highest priority. For those nodes (n_1, n_3, and n_5) with the same number of lack pieces, compare the value of weight of lack pieces. Because w_1 equals to 2, n_1 has higher priority. If we use method 2 to assign priority, then the receivers queue changes to $\{n_2, n_5, n_4, n_3, n_1\}$.

63.3.3 Conflict of Node

The difficulty in peer selection is how to avoid the sender exceeding its capacity when several receivers simultaneously attempt to download from it. We call this phenomenon 'conflict of node'. Each node downloads independently. They do not inform each other of which sender they will select. When multiple nodes simultaneously select the same node as sender, due to constraint of the sender's capacity the selected sender may be cannot simultaneously meet all of downloading requirements; it will just upload the content to some of nodes. For those nodes failing to download, they have to wait for next time slot and this time slot is wasted.

In order to avoid the conflict among nodes, in peer selection, each node has to conjecture other nodes' activities with respect to selection sender at any time slot when it wants to download. The process of conjecture is as follows.

We assume that nodes can exchange information with each other so that each node knows which pieces any other node owns or does not. Under this assumption, each node knows the number and the popularities of lack pieces of all of nodes, and which nodes lack pieces have been distributed to. Based on these information and two queues mentioned above, each node can conclude it will be favored by which nodes owning its lack pieces. Then, it will select them as its own sender peers.

Table 63.3 Sender queues

Node	Sender queue
n_2	n_0, n_5, n_3, n_1, n_4
n_4	n_0, n_5, n_3, n_1
n_1	n_0, n_5, n_3
n_3	n_0, n_5, n_1
n_5	n_0, n_3, n_1, n_2, n_4

See the example shown in Fig. 63.1. Nodes and their corresponding sender queues are shown in Table 63.3. Sender nodes queue on descending order of their priorities. In sender queues of n_2 and n_4, n_3 has higher priority than n_1 because piece 5 it owns is rarer. It can be seen that conflicts of n_0 and n_5 occur for n_2, n_4, and n_1. In the receiver queue, n_2 is at the header and is the most favored; it can safely select n_0 and/or n_5 as its sender peers. For n_1, it can conjecture that both n_2 and n_4 have high possibility to select n_0 and n_5 as sender peers, so it will select n_3 as sender peer to avoid node contention.

If we use method 2, n_1 is at the tail of the receivers queue ($\{n_2, n_5, n_4, n_3, n_1\}$). So, n_1 is the least favored one among all of nodes. It can conjecture how other nodes chose their sender peers, and that it is possible that it has no chance to select a peer, although there are nodes owning its lack pieces, as sender. In this example, one less piece is distributed in this time slot because n_1 fails to download, shown in the right of Fig. 63.1. Compared with method 1 (shown in the left of Fig. 63.1), although method 2 losses a fraction of performance, it is simpler.

63.4 Performance Evaluation

We do simulation to evaluate our solution under small- and large-scale network scenarios, compared with that of the optimal model and BitTorrent-like (BT-like) scheme.

63.4.1 Small-Scale Networks

In this subsection, we compare our solution with that of the optimal model. The solution of the model is solved by Lingo 8.0. The results of our solution are the average of 100 tests with random seeds. Both the values of C and M are 2. The comparison is shown in Fig. 63.2. There are some red circles overlapped with black squares because the time it takes to distribute all pieces across all of nodes is same. In most of scenarios, with our solution, it takes one more time slot to complete file distribution compared to the optimal results.

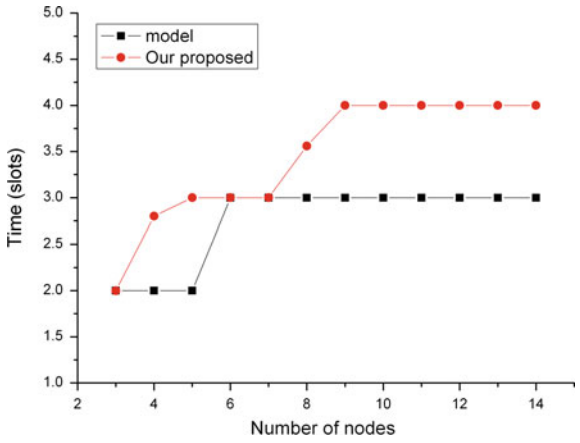


Fig. 63.2 Comparison of our solution and that of the optimal model

63.4.2 Large-Scale Networks

In BT-like, any node selects its sender peers at random among those nodes that own pieces which it wants to get, omitting the signal mechanism. The piece selection is to select the rarest pieces first.

If the time it takes to transmit the file is unit time, the duration of a time slot is unit time divided by M. We use normalized distribution time, i.e., the total duration of time slots it takes to distribute the file across the network. The value of C is 4, equal to the default concurrent connections number of real BitTorrent protocol.

We do test in two cases: (1) $M = 50$ and (2) $M = N$. The comparisons are shown in Figs. 63.3 and 63.4, respectively. It can be seen that in both cases, our proposal outperforms BT-like scheme.

In BT-like scheme, it is possible that multiple nodes attempt to simultaneously download from the same peer because of random peer selection. Limited to

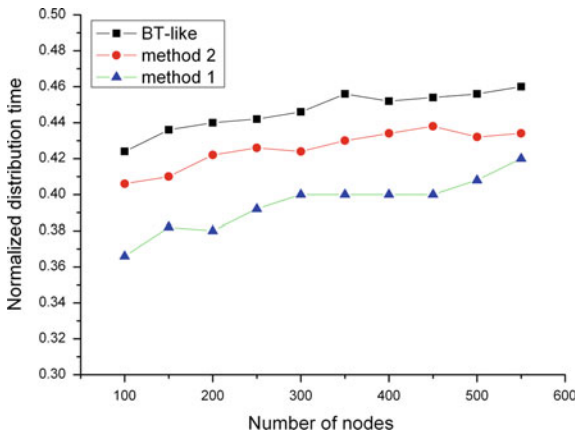


Fig. 63.3 Comparison with BT-like: case 1

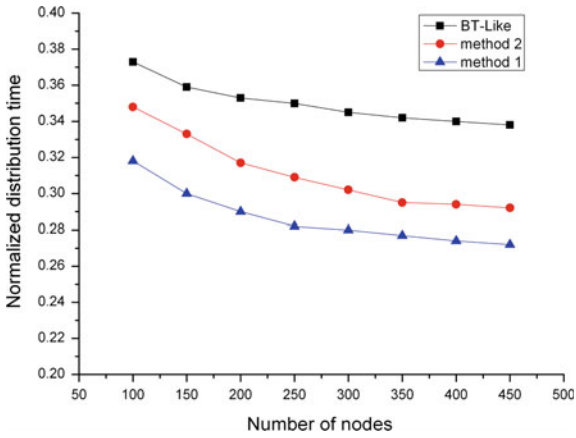


Fig. 63.4 Comparison with BT-like: case 2

capacity, some nodes may be refused at that time slot and have to be downloaded at next time slot, resulting in the current time slot that is wasted.

Besides conflict of sender peer among receiver nodes, BT-like scheme introduces 'conflict' of pieces as well. It is possible that those selected sender peers own same pieces such that the goal of getting lack pieces as more as possible cannot be guaranteed. We call this phenomenon 'conflict' of pieces.

In contrarily, our solution can avoid the conflicts of both of senders and pieces due to taking peer and piece selection into consideration together. When a node selects its sender, it can infer any other node's peer selection,and it will select its sender peers with no conflict as far as possible. Moreover in our algorithms, a node always selects those nodes that own more its lack pieces as its sender peers to guarantee that the receiver node can download more lack pieces in a time slot.

The performance with method 2 is worse. The reason is the weight of lack pieces used in method 2 cannot guarantee the rarest pieces always be selected to send. Consider two scenarios. One is that a receiver node wanders a small quantity of rare pieces, which may be result in a small value of weight of lack pieces. The other is that the receiver wanders lots of popular pieces; the value of the weight of lack pieces can be high. In these two cases, it is possible that the rarest pieces will not be selected to be distributed.

63.5 Conclusion

Minimizing the time it takes to distribute all pieces of a file across all of nodes with constraints of nodes' storage and computing capacities is addressed. We formulate it as an ILP problem and propose a distributed solution. Based on the proposed

algorithm, peer-assisted mechanism can be implemented for file dissemination. The proposed scheme outperforms BT-like scheme.

Acknowledgments This work is supported in the Fundamental Research Funds in Beijing Jiaotong University under grant W11JB00630.

References

1. Liu J, Rao SG, Li B, Zhang H (2008) Opportunities and challenges of peer-to-peer internet video broadcast. In: Proceedings of the IEEE special issue on recent advances in distributed multimedia communications, pp 11–24
2. PPLive (2012) http://www.pptv.com. Accessed 2012
3. PPStream (2012) http://www.ppstream.com. Accessed 2012
4. Cohen B (2003) Incentives build robustness in BitTorrent. In: First workshop on economics of peer-to-peer systems 2003
5. BSP Website (2012) Available: http://www.bsp-worldwide.org. Accessed 2012
6. Zhao Y, Wu J, Liu C (2014) On peer-assisted data dissemination in data center networks: analysis and implementation. Tsinghua Sci Technol 19(1):51–64
7. Twitter Engineering (2010) Murder: fast data center code deploys using bittorrent. http://engineering.twitter.com/2010/07/murder-fast-data-center-codedeploys.html. Accessed 2010
8. Dean J, Ghemawat S (2004) Mapreduce: Simplified data processing on large clusters. In: Operating systems design implementation
9. Kumar R, Ross KW (2006) Peer-assisted file distribution: the minimum distribution time. In: Hot topic in web systems and technologies, Boston, pp 1–11
10. Sweha R (2009) Angels: in-network support for minimum distribution time in P2P overlays [ME], Boston University, Boston
11. Liu Y (2007) On the minimum delay peer-to-peer video streaming: how real time can it be? In: ACM multimedia
12. Liu Y (2011) Reducing or minimizing delays in peer-to-peer communications such as peer-to-peer video streaming. US Patent # US 8,005,975 B2
13. Zhang D, Wang L, Yang H (2013) An effective data scheduling algorithm for mesh-based P2P live streaming. In: Information science and technology 2013, Yangzhou, pp 1221–1224
14. Kim J, Srikanty R (2012) Achieving maximum throughput and minimum delay in heterogeneous peer-to-peer streaming networks. In: 2012 conference record of the forty sixth Asilomar conference on signals, systems and computers, pp 1184–1188
15. Kim J, Srikanty R (2012) Real-time peer-to-peer streaming over multiple random hamiltonian cycles. IEEE Trans Inf Theor 59(9):5763
16. James S, Crowley P (2011) Fast content distribution on data center networks. In: 2011 ACM/IEEE seventh symposium on architectures for networking and communication systems, pp 87–88
17. Mundinger J, Weber R, Weiss G (2008) Optimal scheduling of peer-to-peer file dissemination. J Sched 11(2):105–120 (8)
18. Xie B, van der Schaar M, Courtade T, Wesel RD (2011) Minimizing weighted sum finish time for one-to-many file transfer in peer-to-peer networks. In: Communication, control, and computing, pp 1771–1778
19. Abeni L, Kiraly C, Cigno RL (2009) On the optimal scheduling of streaming applications in unstructured meshes. In: the 8th international IFIP-TC 6 networking conference, pp 117–130

Chapter 64
Reliability Allocation of High-Speed Train Bogie System

Wantong Li, Yong Qin, Shuai Lin, Limin Jia, Min An, Zhilong Zhang, Xiaojun Deng and Hengkui Li

Abstract In order to improve the accuracy reliability of the high-speed train bogie system, fuzzy comprehensive assessment matrix was applied to allocate the reliability of bogie system. The effects of the reliability allocation of the high-speed train bogie system have a multilayered and multifactor characteristic which includes the quantitative and qualitative indexes. A fuzzy analytic hierarchy process (AHP) with five-point scale was proposed. The influence factors were considered and endowed with different weights according to the influence degree. A structural model of AHP is proposed for bogie system, which used interval number instead of fuzzy assessment matrix to express uncertain information. It can make use of expert's experience and the field test information comprehensively. A high-speed train bogie system was taken as example to illustrate the application of this method.

Keywords Fuzzy AHP · Reliability allocation · High-speed train bogie system

64.1 Introduction

Reliability reflects the ability of a system to accomplish required tasks under certain circumstances and scheduled time. Reliability allocation is a very important step of product development process for setting achievable reliability goals. It is that the reliability index in design specification is appropriately allocated in the several subsystems and components which make up the system. Results of reliability allocation

W. Li · S. Lin
School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China
e-mail: 13125709@bjtu.edu.cn

Y. Qin (✉) · L. Jia · M. An
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
e-mail: qinyong2146@126.com

Z. Zhang · X. Deng · H. Li
CRRC Qingdao Sifang Co., Ltd, Qingdao, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_64

would provide essential references to design, manufacture, and inspection of the system. Some of the traditional approaches such as ratio method [1] and the AGREE [2] are specific to the series system. Other methods such as the AGREE cannot consider the factors influencing reliability comprehensively. Genetic algorithm might plunge into local optima easily [3]. Lagrange multiplier method, dynamic programming, and direct seeking method need many assumptions and simplifications when the system has lots of subsystems and components [4]. In addition, the calculating amount of these methods is quite large and these approaches cannot have good quality of solutions. As a consequence, using the traditional approaches of reliability allocation for complex systems like high-speed train bogie system cannot analyze the reliability allocation accurately and scientifically.

AHP is a system analysis method which was proposed by Saaty in 1980s. It is a systematic decision-making method which includes both qualitative and quantitative techniques. It is being widely used in many fields for a long time. The traditional AHP method has the following disadvantages:

1. Its judgment matrix uses the 9-point scale. Sometimes, the decision-maker might find difficult to distinguish among them [5].
2. The difference between the subjective opinions and the objective facts and the inaccuracy of the scales measuring the problem lead to the consequence that the judgment matrix cannot satisfy the consistency conditions. Therefore, the result should be amended [6].
3. When using the row vectors sum uniformization method and the square root method to rank the importance of each program, the effect of one corresponding row of elements in the judgment matrix is only considered. The traditional AHP exhibits low accuracy of calculation, and it cannot control the precision as required. Therefore, this method cannot reflect the actual situation [7, 8].

In order to eliminate this limitation of the traditional AHP, this paper combines the fuzzy evaluation to AHP [9–11]. A fuzzy analytic hierarchy process (AHP) with five-point scale was proposed to define the index weight of the index system. Using this method, the reliability allocation of high-speed train bogie system is analyzed.

64.2 The Fuzzy AHP

The fuzzy AHP judges the matrix scale with five-point scale. It is easy for expert to judge the relative importance between two factors. And the fuzzy consistent matrix satisfies the consistency conditions, and as a result, it does not need consistency examination. And it may induce the iteration times enormously and improve the convergence speed so that the calculation precision can be ensured.

The following gives the steps:

1. Build the complementary fuzzy judgment matrix F (or precedence judgment matrix) with five-point scale.

$$F = (f_{ij})_{n\times n} = \begin{cases} 0.1 & A_j \text{ is strongly more important than } A_i \\ 0.3 & A_j \text{ is moderately more important than } A_i \\ 0.5 & A_j \text{ is as important as } A_i \\ 0.7 & A_i \text{ strongly more important than } A_j \\ 0.9 & A_i \text{ strongly more important than } A_j \end{cases}$$

where A_i is the index in the ith line. A_j is the index in the jth row. n is the number of the elements in the criterion layer.

2. Transform the fuzzy judgment matrix F to the fuzzy consistent judgment matrix Q. Then, get the summation of each line and use the transforming equation

$$q_{ij} = \frac{q_i - q_j}{2n} + \frac{1}{2} \tag{64.1}$$

To obtain the fuzzy consistent judgment matrix $Q = (q_{ij})_{n\times n}$.

3. Get the weights by the row vectors sum uniformization method. The elements' summation (except the comparison value of themselves) of each line in the fuzzy consistent judgment matrix and the summation of the elements except the elements in the diagonal line are

$$l_i = \sum_{j=1}^{n} q_{ij} - 0.5 \tag{64.2}$$

$$\sum_{j=1}^{n} l_j = n(n-1)/2 \tag{64.3}$$

where l_i is the importance of index i relative to the goal of the upper layer. Then, the weights of each index can be obtained by uniformization.

$$W_i = l_i / i = \sum_{j=1}^{n} \frac{2l_j}{n(n-1)} \quad i = 1, 2, \ldots, n \tag{64.4}$$

And the weight vector is $W = (W_1, W_2, \ldots W_n)$.

4. Allocate the reliability index according to the comprehensive weight. The comprehensive weight B of the elements in the alternative layer is introduced to allocate the failure rate λ required by each subsystem. Then, allocate the reliability based on λ.

$$B = W \cdot Z = [W_1, W_2, \ldots, W_n] \begin{bmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & \ddots & \vdots \\ z_{m1} & \cdots & z_{mn} \end{bmatrix} = [b_1, b_2, \ldots, b_n] \tag{64.5}$$

$$\lambda = \frac{\lambda}{w_1} \Big/ \sum_{i=1}^{n} \frac{1}{w_i} \tag{64.6}$$

where n represents the alternative layer that consists of n elements. m represents that there are m fuzzy effects.

5. Calculate the weight of each element in the alternative layer, and allocate the reliability index based on the weight.

$$R(t) = \mathrm{e}^{-\lambda t} \quad i = 1, 2, \ldots, n \tag{64.7}$$

64.3 The Hierarchical Structure of High-Speed Train Bogie System

High-speed train bogie system is a compound system. Simplify the parallel system in the compound system as a series system to simplify the computation. Therefore, the high-speed train bogie system consists of the framework, the wheelset assembly, the axle boxes, the primary suspension, the secondary suspension, the driving device, and the foundation brake system. The reliability block diagram of high-speed train bogie system is shown in Fig. 64.1.

While allocating the reliability of the bogie system, the weight of each subsystem should be determined first. The following effects are commonly considered: (1) the damage degree when the subsystems have faults, (2) the complexity degree of the subsystems, (3) the technical merit of the subsystems, and (4) the working condition of the subsystems.

The total reliability of the bogie system is the goal layer E. The damage degree, the complexity degree, the technical merit, and the working conditions which effect the reliability allocation of the system form the criteria layer B. And take the framework, the wheelset assembly, the axle boxes, the primary suspension, the secondary suspension, the driving device, and the foundation brake system as the

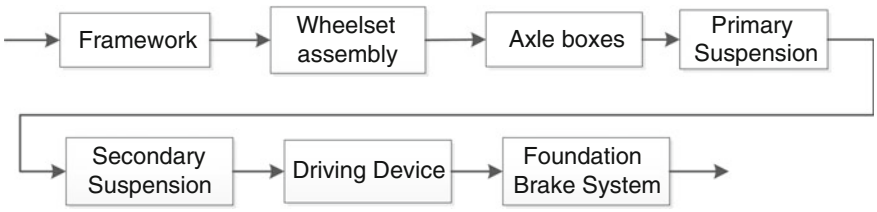


Fig. 64.1 The reliability block diagram of high-speed train bogie system

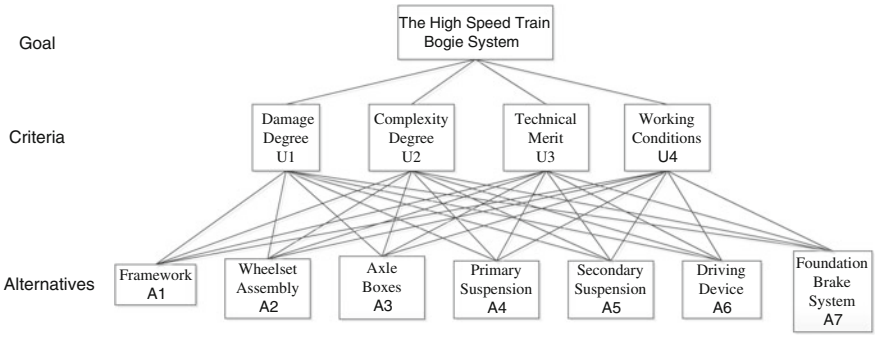


Fig. 64.2 The hierarchical structure model of high-speed train bogie system

alternative layer. Then, build the analytic hierarchy model and make the reliability allocation program which satisfies the required reliability of the total bogie system. Its hierarchical structure model is shown in Fig. 64.2.

64.4 The Reliability Allocation of the High-Speed Train Bogie System

According to the hierarchical structure model and the preferential relationship with 5-point scale, compare the influence degree of reliability allocation in the criteria layer. Based on the result proposed by the experts, the weight relationship of reliability impacting the bogie system is obtained: the damage degree > the complexity degree > the working conditions > the technical merit. Transform it into the precedence relation matrix and is shown in Table 64.1.

According to Table 64.1, obtain the summation of the lines.

$$q_1 = 3.2 \quad q_2 = 2.0 \quad q_3 = 1.2 \quad q_4 = 1.6$$

Transform the preference relation matrix to the fuzzy consistent matrix by Eq. (64.1)

$$Q = \begin{bmatrix} 0.50 & 0.65 & 0.75 & 0.70 \\ 0.35 & 0.50 & 0.60 & 0.55 \\ 0.25 & 0.40 & 0.50 & 0.45 \\ 0.30 & 0.45 & 0.55 & 0.50 \end{bmatrix}$$

Figure out the weight of the elements in the criteria layer relative to the goal layer by Eqs. (64.2), (64.3), and (64.4).

Table 64.1 *C–B* preference relation matrix

C	*U*1	*U*2	*U*3	*U*4
*U*1	0.5	0.9	0.9	0.9
*U*2	0.1	0.5	0.7	0.7
*U*3	0.1	0.3	0.5	0.3
*U*4	0.1	0.3	0.7	0.5

Table 64.2 The preference relation matrix of the damage degree

*U*1	*A*1	*A*2	*A*3	*A*4	*A*5	*A*6	*A*7
*A*1	0.5	0.9	0.9	0.9	0.9	0.9	0.9
*A*2	0.1	0.5	0.9	0.9	0.9	0.9	0.7
*A*3	0.1	0.1	0.5	0.7	0.7	0.7	0.3
*A*4	0.1	0.1	0.3	0.5	0.7	0.7	0.3
*A*5	0.1	0.1	0.3	0.3	0.5	0.3	0.1
*A*6	0.1	0.1	0.3	0.3	0.7	0.5	0.3
*A*7	0.1	0.3	0.7	0.7	0.9	0.7	0.5

Table 64.3 The preference relation matrix of the complexity degree

*U*2	*A*1	*A*2	*A*3	*A*4	*A*5	*A*6	*A*7
*A*1	0.5	0.3	0.1	0.1	0.3	0.1	0.7
*A*2	0.7	0.5	0.1	0.1	0.7	0.1	0.7
*A*3	0.9	0.9	0.5	0.3	0.9	0.7	0.9
*A*4	0.9	0.9	0.7	0.5	0.9	0.7	0.9
*A*5	0.7	0.3	0.1	0.1	0.5	0.1	0.7
*A*6	0.9	0.9	0.3	0.3	0.9	0.5	0.9
*A*7	0.3	0.3	0.1	0.1	0.3	0.1	0.5

Table 64.4 The preference relation matrix of the technical merit

*U*3	*A*1	*A*2	*A*3	*A*4	*A*5	*A*6	*A*7
*A*1	0.5	0.9	0.9	0.9	0.9	0.9	0.1
*A*2	0.1	0.5	0.9	0.9	0.9	0.9	0.1
*A*3	0.1	0.1	0.5	0.7	0.7	0.7	0.1
*A*4	0.1	0.1	0.3	0.5	0.3	0.3	0.1
*A*5	0.1	0.1	0.3	0.7	0.5	0.3	0.1
*A*6	0.1	0.1	0.3	0.7	0.7	0.5	0.1
*A*7	0.9	0.9	0.9	0.9	0.9	0.9	0.5

$$W = (0.350, 0.250, 0.183, 0.217)$$

The result of the experts' judgment shows $p_i (i = 1, 2, \ldots, 7)$ (Tables 64.2, 64.3, 64.4, and 64.5).

Calculate the weight of the elements in the alternative layer against the elements in the criteria layer.

Table 64.5 The preference relation matrix of the working conditions

$U4$	$A1$	$A2$	$A3$	$A4$	$A5$	$A6$	$A7$
$A1$	0.5	0.1	0.3	0.9	0.9	0.9	0.9
$A2$	0.9	0.5	0.9	0.9	0.9	0.9	0.9
$A3$	0.7	0.1	0.5	0.9	0.7	0.9	0.9
$A4$	0.1	0.1	0.1	0.5	0.7	0.9	0.9
$A5$	0.1	0.1	0.3	0.3	0.5	0.9	0.7
$A6$	0.1	0.1	0.1	0.1	0.1	0.5	0.1
$A7$	0.1	0.1	0.1	0.1	0.3	0.9	0.5

$$
\begin{aligned}
b_1 &= (0.200 \quad 0.162 \quad 0.138 \quad 0.129 \quad 0.100 \quad 0.120 \quad 0.152) \\
b_2 &= (0.113 \quad 0.132 \quad 0.163 \quad 0.194 \quad 0.122 \quad 0.174 \quad 0.103) \\
b_3 &= (0.176 \quad 0.167 \quad 0.129 \quad 0.100 \quad 0.110 \quad 0.119 \quad 0.200) \\
b_4 &= (0.167 \quad 0.200 \quad 0.171 \quad 0.138 \quad 0.129 \quad 0.086 \quad 0.110)
\end{aligned}
$$

Express them to a matrix.

$$
B = \begin{bmatrix}
0.200 & 0.162 & 0.138 & 0.126 & 0.100 & 0.120 & 0.152 \\
0.113 & 0.132 & 0.163 & 0.194 & 0.122 & 0.174 & 0.103 \\
0.176 & 0.167 & 0.129 & 0.100 & 0.110 & 0.119 & 0.200 \\
0.167 & 0.200 & 0.171 & 0.138 & 0.129 & 0.086 & 0.110
\end{bmatrix}
$$

The comprehensive weight is

$$
B = W \cdot Z = [0.167 \quad 0.163 \quad 0.150 \quad 0.142 \quad 0.113 \quad 0.126 \quad 0.139]
$$

When the failure rate of the high-speed train bogie system $\lambda = 1.86 \times 10^{-5}$ and the working time is 12,000 h, the reliability of each subsystem is obtained by Eqs. (64.6) and (64.7).

$$
\begin{aligned}
&\lambda_1 = 0.224 \times 10^{-5},\ \lambda_2 = 0.228 \times 10^{-5},\ \lambda_3 = 0.249 \times 10^{-5}, \\
&\lambda_4 = 0.264 \times 10^{-5},\ \lambda_5 = 0.329 \times 10^{-5},\ \lambda_6 = 0.297 \times 10^{-5}, \\
&\lambda_7 = 0.268 \times 10^{-5}
\end{aligned}
$$

The reliability of the bogie system is 0.8. The result of the allocation reliability is shown in Table 64.6.

According to the result of the calculation, the framework and the wheelset assembly should be allocated more reliability and it accords with the fact that the framework is the most important component in the bogie system and the wheelset assembly has the most failure rate. It follows that the fuzzy AHP method is practical on the reliability allocation of the high-speed train bogie system.

Table 64.6 The result of the allocation reliability of the high-speed train bogie system

Reliability index	Framework	Wheelset assembly	Axle boxes	Primary suspension	Secondary suspension	Driving device	Foundation brake system	Total
Failure rate/10^{-5}	0.224	0.228	0.249	0.264	0.329	0.297	0.268	1.86
Reliability	0.9734	0.9730	0.9705	0.9689	0.9613	0.9650	0.9684	0.8

64.5 Conclusion

The fuzzy AHP uses the information of the fuzzy elements which have an important effect on the reliability allocation. It uses the human brain to have a correct judgment of the fuzzy phenomenon and makes full use of the experts. Moreover, the fuzzy AHP reduces the drawback which is caused by the personal subjective judgment. It is a scientific and reasonable method to allocate the reliability.

This paper considers the high-speed train system as a series system and considers 4 elements affecting the system. It is feasible by this algorithm, and the result is similarly consistent to the practical situation. However, there is still a certain gap against the fact. It is also a problem to consider all the uncertain factors and have a quantitative assessment so that the model of reliability allocation is more tally with the actual situation and more reliable.

Acknowledgments The project is supported by the independent subject (Grant No. I14K00451) and National High Technology Research and Development Program of China (863 Program) (Grant No. T13B200011).

References

1. Zhang Q, Huang Q, Zhang Y (2010) Research on the method of reliability allocation based on exponential distribution. Ship Electron Eng 30(3):151–154 (in Chinese)
2. Xiang Y, Huang D, Huang L (2010) Reliability allocation of BA system based on grey relative theory and AGREE. Appl Res Comput 27(12):4489–4491 (in Chinese)
3. Li F, Liu S, Chen B (2004) System reliability allocation model based on genetic algorithms. J Inst Command Technol Equipment 15(4):982100
4. Zhang X, Wu X (2012) Research progress of system reliability allocation. Fire Command Control 37(8):1–4
5. Li Y, Hu X, Qiao J (2005) An improved fuzzy AHP method. J Northwest Univ 35(1):11–13 (in Chinese)
6. Shaw K, Shankar R, Yadav SS (2012) Supplier selection using fuzzy AHP and fuzzy multi-objective linear programming for developing low carbon supply chain. Expert Syst Appl 39(9):8182–8192
7. Sun L, Ren F, Yang C (2006) Study on the indicator weight and system on feasibility of expressway construction item. J Beijing Univ Technol 32(7):596–599 (in Chinese)
8. Chan FTS, Kumar N, Tiwari MK (2008) Global supplier selection: a fuzzy-AHP approach. Int J Prod Res 46(14):3825–3857
9. Zhu D, Liu H (2011) Reliability CC. China Mech Eng 22(24):2923–2927 (in Chinese)

10. GU Yingkui, WU Lucheng (2007) A fuzzy AHP approach to the determination of weights of evaluation process. China Mech Eng 18(9):1052–1055, 1067 (in Chinese)
11. Wang S, Liu H (2013) Reliability allocation of NC machine accuracy based on analytic hierarchy process and fuzzy comprehensive assessment. Mech Sci Technol Aerosp Eng 10:1441–1445 (in Chinese)

Chapter 65
Optimizing Adaptive Clustered Bitmap Index Based on Medical Data Warehouse

Yanfei Liu

Abstract As a high-performance index structure, bitmap index has been widely used in data warehouse (DW). However, due to the particularity of medical data, the efficiency is different between conventional DW and medical data warehouse (MDW). The key challenge is the uneven distribution of medical data in MDW. Then, the user's queries are generally for sparse region data. Therefore, the new index mechanisms that can provide efficient multidimensional query are necessary. This paper fully considering the distribution characteristics and data type of medical data, a dynamic self-adaptive optimization bin boundary bitmap indexing algorithm is proposed. It can accelerate data access by automatically collecting the user query points to optimize the boundary of binning scheme and reduce the number of candidate check. Meanwhile, newly adding two-dimensional index structure gathers the above processing sparse bitmap to speed up the response of multidimensional query operations. In performance test, with the increase of query boundary points to readjust the bin boundaries, the query performance optimization is improved by almost 60 % compared with that before adjustment.

Keywords Medical data · Bitmap index · Candidate check · Dynamic self-adapting

65.1 Introduction

With the rapid development of database technology and various devices to popularize in hospital, huge amounts of medical data are recorded. These data have very important research value for disease diagnosis, treatment, health warning, remote medical treatment, and scientific experiments. However, in terms of too large growth of data, the average data growth per year will reach dozens of TB in a big general hospital, and it is impossible to rely on hospital storage capacity to solve the

Y. Liu (✉)
College of Information Engineering, Capital Normal University, Beijing, China
e-mail: merrylyf@163.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_65

problem. The cloud of medical data migration and creating MDW (i.e., medical data warehouse) in the cloud are good ways. However, in the process of migration, the basic characteristics of medical data bring the new challenge for data access and query. Different from traditional data, the value of medical data has complex data type and data structure. Then, the value is uneven among the same attributes that it generally contains the value range and the standard range. Each of the data has a predetermined value domain. Meanwhile, most of the data are present within a certain range. For example, the value range of the data of diastolic blood pressure is [0,100], among which anyone may not exceed the range, while the standard range is [60–90], among which most of the data are in this range. Because the value of fact table represents some human characteristics, the change scope of each data attribute is relatively small. The feature makes it possible that some new approaches must be taken to improve the existing index method. Before that, the bitmap indexes [1] can be generally introduced in DW. It is efficient to low-cardinality attributes and inefficient to high-cardinality attributes due to the large storage requirements for the bitmap indices. In order to overcome this limitation, a variety of methods had been offered to reduce the storage complexity of bitmap indexes. To sum up, they are mainly divided into 3 categories, namely compression, bitmap encoding, and binning.

Of these methods, binning resembles simple bitmap index, but it partitions the attribute values into a series of ranges, known as bins, and uses bitmap vectors rather than bitmap bit to represent bins. But the query performance is restricted by check candidate and performance is unstable.

According to [2], candidate check is a main factor affecting performance. As to the uneven distribution of data in MDW, the high density of data in standard range, and the sparse density in non-standard range, the paper mainly contributes to the following aspects:

- Give the standard range and min–max range, automatically calculate the units of non-standard range which is partitioned into several bins according to unit, collect the user's query point, and propose one-dimensional bin optimization algorithm under the restrictions on the number of given bins.
- In view of the multidimensional dataset, a two-dimensional structure is established to achieve the gathering of sparse attributes. The value's sequence of the structure is in accordance with the position of bins.
- Using the actual medical datasets (serious disease early warning dataset) complete performance evaluation on time and space.

65.2 Related Works

Now, the research of MDW and the corresponding indexing in view of medical data is in the primary stage. In medical institutions, this kind of work is built with the use of existing tools. Xiangwu Ding et al. implemented the creation of MDW using the

Microsoft BI in [3]. But they do not change the indexing according to the medical data features. Generally, the changes of index structure are all based on binning algorithms. The revised classified index and inverted index to full-text search are devised to accelerate the access speed for text messages. However, they cannot evidently promote the retrieval speed of measurement data of clinical or health check. In this paper, the new indexing is created according to the features of measurement data and corresponding user's queries, which also can coordinate for specific queries. Compared with other indexing technologies, bitmap index improves the query performance by converting the operation, such as comparison, connection, or data gathering into logic bit operation. Typically, in FastBit, an open source NOSQL database, developed by Wu et al., he offered plenty of bitmap coding strategies which provide an optimizing direction for bitmap indexes of high-cardinality attribute and pointed out binning is the most promising technique [4] which partitions values into bins according to vague range. In [5], it indicates that on the premise of even data distribution, 80 % of query time is spent on candidate check. Wu adds a new indexed structure OrBic [6] and establishes secondary index to collect the data in one bin. All these methods intend to reduce the number of candidate check for some attribute. Now, they have been expended to support multidimensional queries. Siqueira et al. [7] and Howison et al [8] also create new indexing structure to multidimensional cases. But all these performances proceed on the premise of the uniformity of data distribution. Some of the characteristics of medical data can be summarized as follows:

- Huge data: MDW is typically very large and grows over time to be terabytes or petabytes. Generally, the fact table consists of a large number of attributes and the data keep relatively stable, updating less.
- Uneven distributed data: Medical measuring data are main components of medical information, including human physiological index data and measuring index data, most of which represent some of the characteristics of human indicators and generally are numeric. All these data have two distributed scopes, standard range and sparse range. But the queries of the sparse range are popular because these data have research value in some field, such as disease alerting systems. This point is different from all other datasets.
- Uncertain query sets: The relation between illness and concrete data on MDW could not be determined beforehand. Despondently, the endpoints of query sets cannot be gotten.

In this paper, according to the concrete characteristics of medical data, a new algorithm will be put forward to automatically collect the query endpoints, optimize the bin boundaries, and subdivide the sparse range about the one-dimensional attributes. Thus, another procedure will be adopted to gather the data of sparse range.

65.3 Optimizing Dynamic Adaptive Binning Algorithm

In view of Sect. 65.2, data distribution and query range are the two major factors that affect the read database data. As to the data in the fact table, the distribution is seriously lack of balance, with more data falling into standard range and other range sparsely. In this section, an optimizing algorithm is provided to solve the query problem for medical data under the premise that the set of queries is unknown and standard range and min–max endpoints are known. There are three parts: Firstly, the theoretical basis of algorithm will be offered in view of the correctness of our proposition, then, the detailed design of algorithm will be presented, and finally, the gathering of data will be implemented on account of the dependency between illness and multiattributes.

65.3.1 Theoretical Principle

In the binning, a little candidate check may spend a lot of I/O time. Bin boundaries and query endpoints influence the number of candidate check. Give a hypothesis, the attribute A is in database D, and $A = \{v_1, v_2, v_3,\ldots, v_n\}$. There are n different attribute values, $v_i \in A$, in a typical kind of value of attribute A. Bin bitmap index will be established about A. There are k bins denoted by B, and $B = \{b_1, b_2, b_3,\ldots, b_k\}$. Generally, the value of n is far greater than that of k, which substantially reduces the binary bits of bitmap index. The $k - 1$ boundaries are selected among $n - 1$ different values except the lowest value in A, which is a question of combination, and there are a total of C_{n-1}^{k-1} methods among which there is undoubtedly a kind of method with the minimum check candidates. Candidate check is related to the endpoints of query sets and bins. V denotes a vector that automatically collects the query endpoints which the user inputs. The beginning of V is empty, and it will store the endpoints in ascending order and eliminate the same points. The combination of two adjacent consecutive values in V produces a query in queries collection Q, so that each $q \in Q$ defines a range q, and $q = [l_q, u_q)$, open on the right, q denoting that a query can be launched to find values which are equal or greater than l_q and less than u_q. The number of check candidates is shown in Fig. 65.1, which can be predicted from the two endpoints falling into the bin numbers.

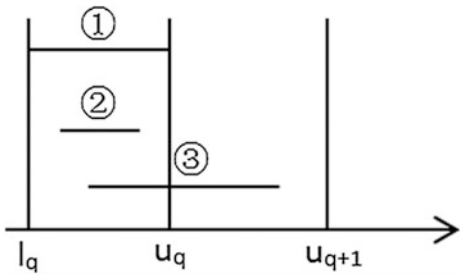


Fig. 65.1 The mark of check candidates

The mark number 1 depicts that query endpoints fall exactly on the two boundary points. In this position, the number of check candidates is zero. The mark number 2 indicates that one of query endpoints coincides with a boundary endpoint, and another endpoint of query is within the bins. This time the number of check candidates is 2. The two endpoints of query q and points of bins do not overlap in the mark number 3. The more bins the candidate check needs to check, the greater the corresponding I/O costs. Queries on each data need to be determined its location and position in all bins, but these bins are not always in the same disk page, need to be read constantly. On the premise of even data distribution, the disk pages of a bin can be roughly estimated. Assuming the attribute A_i occupies the number of p disk pages, if there are n_b values in this bin, the bin will occupy P_b disk pages. And P_b satisfies the formula [9] as follows:

$$P_b = P\left(1 - e_b^{-n/P}\right) \tag{65.1}$$

Under the same conditions, the optimizing binning in C_{n-1}^{k-1} is related to the endpoints of query sets. $EP(Q) = \langle e_1, e_2, \ldots, e_r \rangle$ denotes the auto-collecting query endpoints. Rotem et al. [1] etc. put forward the optimizing theorem:

Lemma 1 *If any binning $B \in C_{n-1}^{k-1}$ uses a boundary point xi that is not in EP(Q), then a binning of equal or smaller cost can be found by replacing x_i with some point in EP(Q).*

As to medical database, the sparse range and standard range are known. Thus, the unit can be inferred from standard range to detailing the sparse range. According to Lemma 65.1, the later detailing sparse range can act as optimizing bins as following corollary. First, a constraint must be given to the number of bins, assuming k.

Lemma 2 *In all the strategies of binning $B \in C_{n-1}^{k-1}$, the number of bins is less than or equal to k; if a bin including the number of query endpoints is greater than or equal to 1, it will be portioned into 1, 2, or 3 bins. Query performance will not be lowered.*

Proof Assume w_i denotes the cost of check b_i, x is the query endpoint in b_i, the points of b_i are x_i and x_{i+1}, b_i will be divided into two parts, namely the range of $[x_i, x)$ and $[x, x_{i+1})$ with the cost, respectively, w_x and w_y, and the relation is $w_x + w_y <= w_i$. So the preceding cost of division can be denoted as follows:

$$\text{cost}(b_i, Q) = P\left(1 - e_b^{-n}\right) w_i \tag{65.2}$$

It is mutilation of the number of falling into b_i and the cost w_i. The later cost of division is as follows:

$$\text{cost}(b_x, Q) + \text{cost}(b_y, Q) = P\left(1 - e_x^{-n}\right) w_x + P\left(1 - e_x^{n-n}{}_b\right) w_y \tag{65.3}$$

The difference value is as follows:

$$\begin{aligned}\mathrm{cost}(b_x, Q) + \mathrm{cost}\left(b_y, Q\right) - \mathrm{cost}\left(b_{i,}Q\right) <= Pw_x\left(e_b^{-n} - e_x^{-n}\right) \\ + Pw_x\left(e_b^{-n} - e_x^{n\ -n}{}_b\right) <= 0\end{aligned} \tag{65.4}$$

The division cannot be added too much, because it will decay simple bitmap index and extend the length of bits of bitmap index.

65.3.2 Dynamic Adaptive Bin Boundaries

In medical scientific applications, query range almost concentrates on the sparse range for its research meaning. So it is difficult to find the existing optimizing binning model to realize bitmap index. In this section, based on vector array which will act as the collection of query endpoints and the subdivision of the range of bins, an algorithm will be offered to formulate optimizing bins for medical database check. Firstly, as to an attribute A of dataset, $A = \{v_1, v_2,\ldots, v_n\}$, the standard range $[v_c,v_m]$ partitions the range of A into two parts, respectively, $[v_1, v_c)$ and $(v_m,v_n]$, where the data in some range may be empty. Set the query vector Q, its endpoints EP(Q) = $\langle e_1, e_2,\ldots, e_r\rangle$, and at the beginning, the EP(Q) = $\langle\ \rangle$, collecting user's query data automatically. Then, formulate the binning algorithm. Firstly, hardly does any endpoint fall into the standard range, so it will act as the bin located at the end of other bins. Now, consider other bins. According to the standard range, the unit of data variation can be inferred in advance, and then, the range $[v_1,v_c)$ and $(v_m,v_n]$ is divided into the beginning bins using the unit, respectively, counting the numbers in which the query endpoints fall into all bins. Based on the falling query endpoint, the beginning bins are optimized. If many endpoints fall into it, the bin is partitioned into several bins at the constraint bin number k. The specific processing algorithm is depicted as follows:

1. Use the tools of DBMS to extract the attribute column value that is vector U={v_1,v_2,…v_c}. Then, V={v_1}, the optimizing bins endpoints, and EP={} the vector of collecting user's query sets endpoints.
2. EP={e_1,e_2,…e_r}//automatically collecting the endpoints and getting rid of the same endpoints, sorting.
3. for i=1 to r// count the number which falls into bins.
   for j=1 to c-1 //c denote the max range.
   if EP[i]<=U[j] {count[j-1]++; break;}
4. //set j=1, denote the index in V.
   for i=1 to c-1 //among the c-1 range, adding bins according to the vector count.
   switch (count[i]){

   case 0:// U[i+1]=0;
   case 1: //If U[i] is not the first value, set U[i]=0, fill the corresponding query endpoint in U[i+1] and V[++j], or else keep the U[i].

RID	A_1	b_{11} [0,52)	b_{12} [52,70)	b_{13} [70,90)
1	45	1	0	0
2	78	0	0	1
3	80	0	0	1
4	39	1	0	0
5	72	0	0	1
6	83	0	0	1
7	76	0	0	1
8	80	0	0	1
9	52	0	1	0

RID	A_2	b_{21} [0,2.4)	b_{22} [2.4,4)	b_{23} [4,6)
1	1.2	1	0	0
2	5.6	0	0	1
3	5.7	0	0	1
4	0.8	1	0	0
5	2.9	0	1	0
6	4.4	0	0	1
7	4.9	0	0	1
8	5.3	0	0	1
9	2.4	0	1	0

Fig. 65.2 Bitmap index bins of A_1 and A_2

case 2: // Compared with the two different values, if the result is less than the 1/10 unit, act on case 1, else if larger than 1/10 unit, partition the bin into three subbins and add V[++j] if U[i]<>0 and all the endpoints that fall into this range.
case 3: // Pairwise comparison, if the results are all less than 1/10 unit, act on case 2, one of results larger, act on case 2, else add the U[i] and endpoints that fall into the range to V.
default: //translate into above-mentioned using combination.}

Record every bin address and return the signal of success. The values of vector V are the endpoints of bins. All the values will be combined based on the creating order to form the bins. Add the standard range as the last bin.

The Fig. 65.2 shows the bins of attributes of A_1 and A_2 in the parts of data of MDW. There are three bins. It can clearly demonstrate that the standard ranges of A_1 and A_2 are b_{13} and b_{23}. The ranges are not commonly used compared with other ranges in queries. So aggregation is the next step to the sparse ranges.

65.3.3 Data Aggregation of Multidimensional Queries

In fact table of MDW, the attributes of sparse range have higher operating frequency, which makes more difficult for multidimensional queries. So gather all sparse attributes to further improve query performance, and this collecting structure will be stored in memory which includes attribute, standard range, min–max value, the address of query vector, bin boundaries, the address of bins and the address of gathering table, etc.

Two-dimensional structures are added, such as in Table 65.1, which describes the starting position of the corresponding attribute values gathering and the number of elements, and Table 65.2 shows attributes gathering. The columns represent different attributes and rows indicate the same ID.

Table 65.1 Starting position

Bin1	Bin2
1	1
2	2
1	2

Table 65.2 Result of gathering

RID	A_1	A_2
1	45	1.2
4	39	0.8
5		2.9
9	52	2.4

When different attributes gather, the column value will be inserted directly due to the same ID row; otherwise, a new row will be added in a proper position, and the value will be filled in except the first column. It certainly is inevitable to waste some memory space in this way. But there is a certain relationship between the abnormal values and some ill signals. If one value in the same ID is unusual, the others are also abnormal, which can be neglected in medical database. And in Sect. 65.4.2, the special assessment is conducted in terms of this situation.

65.4 Experimental Results

The experiment environment is an Intel(R) Core™ Q9550 CPU @2.83 GHz. The operating system is Windows 7. FastBit and Visual Studio 2010, respectively, are database and debugging environment. A set of experiments are run using the illness early warning datasets to verify the algorithm performance. There are nearly 1,10,000 data with multiple attributes columns. The gathering Table 65.2 will be created including the Extracting 5 attributes columns, the standard range and min-max values that users enter the data user input values. According to these known conditions, different bins of various attributes have been initialized. For the particularity of medical data, the numbers of bins are less than 50. So the constraint k which denotes the max numbers of bins is 50. They are the starting conditions of our experiments. At first, the elapsed time for building the single attribute value and processing query will be shown in Tables 65.3 and 65.4. It can be found that using index could achieve nearly 60 % performance. For the time performance that will be improved with the check candidates reducing, the number of check candidates is

Table 65.3 Elapsed time for building index

Dataset	Read data (s)	Building index (s)	Write index (s)
10.2M	0.09	7	0.08
8G	86	1732	74

Table 65.4 Elapsed time for processing query

Kind of query	Success (ave) (s)	Fail (ave) (s)	Success (index) (s)	Fail (index) (s)
Single	7	32	0.09	0.03
Range	12	32	0.01	0.03

also recorded and the space is tested after the indexes of 5 attributes were built and gathered.

65.4.1 Time Performance Analysis

In the early warning of a serious illness known datasets, two bitmap indexes are created, among which one is equi-width binning, and the other is self-adaptive boundary binning. The numbers of check candidates in the two kinds of algorithms are compared, for they are the main factors to affect the efficiency of query. Twenty queries are offered according to the 5 attributes. Method one creates equi-width bins after reading the min–max values. Count the numbers after scanning data. Method two initializes the bins in units and optimizes. Sum the numbers too. The comparison is shown in Fig. 65.3.

It is clearly that the check candidates dramatically reduce after optimizing. In particular, some usual queries hardly need the examination. The improvement has achieved good results.

65.4.2 Space Performance Analysis

In medical database, the probability is relatively high that a data are abnormal, while others of the same ID are also unusual. This is because all these data denote

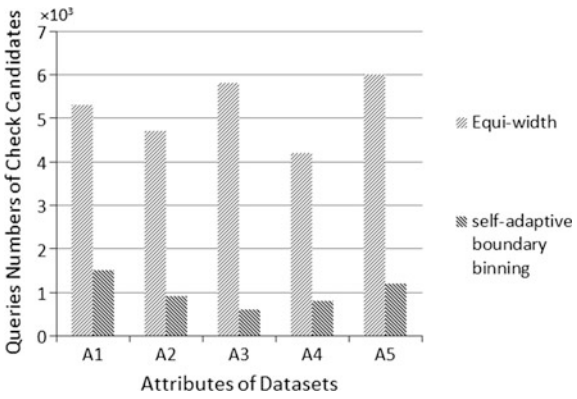


Fig. 65.3 Comparisons between equi-width and self-adaptive boundary binning

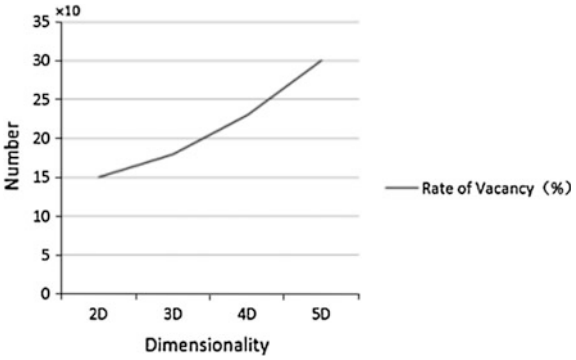


Fig. 65.4 The rate of vacancy after gathering

the status of an object. So in this paper, multidimensional sparse data gathering structure is easy to multiattribute query and it raises the data access efficiency. But it is inevitable to waste storage space. In the experiment, the numbers of empty domain are calculated out after gathering 2*d*, 3*d*, 4*d*, and 5*d* in the extracted records. The trend is shown in Fig. 65.4.

As you see with the increase of the attributes columns, the memory of a gathering table inevitably have some waste. But from the experimental data can be analyzed the waste can be ignored compared with the advance of multiquery efficiency.

65.5 Conclusions

The reasonable application of bitmap index can to a great extent optimize the performance of data warehouse (DW). But the distribution of medical data is uneven in fact table of database and most of data fall into the standard range, and a number of queries aim at sparse range. There has not been a proper algorithm to solve this problem. In this paper, an automatic adaptive optimizing bin boundary algorithm is put forward to accomplish one-dimensional check and two new structures to realize the gathering of sparse value and enhance data access efficiency.

Acknowledgments This research was supported by China National Key Technology R&D Program (2013BAH19F01), Beijing Engineering Research Center of High Reliable Embedded System, Beijing Key Laboratory of Electronic System Reliability and Prognostics, The Project of Construction of Innovative Teams and Teacher Career Development for Universities and Colleges Under Beijing Municipality.

References

1. Rotem D, Stockinger K, KeSheng W (2005) Optimizing candidate check costs for bitmap indices. In: Proceedings of the 14th ACM international conference of information and knowledge management, pp 648–655
2. Ding X, Huang Z (2013) Building medical data warehouse on microsoft business intelligence platform. Comput Appl Softw 30(11):238–242
3. Wu J (2010) FastBit: an efficient compressed bitmap index technology [EB/OL]. 22 Jan 2015, http://crd-legacy.lbl.gov/∼kewu/fastbit/index.html
4. Wang L, Cheng X, Chai Y (2010) The multidimensional query optimization algorithm based on bin bitmap index. Comput Appl 30(8) (in Chinese)
5. Wu K, Stockinger K, Shoshani A (2008) Breaking the curse of cardinality on bitmap indexes. In: Ludäscher B, Mamoulis N (eds) Scientific and statistical database management conference. Springer, Hong Kong, pp 348–365
6. Siqueira TLL, de Aguiar Ciferri CD, Times VC, Ciferri RR (2012) The SB-index and the HSB-index: efficient indices for spatial data warehouses. GeoInformatica 16(1):165–205
7. Howison M, Wu K, Qiang J, Wes Bethel E, Prabhat P, Chou J. et al (2011) Parallel index and query for large scale data analysis. In: Proceedings of 2011 international conference for high performance computing, networking, storage and analysis, pp 1–11
8. Huang Z (2013) Statistical analysis of clinical data in the UIS system design and implementation. East China University Of Science and Technology
9. Otoo E, Wu K (2010) Accelerating queries on very large datasets. In: Scientific data management: challenges, technology, and deployment. Chapman & Hall/CRC Computational Science, USA

Chapter 66
Importance Computing-Based SWoT Ontology Summarization System Design

L. Deng, N. Liu and G. Li

Abstract It can be foreseen that the volume of ontology in Semantic Web of Things (SWoT) is always too large to understand or reason about ontology by machines and users. To solve this problem, an importance computing-based SWoT ontology summarization system (ICSOSS) is proposed, and its framework is designed. This system employed an efficient pragmatic importance and structural importance combined method to compute the importance of RDF sentence, and summaries are produced by extracting the importance of RDF sentence in ontology under a reranking strategy. Some experimental result analyses show that this system can provide users with the best ontology summaries to meet their preferences; hence, it can be more convenient for ontology-related operations in the process of SWoT semantic collaboration, and achieving user-centric personalizes ontology summaries.

Keywords Semantic Web of Things · Importance · Ontology summarization · Structural · Pragmatic

66.1 Introduction

Semantic collaboration can be introduced into Internet of Things (IoT) to form Semantic Web of Things (SWoT) [1, 2] to promote semantic interoperability. One of the main purposes of constructing SWoT is to make machine to understand ontology easily and to reason ontology effectively; therefore, ontology is the core of the SWoT. However, most of the ontology's volume is too large, which is not easy for machine to understand or reason about ontology, and is undoubtedly a huge challenge to solve this problem. Ontology summarization [3] can solve these problems, and it is the process of distilling the most important information from an ontology.

L. Deng · N. Liu (✉) · G. Li
Faculty of Information Science and Technology, Dalian Maritime University, No. 1 Ling Shui Lu, Dalian, China
e-mail: liun571@126.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_66

Summarizing ontology has been used in some tasks, but none of them refer to pragmatic feature. Zhang et al. [4] summarizing ontology based on RDF sentence graph, and the salience of each RDF sentence is assessed in terms of its "centrality" in the graph. Peroni and Motta [5] identified key concepts in an ontology, which best summarize what the ontology is about. Jun [6] compressed the ontology by merging equivalent entities in the ontology. This paper take RDF sentence as the basic unit of ontology summarization; General Bipartite Model and Expanded Bipartite Model of an ontology are proposed to characterize the structure in an ontology; structural importance and pragmatic importance for each RDF sentence are computed respectively. The pragmatic information is a good complement to structural information in summarization.

The related concepts are discussed in Sect. 66.2. Importance computing-based SWoT ontology summarization method is described in Sect. 66.3. The system ICSOSS is designed and described in Sect. 66.4. Evaluation is presented in Sect. 66.5. In the last section, conclusions are made and future work is present.

66.2 Related Concepts

66.2.1 SWoT

The concept of IoT comes from the Electronic Product Code (EPC) technology developed by MIT Auto-ID Center which was founded in 1999 [7]. The collection model of IoT is expressed as follows: *IoT* = {*material information*|*material information is expressed as UID/EPC coding* ∧ *embedded in RFID tags* ∧ *uploaded in the way of non-contact machine-readable* ∧ *stored on the Internet*}.

The material information is uploaded to the Internet by the information provider in different locations; hence, the material information representation is diversified. And it creates inherent contradictions of IoT, i.e., object diversity and subject limitation. To solve this problem, semantic collaboration can be introduced into IoT to form SWoT, i.e., SWoT = IoT + semantic collaboration. The collection model of SWoT is expressed as follows: *SWoT* = {*material information*|<*concept*>*materialinformation*</*concept*>} concept ∈ ontology = <{concept, partial order> *material information is expressed as UID/EPC coding* ∧ *embedded in RFID tags* ∧ *uploaded in the way of non-contact machine-readable* ∧ *stored on the Internet*}.

66.2.2 Ontology Summarization

By the definition of "summary" in natural language processing given in [8], the features of summary include the following: (a) Summaries should be short, no longer than half of the original text(s) and usually significantly less than that; (b) summaries should preserve important information; (c) summaries may be

produced from a single document or multiple documents. In the context of ontology engineering, it is the second feature that fundamentally differentiates ontology summarization from other similar techniques. For example, ontology partitioning and ontology modularization are aimed to reduce the size or complexity of original ontology, instead of keeping important information. However, ontology summarization is a process of extracting knowledge from ontology to produce an abridged version for a specific user(or users) and task(or tasks) [4].

Ontology summarization is divided into user-driven ontology summarization and task-driven ontology summarization. User-driven ontology summarization allows users to quickly understand the target ontology in a large-scale ontology space, for example, online ontology sharing system: cupboard. Task-driven ontology summarization allows ontology to meet a specific task, for example, summary ontology to reduce the size of its volume. In this paper, we can fit the need of user and task.

66.2.3 RDF Sentence

Definition 1 (*B-linked*): Two RDF statements are B-linked if they share common blank nodes. Two RDF statements are said to be B-linked if they are both B-linked to another RDF statements; therefore, the B-linked relation is transitive.

Definition 2 (*RDF sentence*): For an RDF graph $G(T)$, an RDF sentence s is a set of RDF statements, which satisfies the following conditions:

- $S \subseteq T$;
- $\forall i, j \in s, i, j$ are B-linked;
- $\forall i \in s, j \notin s, i, j$ are not B-linked;

RDF sentence is divided into two types: The first is general RDF sentence, and it is formed by a main RDF statement and all the other RDF statements B-linked to it; the second is special RDF sentence, and it has zero or more than one main statements. We take RDF sentence as basic unit of ontology summarization in this paper.

66.2.4 Bipartite Model

Bipartite model (BM) is divided into General Bipartite Model (GBM) and Expanded Bipartite Model (EBM). GBM and EBM are proposed to characterize the structure in an ontology and link between ontologies. The more ontologies the EBM covers, the more abundant information can be utilized in summarization, and the model is controllable in its scale. An RDF sentence is an integrated information unit, which can be a single RDF statement without any blank node or a set of RDF statements connected by blank nodes. Terms are classes and properties defined in the ontology. The GBM and EBM of ontology O are defined as follows:

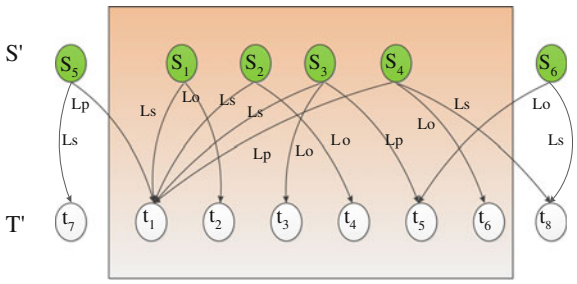


Fig. 66.1 BM of ontology O

Definition 3 (*GBM of an ontology*): $GBM(O) = \langle T, S, L_s, L_p, L_o \rangle$, where T is the set of terms representing all the terms defined in O; S is the set of RDF sentences; L_s, L_p, and L_o are three sets of links from S to T, where L_s denotes a RDF sentence that refers to a term as its subject, $L_p(L_o)$ denotes a RDF sentence that refers to a term as one of its predicates (objects).

Definition 4 (*EBM of an ontology*): $EBM(O) = \langle T', S', L'_s, L'_p, L'_o \rangle$ is derived from $GBM(O)$ under an expansion rule with a specified expansion steps n. After step k ($0 < k \leq n$), the expanded set of terms and RDF sentences are as follows: $T'_k = T'_{k-1} \cup \{t | t \in \text{Term}(S'_{k-1})\}$ and $S'_k = S'_{k-1} \cup \{s | s \in \text{Sentence}(T'_{k-1})\}$. At the beginning, $T'_0 = T, S'_0 = S$, and finally, $T' = T'_n, S' = S'_n$. $\text{Term}(S'_k)$ is the set of terms referred by at least one term in S'_k; and $\text{sentence}(T'_k)$ is the set of RDF sentences referring to at least one term in (T'_k).

The EBM of ontology O is shown in Fig. 66.1. GBM is in the deep colored frame, which includes four RDF sentences and six terms. Terms and RDF sentences outside the frame are the result of one-step expansion. The upper part S_1–S_6 is a set of RDF sentences, and the lower part t_1–t_8 is a set of terms. L_s indicates $t_i (1 \leq i \leq 8)$ is a subject of $S_j (1 \leq j \leq 6)$.

The purpose of building the EBM is to utilize the neighboring structure information to compute the importance of RDF sentence declared in the GBM. The BM is more intuitive compared to the RDF sentence graph.

66.3 Importance Computing-Based SWoT Ontology Summarization Method

Importance computing-based ontology summarization method takes RDF sentence as the basic unit in ontology summarization, characterized ontology by building a BM, as shown in Fig. 66.1. The importance of each RDF sentence is a linear summation of both structural and pragmatic importance, summaries are produced by extracting the important of RDF sentence in ontology under a reranking strategy,

and a user may provides the expected length of the final summary. The importance of each RDF sentence is computed as follows.

66.3.1 Compute Structural Importance of RDF Sentence

Given a GBM and EBM of ontology O, the structure importance $S(s)$ of each RDF sentence s can be assessed by its hub value $h(s)$. The authority value $a(t)$ of each term t in T′ and the hub value of each RDF sentence in S' are computed respectively. The specific compute method is shown in Eqs. (66.1)–(66.3).

$$S(s) = h(s) \tag{66.1}$$

$$h(s) = \sum_{t_1 \in T', (s,t_1) \in L'_s} w_s * a(t_1) + \sum_{t_2 \in T', (s,t_2) \in L'_p} w_p * a(t_2) + \sum_{t_3 \in T', (s,t_3) \in L'_o} w_o * a(t_3) \tag{66.2}$$

$$a(t) = \sum_{s_1 \in S', (s_1,t) \in L'_s} w_s * h(s_1) + \sum_{s_2 \in S', (s_2,t) \in L'_p} w_p * h(s_2) + \sum_{s_3 \in S', (s_3,t) \in L'_o} w_o * h(s_3) \tag{66.3}$$

where w_s, $w_{p,}$ and w_o are different link weights for L_s, $L_{p,}$ and L_o separately; if the term is the subject of RDF sentence, the $h(s)$ contributes more to the authority of a term. We set the link weights to $w_s = 0.8$ and $w_p = w_o = 0.1$ [9].

66.3.2 Compute Pragmatic Importance of RDF Sentence

Current research on the ontology summarization is mainly limited to the semantic level, many implicit information cannot be exhumed, and only the combination of semantic and specific context can properly reflect the essence of language comprehension, i.e., pragmatic [10]. The biggest feature of processing techniques based on pragmatic is the ability to simulate the human use of language information to solve problems. We will solve many problems that the existing technique cannot be solved [11].

The pragmatic importance $P(s)$ of an RDF sentence s is a weighted average of the popularity of each term t it refers to. For each term in ontology, count the number of Web sites $m(t)$ populating terms t. We refer the pragmatic statistics of a term as the statistics of its instantiation, and we can retrieve the pragmatic statistics of each term from the semantic Web index. The pragmatic information is introduced into ontology summarization to make summaries more humanistic and intelligent, not only can convenient for ontology understanding, but also can meet

user's need. In Eqs. (66.4)–(66.6), $m(t)$ denotes the number of Websites populating terms t and β is the number of terms s refer to.

$$P(s) = \frac{\alpha}{\beta} \tag{66.4}$$

$$\begin{aligned} \alpha = &\sum w_s * m(t_1) + \sum w_p * m(t_2) + \sum w_o * m(t_3) \\ &t_1 \in T', (s, t_1) \in L'_s \quad t_2 \in T', (s, t_2) \in L'_p \quad t_3 \in T', (s, t_3) \in L'_o \end{aligned} \tag{66.5}$$

$$\beta = \{t_1 | (s, t_1) \in L'_s\} \cup \{t_2 | (s, t_2) \in L'_p\} \cup \{t_3 | (s, t_3) \in L'_o\} \tag{66.6}$$

66.3.3 Importance Merger of RDF Sentence

The importance of each RDF sentence $F(s)$ is a linear summation of $S(s)$ and $P(s)$. For a broadly used ontology such as FOAF, the summary of the ontology should be completely determined by the pragmatic importance $P(s)$, because it can be fully understand by the user. In other case, pragmatic should not determine the importance if the ontology is not understand by the user widely. In Eqs. (66.7)–(66.9), σ is a factor balancing the weight of $S(s)$ and $P(s)$, $m(o)$ indicates the popularity of O, and θ is a threshold value for popular ontology. If the popularity of O is greater than θ, pragmatic importance is used to rank RDF sentence. Otherwise, $S(s)$ and $P(s)$ are both considered in ranking.

$$F(s) = (1 - \sigma)S(s) + \sigma P(s) \tag{66.7}$$

$$m(o) = \sum_{t \in T} m(t) \tag{66.8}$$

$$\sigma = \begin{pmatrix} 1 & \text{if} \quad m(o) > \theta; \\ \frac{m(o)}{\theta} & \text{if} \quad m(o) \leq \theta \end{pmatrix} \tag{66.9}$$

66.4 Design of ICSOSS

66.4.1 Framework of ICSOSS

The overall design framework of ICSOSS is shown in Fig. 66.2. This system has four main modules: ontology pretreatment module, importance computing module, ontology summaries optimization module, and semantic reserved detection module.

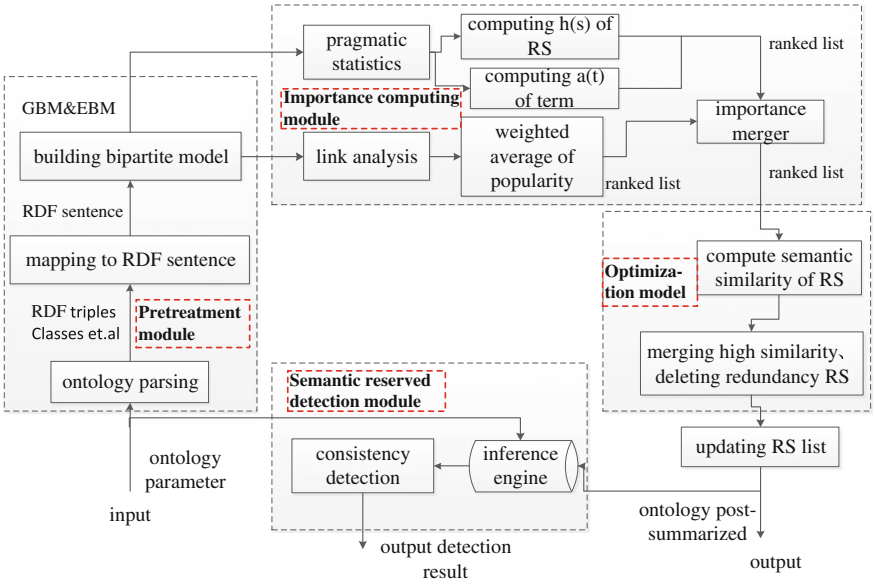


Fig. 66.2 Framework of ICSOSS

66.4.2 Ontology Pretreatment Module

Ontology pretreatment module is mainly responsible for preprocessing the ontology An ontology can be mapped to a set of RDF sentences. Facts or axioms of atomic class are usually mapped to general RDF sentences, and others are mapped to special RDF sentences. BM (including GBM and EBM) is build to characterize the structure in an ontology, in order to facilitate the implementation of importance computing module. The workflow of ontology pretreatment module is shown in Fig. 66.3.

After ontology pretreatment module, BM is build and importance will be computed in the next module.

66.4.3 Importance Computing Module

Importance computing module is mainly responsible for computing the importance of RDF sentences. The summary is formalized as multi-objective optimization problem, and the feature of link analysis and pragmatic statistics are both concerned. The computing methods of $F(s)$, $S(s)$, and $P(s)$ are described in Sect. 66.3. The workflow of importance computing module is shown in Fig. 66.4.

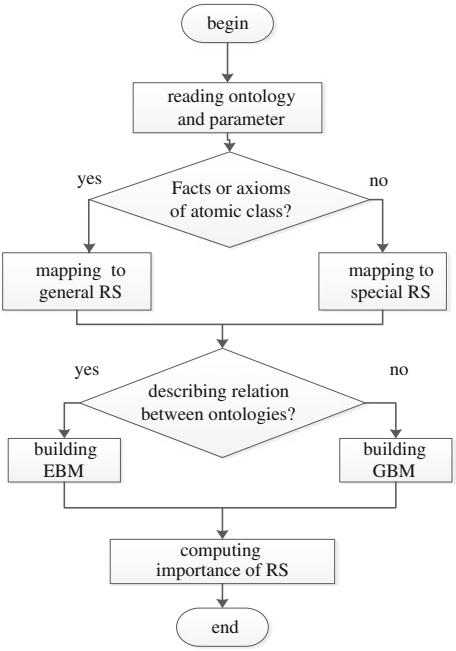


Fig. 66.3 Workflow of ontology pretreatment module

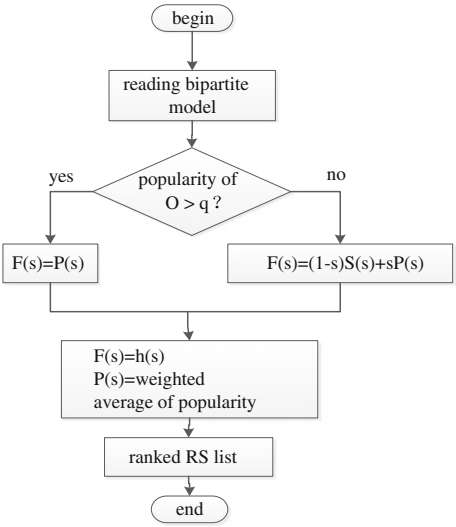


Fig. 66.4 Workflow of importance computing module

After importance computing module, ranked RDF sentence list is produced, and optimization will be implemented in next module.

66.4.4 Summaries Optimization Module

Ranked RDF sentence list is produced after importance computing module; however, just extracting top-ranked RDF sentences is not enough to make the summary informative. Extracted ontology summaries are often redundant if many important RDF sentences talk about similar things in the original ontology. We design a new semantic similarity algorithm SSA to reduce information redundancy in ontology summaries, where w_{ss}, w_{pp}, $w_{oo,}$ and w_{so} are respective weights attributed to subject–subject, predicate–predicate, object–object, and subject–object comparisons. To ensure that our method will stay between 0 and 1, we assume that $w_{ss} + w_{pp} + w_{oo} = w_{pp} + 2 * w_{so} = 1$. Let $\text{sim}_{ss}(s_1, s_2)$, $\text{sim}_{pp}(s_1, s_2)$, $\text{sim}_{oo}(s_1, s_2)$, $\text{sim}_{s_1o_2}(s_1, s_2)$ and $\text{sim}_{s_2o_1}(s_1, s_2)$ be respectively the similarity between Subject(s_1) and Subject(s_2), Predicate(s_1) and Predicate(s_2), Object(s_1) and Object(s_2), Subject (s_1) and Object(s_2), and Subject(s_2) and Object(s_1). SSA is shown as follows, and the workflow of summaries optimization module is shown in Fig. 66.5.

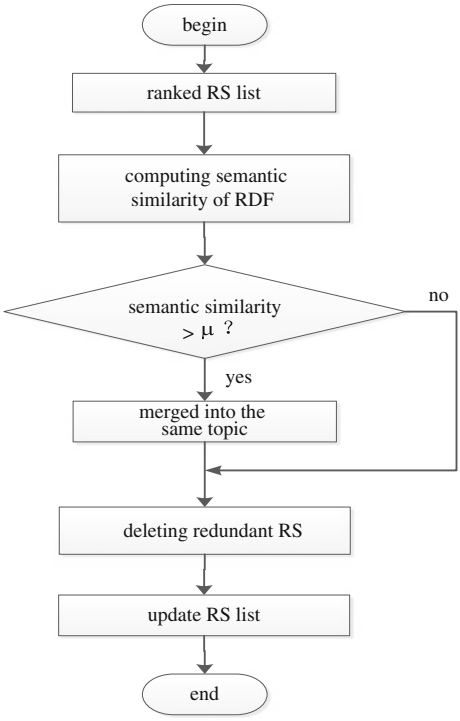


Fig. 66.5 Workflow of summaries optimization module

Algorithm: SSA

INPUT:ranked RDF sentence list R;

OUTPUT:RDF sentence set *S* after optimization;

1 extract RDF sentence set in $O{:}R=\{r_1,r_2,r_3\ldots r_n\}$;

2 $\forall r_m, r_n \in R$

3 if $semantic_{sts}(r_1,r_2)+semantic_{oto}(r_1,r_2)\geq$

4 $semantic_{s1to2}(r_1,r_2)+semantic_{s2to1}(r_1,r_2)$

5 $semantic(r_1,r_2)=V_{sts}*semantic_{sts}(r_1,r_2)+$

6 $V_{ptp}*semantic_{ptp}(r_1,r_2)+ V_{oo}*semantic_{oto}(r_1,r_2)$;

7 else

8 $semantic(r_1,r_2)=V_{sto}*(semantic_{o1to2}(r_1,r_2)+$

9 $semantic_{s2to1}(r_1,r_2))+ V_{ptp}*semantic_{ptp}(r_1,r_2)$;

10 if $semantic(r_1,r_2)>\theta$

11 merger r_1, r_2 ;

12 else put r_i into S;

13 end

14 end

15 end

16 obtain RDF sentence set *S* after optimization;

17 update RDF sentence set.

After summaries optimization module, the process of ontology summarization is over. Inputting original ontology and post-summarized ontology to the inference engine verifies whether they are semantic equivalent or not. This part will be described in next module.

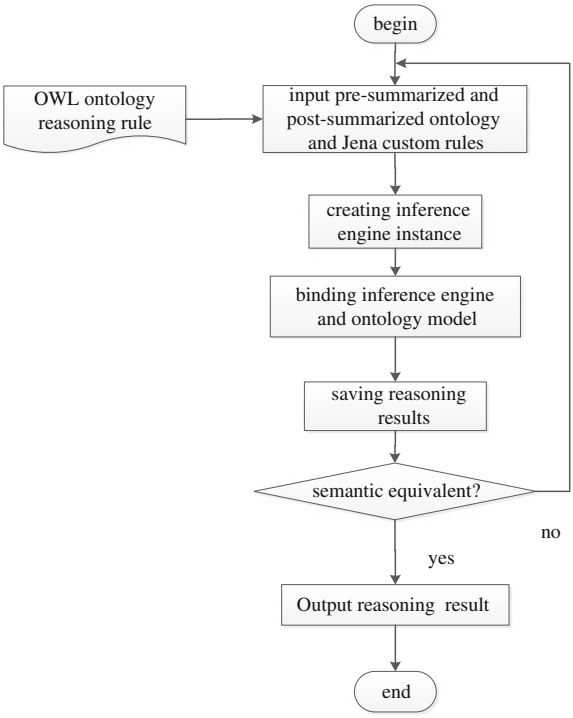


Fig. 66.6 Workflow of semantic reserved detection module

66.4.5 Semantic Reserved Detection Module

The summaries whether can preserve the important information is the main feature that fundamentally differentiates ontology summarization from other similar techniques; hence, it is important to verify whether they are semantic equivalent or not. Jena inference engine is used to reasoning, and the workflow of semantic reserved detection module is shown in Fig. 66.6.

Detection is over if the result is consistent; otherwise, restart the process of ontology summarization.

66.5 Analysis and Evaluation

66.5.1 Comparison Analysis

This paper describes the hierarchical structure and function of ICSOSS, and this system is analyzed and compared with other ontology summarization system in the following aspects.

Table 66.1 Comparison of ICSOSS and other system

System name	Method	Feature selection	Semantic reserved detection	User customize size
PDMS	Summarizing ontology-based schemas	Centrality frequency	No	No
Ontology's Abox summarization system	Equivalent relation based	Equivalent relation	Yes	No
Iwin	Semantic analysis based	Relevance of the information	No	No
OWLSumBRP	BRP based	Relevance of the concepts	No	Yes
ICSOSS	Importance based	Structural Pragmatic	Yes	Yes

As shown in Table 66.1, ICSOSS has obvious advantages. Only ICSOSS with a combination of multi-objective optimization, semantic reserved detection and personalized user customize size. This system can provide users with the best summaries to meet user's preferences.

66.5.2 System Evaluation

This system is evaluated by summarization time and consistency check. Summarization time represents the time that the system executes the importance computing module used, which can be drawn by the system. Consistency check is obtained by the results of semantic reserved detection module. We performed an evaluation using four different ontologies: *financial*, *music*, *aktorsportal,* and *biosphere*. The total number of terms in the test ontologies are 188, 91, 247, and 87, respectively. The evaluation of ICSOSS is shown in Table 66.2.

Table 66.2 Evaluation of ICSOSS

Test ontology	Summarization time t (s)	Consistency check
Financial	1.01	Yes
Music	0.57	Yes
Aktors portal	1.69	Yes
Biosphere	0.41	Yes

(http://www.larflast.bas.bg/ontology; http://pingthesemanticweb.com/ontology/mo/musicontology.rdfs; http://www.aktors.org/ontology/porta; http://sweet.jpl.nasa.gov/ontology/biosphere.owl).

The experimental results show that the summarization time is proportional to the number of terms in ontology. The fewer the number of terms, the shorter the time is needed; after reasoning consistency check, the results confirmed that the presummarized ontology and post-summarized ontology are semantic equivalent. From the above-mentioned results, we can conclude that the ICSOSS is effective and feasible.

66.6 Conclusion

Based on existing abroad research on SWoT ontology and ontology summarization, a SWoT ontology summarization method based on importance computing is proposed, and a system (ICSOSS) based on this method is designed. Importance computing module is the core of this system.

This system employed link analysis and pragmatic statistics to compute the importance of RDF sentences. Public ontology is used to test this system. Analyses of some experimental results show that this system under importance computing method has greater improvement than current system in functionality and performance, which makes more convenient for ontology-related operations in the process of SWoT semantic collaboration. In addition, the study needs to be further improved: firstly, how to modify the ICSOSS to improve the accuracy of summaries; secondly, how to improve the summarization based on the feedbacks from the users.

Acknowledgment This work is supported by the National Natural Science Foundation of China (Grant No. 61371090).

References

1. Huang Y, Li G (2010) SWoT: strategy for internet of things' intrinsic contradiction. Appl Res Comput 27(11):4087–4090 (in Chinese)
2. Pfisterer D, Romer K, Bimschas D et al (2011) SPITFIRE: toward a semantic web of things. IEEE Commun Mag 49(11):40–48
3. Queiroz-Sousa PO, Salgado AC, Pires CE (2013) A method for building personalized ontology summaries. J Inf Data Manage 4(3):236–250
4. Zhang X, Cheng G, Qu Y (2007) Ontology summarization based on rdf sentence graph. Proc Int Conf World Wide Web 707–716
5. Peroni S, Motta E (2008) Identifying key concepts in an ontology, through the integration of cognitive principles with statistical and topological measures. Semant Web Lncs 242–256
6. Jun F (2013) Ontology lossless summarization based on equivalence [EB/OL] (2013-01-16)
7. Huang Y, Li G (2011) IoT: semantic, nature and classification. Comput Sci (1):31–33 (in Chinese)

8. Li N, Motta E, d'Aquin M (2010) Ontology summarization: an analysis and an evaluation. Int Workshop Eval Semant Technol
9. Zhang X et al (2009) Summarizing vocabularies in the global semantic web. J Comput Sci Technol 24(1):165–174
10. Zhao L, Zhao K, Xu W (2007) Research and realization of pragmatics based on given domain. Discrete mathematics. Comput Technol Dev 6(17):47–52 (in Chinese)
11. Li L, Zhou Y, Zhong Y (2006) Pragmatic information based NLP research and application. CAAI Trans Intell Syst 2(1):1–6 (in Chinese)

Chapter 67
Survey on the Railway Telematic System for Rolling Stocks

Xiangyang Lu, Sheng Shan, Guoping Tang and Zheng Wen

Abstract This paper presents a survey on the development and application of the railway telematic system for rolling stocks. The usage of the system among main line locomotives, high-speed trains, and metros in Europe, America, Japan, and China is reviewed. In addition, this paper predicts the developing trend of the system.

Keywords Rail transit · Telematic · Intelligent · Big data · PHM

67.1 Introduction

China's rail transit is in a period of rapid development. In accordance with the national *Medium- and Long-term Railway Network Plan* (revised in 2008) [1], the total length of China's high-speed railways will reach 16,000 km by 2020, while the total railway operating mileage is expected to reach 120,000 km. Similar prosperity is witnessed in the metro market: By the end of 2014, the total 3173-km metro lines are in operation in 22 cities and more than 4073-km metro lines are under construction in 40 cities, which counts for a total investment of 300 billion RMB in 2014 [2]. The growth in rail mileage, the increase of railway vehicles, and the improvement of operating speed imply the ever-increasing requirements on reliability, availability, maintainability, and safety (RAMS) of railway vehicles as well as the onboard equipment. Railway operators are in great need of means to assist drivers, maintenance personnel, and management personnel in comprehensive monitoring the running status of railway vehicles and onboard equipment in real time; processing the online faults and events in time; enhancing the utilization ratio of the railway vehicles and the rail network; and improving the maintenance proficiency of railway vehicles and onboard equipment. These are the important goals of the railway telematic system construction both in China and abroad. With the

X. Lu · S. Shan · G. Tang · Z. Wen (✉)
CSR Zhuzhou Institute, Shidai Rd., Zhuzhou 412001, Hunan, China
e-mail: wenzheng@csrzic.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_67

in-depth integration and well utilization of the emerging information and communication technologies, i.e., Internet of things, cloud computing, and big data, railway telematic applications such as the remote monitoring, the fault diagnosis, the prognosis and health management of the rolling stocks, and onboard equipment have been evolving. And, in fact, the research and application of railway telematic system is a hot topic in the railway industry [3].

67.2 Development and Application of Railway Telematic System for Rolling Stocks Abroad

Innovation and application of the railway telematic technologies, especially the remote monitoring, fault diagnosis, and maintenance support, are started relatively early in Europe, the USA, and Japan. Network-based telematic systems have also been widely adopted in these countries. Recently, emerging technologies, such as the mobile Internet, big data, 4G wireless communications, and cloud computing, are also gradually adopted in the field of railway telematics, which improves the telematic system performance and increases the convenience of the system usage.

67.2.1 Europe

Since the 1990s, driven by the Association of the European Rail Industry (UNIFE [4]) , European states carried out a series of research and applications of the railway telematic system for rolling stocks.

As early as 1996, a railway open system interconnection network (RoSin [5]) was built in Europe, in which an open architecture of train network system was established. From 1998 to 2002, an open railway maintenance system was raised by the European railway open maintenance system (EuroMain [6]) project, in which vehicles were maintained through the Internet; maintenance personnel were able to monitor and manage the railway locomotive and equipment at any place in any time; and the experts in different places could cooperate and schedule the maintenance tasks in advance, so as to maintain railway equipment timely and efficiently.

From 2000 to 2004, European countries carried out the TrainCom [7] project and focused on the implementation of the telematic systems. By integrating onboard network, GSM wireless connection, and the Internet networking technologies, a train-to-ground communication framework was established, which provided a general system architecture prototype and some basic elements for the remote monitoring and maintenance support system.

From 2005 to 2010, the EU invested 20 million Euros to carry out the InteGRail [8] project which integrates the surveillance data and diagnostic data with other

telematic systems based on modular system architecture and enables information sharing among them.

In the most recent SHIFT2RAIL [9] project, launched by the EU between 2014 and 2020, 920 million Euros will be invested into the field of rail transit in order to create an intelligent, green, and integrated transport system through research and application of innovative technologies.

Benefited from the railway telematic systems' construction experience and interoperability design consensuses, European countries also positively promoted the international standardization process of the train network and telematic systems through the Ninth Technical Committee (TC9 [10]) of the International Electrotechnical Commission (IEC). WG43 working group, under the leadership of the IEC/TC9, has been filing the IEC61375 series standards for train communication network (TCN), systematically defining the general structure of TCN, its functionality and interoperability. IEC61375-2-6 and IEC61375-2-3 provide the basis for data communication of railway telematic systems from the train-to-ground communication and onboard communication profile's prospective, respectively. WG46 working group has been publishing the IEC62580 series standards for onboard multimedia and telematic system. Specifically, IEC62580-1 standard defines the general architecture of onboard multimedia and telematic system, and IEC62580-2–IEC62580-5 series standards define the functional architecture, performance requirements, and interoperability requirements of various services in detail, including video surveillance/CCTV, driver- and crew-orientated services, passenger-orientated services, and train operator- and maintainer-orientated services. Figure 67.1 depicts a summary of IEC international standards related to the railway telematic system for rolling stocks.

Most of the well-known European rail suppliers have been rolling out their telematic system in accordance with the IEC 61375 and IEC62580 series standards. Adtranz developed the Mitrac CC remote system for remote monitoring, fault diagnosis, and maintenance support of their traction and control system-Mitrac. Mitrac CC remote system is composed of a client terminal, ground stations, and mobile stations, wherein the communication between mobile stations and ground stations is connected through wireless communication networks, i.e. GSM, WLAN, and CDPD. The ground stations remotely obtain the diagnostic data, environmental data, and status data of trains through the wireless network, and store the train information in the database at the ground stations for user access. End users at different places mainly log onto the ground stations through the Internet and intranet to monitor a specific train or to get the record data using Web service.

In 2010, Bombardier has launched a set of Web-based telematic service: MyBTfleet [11] and released its APP version in 2013. MyBTfleet provides GPS-based train location, trace tracking, train maintenance plan, and customer record services for operators; report of intervention, catalog, and manual achieve service for maintenance personnel; and online fault diagnostics, real-time monitoring, and Orbita data center access for engineers. In addition, it is possible for MyBTfleet to provide energy management and driver desk simulator service. In January 2015, Bombardier further introduced its MyBTfleet gateway 2.0 which can

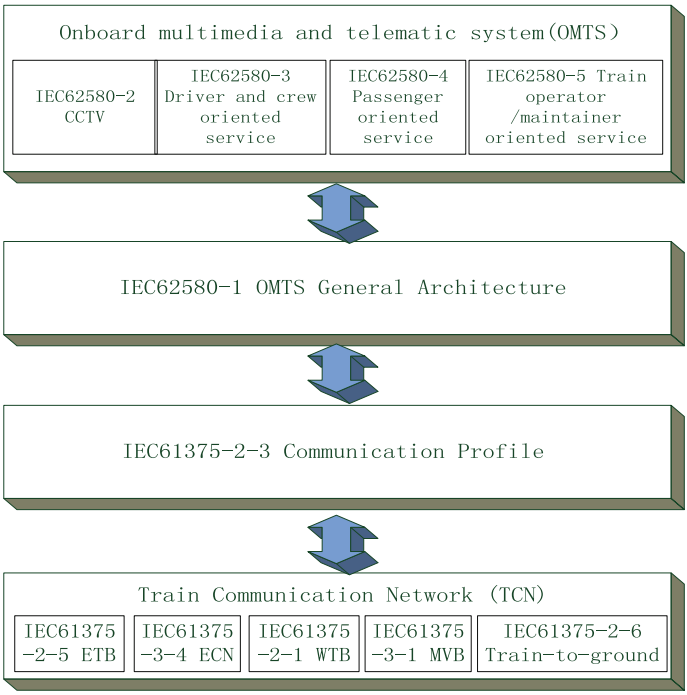


Fig. 67.1 IEC series international standards of railway telematic system for rolling stocks

seamlessly access the onboard TCMS system and provide two-way real-time data interaction between the train and MyBTfleet control center. As a result, technical personnel at the control center could acquire the operating status in real time, upload dedicated alerts to the monitor, and get immediate feedback in case of parameter deviation.

Siemens launched a set of wireless data transmission, remote monitoring, and analysis system (EFLEET system) in 2007. The system consists of an onboard terminal, train-to-ground communication networks, and a ground processing system. EFLEET system provides functions, including locomotive GPS positioning, remote faults alarm and remote operating data reviewing, data records downloading, offline analysis, and expert diagnosis of the faults and event records. In 2014, Siemens published a set of more comprehensive system, the 100 % RAILability. By adopting remote monitoring, intelligent data processing, big data analysis, and predictive maintenance technologies, this system is extended to provide product maintenance, spare parts service, qualification certification, and operations management services.

Alstom's early product was as follows: AGATE LINK system provides seamless connection of the onboard control systems and ground systems. In 2006, Alstom launched TrainTracer for real-time remote monitoring of train onboard system to meet the maintenance requirement. In 2008, Alstom further launched the ETRAIN

system for remote monitoring and analysis. ETRAIN is composed of onboard terminal COMBOX, train-to-ground data transmission networks, i.e. GSM/GPRS and WLAN, and ground processing system. ETRAIN provides the onboard data records downloading, off-line analysis of locomotive failure and event records, and locomotive GPS positioning functions. In 2014, Alstom launched a predictive maintenance tool called HealthHub which integrates TrainScanner, TrainTracer, Motes, and Alstom cloud services. In this system, TrainScanner is a diagnostics portal that automatically measures the condition of wheels, brake pads, pantograph carbon strips as well as train integrity when a vehicle passing through it. Motes monitors the bearing vibration and temperature using wireless sensors. TrainTracer remotely monitors the onboard systems, the track, and the catenary. All different type of data are collected and analyzed in the Alstom cloud, which in turn predicts the health condition and remaining useful life of the given parts and assists in planning spare parts and maintenance tasks. It is worth noting that all results are published by Alstom cloud services, and customers can also access to the system via the Internet. HealthHub has been providing comprehensive maintenance service for the Virgin Trains running on the West Coast main line of the UK.

67.2.2 The United States of America

General electric (GE) takes service as an important driving force for the development of the company. As one of the most popular technology, remote monitoring, and fault diagnosis, developed by GE research center, has been applied in various fields including air plane, rolling stocks, and medical equipment.

Early in 2000, GE introduced the RM&D system for locomotive remote monitoring and fault diagnosis. GE's data center brings together the status monitoring information of ten thousands of locomotives worldwide, while GE's technical experts are able to provide remote online technical support based on real-time status, performance information, GPS positioning, and weather and geographical data information. In addition, RM&D can also provide customer-specific maintenance recommendations according to customized alerts of customers (such as geographical location and energy consumption), plan in advance the resources and materials needed along the routes, and provide condition-based predictive maintenance. RM&D system has been applied in North America and also in the Qinghai–Tibet railway line in China. RM&D has been providing remote monitoring and fault diagnosis services for more than 10 years and has been deployed in large number of locomotives and approximately 100 maintenance facilities. It handles more than 100,000 message data processing per day and generates more than 100 maintenance instructions.

In 2014, GE integrated its railway transportation information-related products and proposed a comprehensive set of solutions called RailConnect 360, aiming at assets management, network, supply chain, and operations. In the field of asset management, many solutions were rolled out: The ExpressYard is provided for

locomotive maintenance management, repair, and inspection; the LocoCAM is launched for real-time video surveillance; the RailDOCS is proposed for wayside asset management; the Trip Optimizer is the onboard train control system for energy efficiency management, etc. This set of solutions manages and optimizes the utilization, performance, running, stopping over, maintenance, and life cycle of GE locomotives. It is worth to note that remote monitoring of locomotive health status (including locomotive traction, control, cooling system, and third-party electronic devices), locomotive tracking, online fault prevention, maintenance scheduling, and spare parts and maintenance task planning are all accessible through the Internet.

67.2.3 Japan

Mitsubishi's train information management system (TIMS) has sophisticated monitoring and diagnosis capabilities on the multiple unit trains, the onboard equipment, and even some electric lines and components. Besides TIMS, Mitsubishi has also developed maintenance-oriented systems, including depot information management system (DIMS), mainly completing the deploy plan, depot job plan, route control and automatic dispatch; maintenance information management system (MIMS), mainly completing account management, maintenance plans, maintenance management, fault management, as well as materials and spare parts management; remote information management system (RIMS), mainly completing the information transfer between the train and the ground, train state monitoring, and fault recovery support; automatic detection and automatic testing equipment for pantograph, wheel, axle, and integrated circuit testing; and the education and training system (CAI), i.e., simulation driving.

Toshiba's traffic management system (TMS) is composed of the traffic control system and the railway operation management system. The former one is used for monitoring and dispatch, scheduling, and engineering work management, the latter one is used for rail operation plan, fleet maintenance, and information publishing. Remote monitoring, fault diagnosis, maintenance scheduling, and record management services are integrated in fleet maintenance subsystem. In this subsystem, fault prognosis and predictive maintenance based on historical data analysis and trend prediction are also emphasized.

67.3 Development and Application of Railway Telematic System for Rolling Stocks in China

With the development of Beidou satellite and wireless communication technology, the performance of the railway telematic system for rolling stocks in China has been developed to meet the rail operator's operation, maintenance, and management

requirement. China's railway telematic system has been applied mainly in the main line locomotives, electrical multiple unit (EMU) trains.

67.3.1 Main Line Locomotives

The development of the railway telematic system for locomotive remote monitoring, fault diagnosis, and maintenance support in China can be traced back to the beginning of this century. In 2004, the Xi'an Railway Subbureau, Zhuzhou Electric Locomotive Research Institute (ZELRI), Xi'an Jiaotong University, and Central South University developed a remote monitoring system to investigate the operating status of SS7E electric locomotives. The system is composed of onboard equipment, train-to-ground wireless communication system, and ground processing equipment. The real-time monitoring data and locomotive positioning information were transmitted using Beidou satellite positioning and communications technology, and an elementary train-to-ground telematic system was built for the operators on the ground to track the locomotive, monitor the status of the locomotive and the leakage current of the main and auxiliary circuit.

In 2005, the Ministry of Railways (MOR) issued the *General Planning for Railway Informationization*. In the same year, the transport bureau and the information technology center of the MOR issued a *General Program for Locomotive Informationization*. This program defined the general architecture, the infrastructure, the network, the software and hardware specification, composition of the application system, the maintenance management, the objective, and the deployment schedule. In this program, locomotive remote monitoring, fault diagnosis, and maintenance support are categorized within the range of locomotive operation resource management system, which laid the foundation for the development of the railway telematic system for rolling stocks in China.

From 2005 to 2007, the Lanzhou Railway Bureau, ZELRI, and the Information Technology Center of the MOR carried out the system development work in the fields of network-based locomotive remote monitoring, fault diagnosis, and maintenance support. As the output of the project, the developed onboard equipment was applied in six SS7E electric locomotives, and the ground system was deployed. The train-to-ground communication of this system is realized by using GSM for online data transmission and using WLAN for offline bulk data transfer. Noticeably, the ground system is connected to the MOR private network, and this preliminary work formed the basic framework for the ground system.

From 2007 to 2009, ZELRI and Wuhan Zhengyuan Electric, under the coordination of the Locomotive Joint Office of the MOR, carried out research on the remote monitoring and fault diagnosis for HX locomotives, and applied commissioning on HXN3, HXN5, and HXD3. The train-to-ground communication of this system is realized using GPRS.

In 2009, under the coordination of the Locomotive Joint Office of the MOR, a project team composed of the Information Technology Center of the MOR, ZELRI,

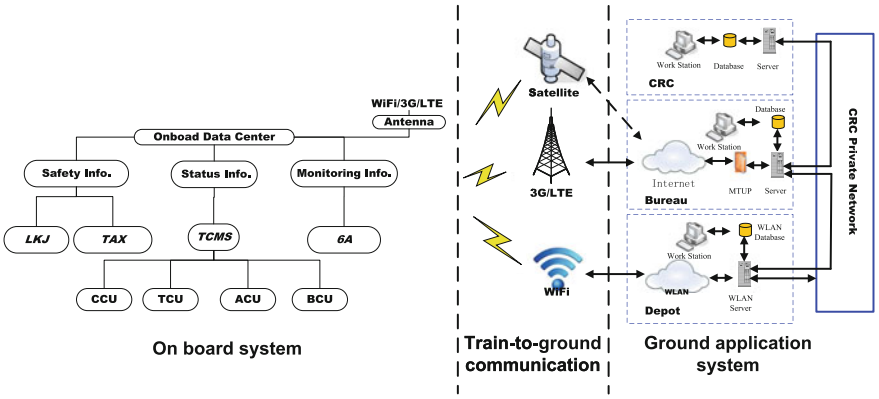


Fig. 67.2 System diagram of CMD system; *LKJ* Locomotive operation monitoring equipment, *TAX* Locomotive safety information integrated monitoring equipment, *6A* Locomotive onboard safety protection system, *TCMS* Train control and management system

and Wuhan Zhengyuan Electric undertook the development and application research of "Chinese locomotive remote monitoring and diagnosis system (CMD system)" project set by the Science and Technology Division of the MOR. The CMD system developed by the project team has been tested on various types of locomotives in Chongqing Depot, Lanzhou Depot, Nanjing Depot, and Wuhan Depot. This series of work formed the technical specifications of the CMD system, enforced the unified framework of the CMD ground system, and built the MOR bureau-depot three-tier network layers as depicted in Fig. 67.2. Thereafter, Locomotive Department of the China Railway Corporation (CRC) has been systematically revising the series of technical specifications for the CMD system. The CMD system has begun commercial application since 2014.

In 2011, under the coordination of Transport Bureau of the MOR, China Academy of Railway Sciences, ZELRI, and Chengdu YD-Tech undertook the development and standardization work of the locomotive onboard safety protection system (6A). The 6A system is composed of 6 safety-related subsystems, including air brake safety monitoring and data record system; fire alarm and data record system; anti-ground detection system for resistance of high-voltage equipment; appliances train power diagnosis and record system; advanced truck defect diagnosis and record system and automatic video display and record system. At present, the 6A system has already been applied in HXD locomotives (see Table 67.1).

Table 67.1 Application of 6A system in the years

Year	New vehicle	Legacy vehicle	Remark
2013	1124	722	Assembly completed
2014	1690	1450	Assembly completed
2015	861	1800	To be completed

67.3.2 Electrical Multiple Unit

In 2009, the MOR proposed to build remote monitoring and diagnosis system for EMU. The China Academy of Railway Science, ZELRI, and other manufacturers undertook the development work of the system and filed the technical specification under the organization of the MOR. The application and commissioning of the system were started in the same year. The system can transmit the status data as well as fault record data of the traction unite, the network and the brake of the EMU to the ground system by means of GPRS, and the other data by means of WLAN to realize remote monitoring and event replay function. At present, this system has been applied in CRH1, CRH2, CRH3, and CRH5 EMUs.

67.3.3 Metro

Metro train operation is mainly managed by use of automatic train supervision (ATS) system. ATS system mainly supervises and controls the running status of trains, including centralized monitoring of running status of trains, automatic route setting, automatic train schedule adjustment, automatic timetable generation, automatic trace records, automatic running data statistics, and automatic report generation. The ATS system also helps the dispatch personnel to manage all trains of the metro line. However, the ATS system lacks some functions, such as rail vehicle as well as onboard equipment operational status monitoring, data recording, and fault diagnosis.

In 2007, Shanghai Shentong Metro Group, Beijing Jiaotong University, Casco, and ZELRI jointly carried out the National 863 Program project: rail transport operation safety critical equipment monitoring, early warning, and emergency response technology. In 2010, Guangzhou Metro Group, Beijing Jiaotong University, ZELRI, and other manufactures jointly carried out the other National 863 Program project: urban rail transit vehicle online monitoring and safety alarming key technology. These projects are approached for the study of remote monitoring and safety alarm in the field of metro system.

The passenger information system (PIS) is also an important part of the railway telematic system in the field of urban metro system. PIS mainly provides travel reference, passenger announcement, and other information services for passengers under normal circumstances, and provide emergency evacuation guidance information for passengers in emergent circumstances.

67.4 Prospect

Along with the improvement of technology, especially the adoption of cloud computing, big data, and intelligent devices, the railway telematic system for rolling stocks has been evolving toward the directions of professional technology advancing, platform integration, and intelligent service upgrading.

67.4.1 Professional Technology Advancing

Traditional railway telematic system for rolling stocks mainly focuses on tracking, dispatching, train-to-ground wireless communication, remote monitoring, and fault diagnosis. These can all be categorized as reactive services. The advance and application of emerging technologies, i.e., artificial intelligence, life prediction, and big data analysis, together with the improvement of the rolling stock, and its onboard equipment design technique have now made the predictive services ready for use. It covers the performance test, fault diagnosis and prediction, and life cycle health management of the rolling stocks and key parts, which is also called the fault prognostics and health management (PHM) for short. This represents the one of the main developing trend of the railway telematic system for rolling stocks. During the transformation from reactive services to predictive services, the beneficial degree of customer will be improved gradually, but the dependency on data and intelligent level of the telematic system shall also be enhanced progressively (see Fig. 67.3).

Additionally, the data collection and transmission capacity improvement of the railway telematic system for rolling stocks accumulate huge amount of data. It is getting feasible for machine learning and big data analysis to leverage these resources and formulate innovative and intelligent techniques, i.e., the optimal control service which jointly considers various factors, i.e., the running status of the locomotive, the characteristics of the track, the safety requirement of the operation, and the efficiency of energy consumption, and assists the driver for locomotive operation automatically. This is also a direction of the railway telematic system professional technology advancing.

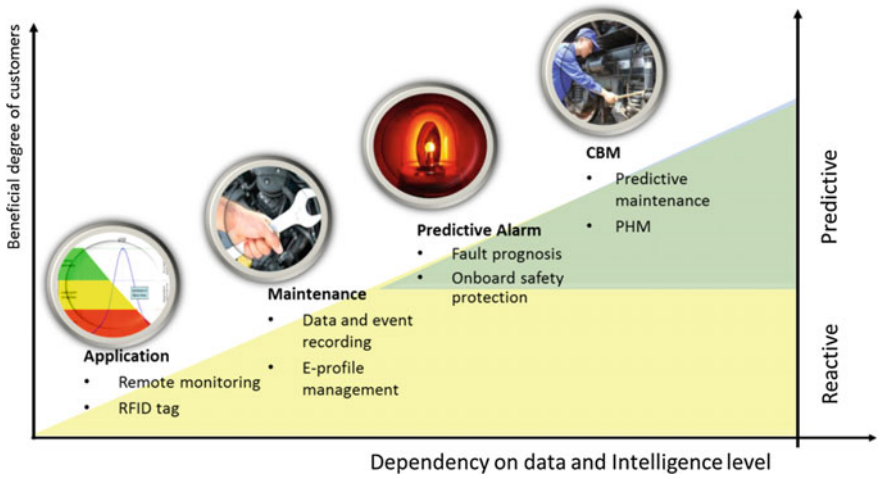


Fig. 67.3 Change from reactive to predictive service

67.4.2 Platform Integration

The other inevitable developing trend of the railway telematic system for rolling stocks is platform integration. Regarding the onboard subsystem, more sensors will be adopted to improve the raw data collection capacity; the train control network and telematic service network are likely to merge so as to improve the bandwidth, reliability, and flexibility of the onboard network; the onboard data storage and analysis capacity will be enhanced; and new technologies will be adopted to improve throughput and reliability of the train-to-ground communication. As for the ground system, offering an open system framework, standardized data interfaces and normative business organization models to achieve the integration and interoperability of the ground infrastructure will be the focus of system construction. Hopefully, the information and functions across different system, domain, and time could be integrated and well utilized.

67.4.3 Intelligent Service Upgrading

As ICT technologies (information, communication, and data) are booming, the applications of such technologies in the rail transit fields will certainly promote the intelligent service upgrading of railway telematic system for rolling stocks. In a long term, this promotion will improve the service convenience and work efficiency, which is likely to transform the vehicle operation, management, and maintenance ecosphere.

The intelligent service upgrading is characterized by adopting big data technology and providing cloud platform for data access, storage, and analysis, so as to enrich the telematic applications along the life cycle management of the rolling stock and its key parts. And of course, Web portal, mobile App, and a variety of other human–computer interactive techniques will be used.

As for the transformation of the vehicle operation, management, and maintenance ecosphere, the intelligent service upgrading of the telematic system will greatly improve the equipment maintainability. The function upgrading and daily maintenance are expected to be more systematic and efficient. With the open platform and communication protocol, providing vehicle operation, management, and maintenance service is getting relatively simplified. Offering such service is likely to become a commercial product, that is, a formation of a new type of business model: “equipment plus services.”

67.5 Summary

The railway telematic system for rolling stocks is built up gradually under the dual driving factors: the advances of computer networks and wireless communication technology, and the rail operator’s demand for rolling stock operation,

management, and maintenance. After years of development and deployment, the general system framework and specifications have been established, and remote monitoring and fault diagnosis service have been applied, though it is still far from satisfying the functional requirement in face of the rolling stocks profession and operation management complexity. There are many works to do to drive the railway telematic system for rolling stocks toward directions of professional technology advancing, platform integration, and intelligent service upgrading. The development of the railway telematic system for rolling stocks is synchronized to the fourth industrial revolution, and it is an indispensable stage of the intelligent rolling stocks involvement. In addition, this development is likely to enrich the rolling stock design techniques, thus revolutionarily transforms the traditional rolling stock operation, management, and maintenance ecosphere.

Acknowledgments This work was supported in part by the China National Science and Technology Support Program under Grant (2015BAG14B00).

References

1. The medium-and long-term railway network plan (revised in 2008) (in Chinese)
2. Urban metro transport report (2014) Available at: http://www.camet.org.cn/hyxw/201505/t20150513_407674.htm (in Chinese)
3. Wang Q, Tang G (2012) Research and design of remote monitoring and diagnosis system for locomotive, locomotive drive, vol 3, pp 42–44
4. UNIFE. http://www.unife.org/
5. Andresi S, Bedogna U, Sacchetto R (1997) User requirements for networked applications in the railway environment, the RoSin project report
6. EuroMain. http://www.euromain.org
7. Renner E (2001) Integrated communication system for intelligent train applications, TrainCom project presentation, Erlangen
8. UNIFE. InteGRail final report (2010)
9. Shift2Rail. http://www.shift2rail.org/
10. http://www.iec.ch/dyn/www/f?p=103:7:0:::FSP_ORG_ID:1248
11. MyBTfleet. http:// www.mybtfleet.com

Chapter 68
Remote Preparation of EPR Entangled Photons and Quantum States

Zhiguo Wang

Abstract Along with the development of quantum information in recent years, people have conducted lots of and thorough researches into it. Because of the non-locality of quantum entanglement, entangled states have achieved an extensive development and application in quantum communication. Entangled photons serve as an effective means to implement quantum communication. This paper introduces the fundamental principles of quantum communication and quantum entanglement and proposes an implementation plan for remote preparation of quantum information.

Keywords Entangled photons · Quantum states · Quantum communication · Remote preparation

68.1 Introduction

The combination of quantum mechanics and information science has led to the emergence of today's quantum information, and two more branches, namely quantum communication and quantum computing. At present, quantum communication is being studied for implementing long-distance quantum communication along the branches of free space and fiber-optical telecommunication [1]. However, by playing a core role in quantum communication, quantum entanglement has become an important resource in quantum information by virtue of its unique quantum properties. Quantum entanglement generally adopts such preparation methods as optical parameters based on the sources of entanglement in a squeezed

Z. Wang (✉)
School of Physics and Electrical and Electronic Engineering,
Huaiyin Normal University, Huai'an 223300, Jiangsu Province, China
e-mail: wzg_hysy@sohu.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_68

state. Along with the preparation of sources of quantum entanglement, the application of quantum entanglement has developed greatly, such as quantum teleportation, quantum dense coding, and entanglement swapping [2]. Now, fiber-optic telecommunication technology has become more and more mature, and fiber-optic transmission technology has reached a considerable level. As found out from long-term experimental and comparative studies, polarized entangled photons enjoy many advantages in quantum communication and quantum cryptography. Therefore, entangled photons are generally applied to long-distance quantum communication and quantum key distribution.

68.2 EPR Paradox and Bell Inequality

In 1935, Albert Einstein, together with B. Podolsky and N. Rosen, published an article entitled "Can Quantum Mechanical Description of Physical Reality Be Considered Completed?" on Physics Review [3]. They completed an ideal experiment and proved that the description of quantum mechanics was incomplete via logical argumentation, which is called EPR paradox. In 1964, Bell gave a quantitative description of this paradox and obtained evidence for the correctness of EPR paradox—Bell inequality.

From the perspective of classic localization theory, Albert Einstein absolutely negated the second viewpoint, thus determining that quantum mechanics was incomplete to describe microcosmos by the use of wave function and state vector. Afterward, Bohr gave a definite answer to the query raised by Albert Einstein et al., and considered they did not understand the above two viewpoints. On the one hand, in quantum mechanics, entangled states formed by two particles were non-localized; entanglement properties made two particles form an inseparable uniform system. On the other hand, measurement of the system in a different way would result in discrepancies in collapse, making the results different. Up to now, the related experiment has been supporting the viewpoint of quantum mechanics.

In 1964, on the basis of local realism, Bell concluded that if local realism is met, then any system should meet the following inequality, i.e., Bell inequality:

$$|P(a,b)| - P(a,c) \leq 1 + P(b,c) \tag{68.1}$$

wherein $P(a,b) = \int p(\lambda)\mathrm{A}(a\lambda)\mathrm{B}(b\lambda)d\lambda$. The sender Alice is abbreviated as A; the receiver Bob is abbreviated as B. λ is a hidden variable; $P(\lambda)$ is its distribution function; $\mathrm{A}(\lambda)$ is the spinning result of relevant particle a as measured by Alice in the direction of a; $\mathrm{B}(\lambda)$ is the spinning result of relevant particle b as measured by Bob in the direction of b.

68.3 Entangled Light Source

68.3.1 Two-Photon Entanglement

When light passes through certain nonlinear crystals, there will be a certain probability that a nonlinear process called parametric down-conversion occurs. In the process of parametric down-conversion, a photon passing through the nonlinear crystal will be split into two photons. Such two photons are called signal light (signal) and idler light (idler), respectively. By selecting the appropriate polarization of incident pumped laser and the moderate optical axis of BBO crystal, BBO used for the experiment makes one emergent photon featured by horizontal polarization (marked "H") and the other featured by vertical polarization (marked "V") [4].

In addition, this process meets momentum conservation constraints and energy conservation constraints.

$$k_{\mathrm{p}} = k_{\mathrm{s}} + k_{\mathrm{i}} \tag{68.2}$$

$$E_{\mathrm{p}} = E_{\mathrm{s}} + E_{\mathrm{i}} \tag{68.3}$$

wherein subscripts p, s, and i represent pump light, signal light, and idler light. In case of incidence of a large number of photons, the distribution of s and i will form a light cone.

By virtue of momentum conservation, aggregate momentum and pump light can be always kept in the same direction. Two possible states $|s\rangle_A|i\rangle_B$ and $|s\rangle_B|i\rangle_A$ exist at the intersections (A and B) of two light cones. From the perspective of quantum mechanics, such two states are actually superposition states as we cannot identify them in principle, namely:

$$\frac{1}{\sqrt{2}}\left(|s\rangle_A|i\rangle_B + |i\rangle_A|s\rangle_B\right) \tag{68.4}$$

Polarized superposition states are:

$$\frac{1}{\sqrt{2}}\left(|H\rangle_A|V\rangle_B + |V\rangle_A|H\rangle_B\right) \tag{68.5}$$

In such a way, two-photon polarization entanglement states commonly seen in the study of quantum information are prepared.

68.3.2 Three-Photon Entanglement

Three-photon entanglement state arises from two spontaneous parametric down-conversion (SPDC) and one sum-frequency generation (SFG), with the detailed

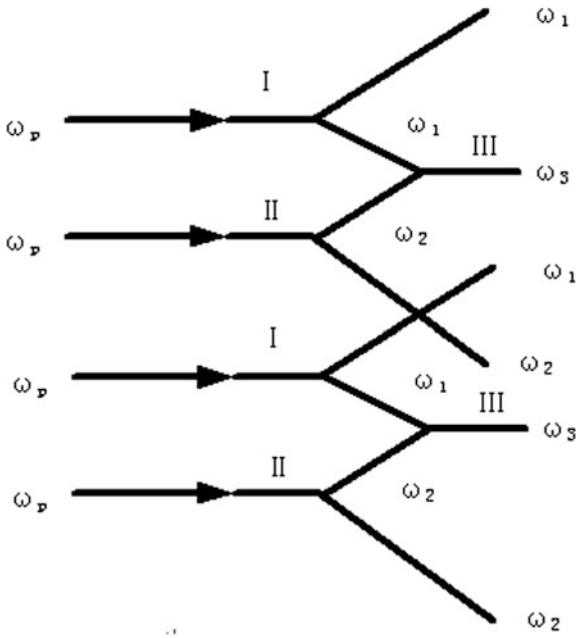


Fig. 68.1 Preparation diagram of three-photon entanglement sources

process shown in Fig. 68.1. By utilizing the phase-matching return behavior in a crystal, a beam of pump light is used to realize three-photon entanglement. In the process of SPDC, pump light intensity is stronger than the signal and idler light intensity. Therefore, pump light can be described with classic plane wave, while signal light and idler light need to be described with quantum. All light fields participating in SFG process need to be described with the quantum viewpoint [5].

In 2010, H. Hubel et al. directly prepared a three-photon entanglement state by the cascading method. 405-nm pump light pumped quasi-phase-matched PPKTP photonic crystals, generating downward conversion photon pairs via collinear SPDC: 405 nm → 775 nm + 848 nm. 775-nm photon beam pumped quasi-phase-matched PPLN waveguides again: 775 nm → 1510 nm + 1590 nm. The resultant three-photon entanglement state was detected by three single-photon detectors.

68.4 Application of Entangled Photons in Quantum Communication

68.4.1 Quantum Repeater and Quantum Memory

In order to eliminate the impact of loss of photon number and noise of detectors during telecommunication, a quantum repeater scheme based on entanglement

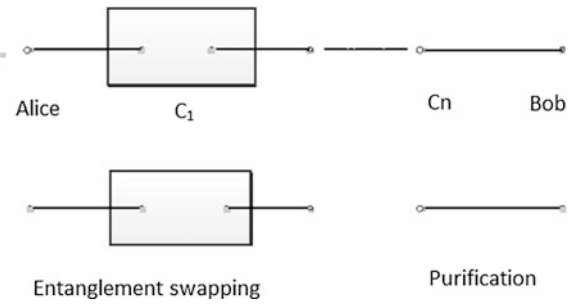


Fig. 68.2 Quantum repeater diagram

swapping and entanglement purification has thus come into being. This scheme can also get rid of the impact of decoherence effect of quantum channels during telecommunication. First, we divide total channels into N segments, prepare entanglement pairs at each segment, and send them to both ends of the segment. Then, we can purify these entanglement pairs and increase the distance of purified entanglement pairs via entanglement swapping between adjacent segments. After entanglement swapping is completed, entanglement degree will be reduced. For this reason, we also need to repeat the purification procedure several times until almost perfect entanglement pairs have been established between two sites far apart (as shown in Fig. 68.2).

In order to realize quantum repeater, entanglement needs to be established between quantum bits far apart and should be kept for a sufficient long period; several quantum bits at adjacent segments should be controlled collectively. Quantum memory is very important because quantum purification is a probability process. When one segment is successfully purified, quantum memory first memorizes quantum state, and segments unsuccessfully purified will continue purification. This is crucial for polynomial growth of telecommunication time. The reason is as follows: without such a quantum memory, all segments should be successfully purified simultaneously, the probability of which is very low, and time consumed will exponentially increase. In 2001, Duan et al. proposed atomic ensemble for quantum memory, that is, the well-known DLCZ scheme, the currently most feasible method recognized internationally. At present, important achievements have also been made in experimental study of quantum repeater based on atomic ensemble.

68.4.2 Quantum Teleportation

On the strength of non-locality of quantum entanglement, we can implement teleportation of quantum states. In 1993, American physicist Bennett and others proposed a scheme of quantum teleportation for the first time: sending the unknown quantum state of one particle to another particle (elsewhere), with the original particle remaining where it is. The original information is divided into two parts.

Classical information and quantum information are transmitted to the receiver via classical channels and quantum channels, respectively. Some measurements of the original particle made by the sender are classical information; any information unextracted by the sender during original measurement is quantum information. According to the above two types of information, people can set up a complete duplicate state based on the original quantum state. Equivalently, the sender does not even know the quantum state to be sent, while the receiver prepares the quantum state of the original by the use of other particles, including ones different from the original particle. This transmission method is not only absolutely safe, but also "invisible."

Alice sent the photon state of photon 1 to Bob by the use of entangled photon pairs 2 and 3. Assuming the original state of photon 1 as follows:

$$|\psi\rangle_1 = a|H\rangle_1 + b|V\rangle_1 \tag{68.6}$$

wherein $|H\rangle_1$ and $|V\rangle_1$ refer to horizontal polarization and vertical polarization of photon 1, respectively; entangled photon pairs 2 and 3 are in entangled states, which are produced by preparation:

$$|\psi\rangle_{23} = \frac{1}{\sqrt{2}}\left(|H\rangle_2|V\rangle_3 - |V\rangle_2|H\rangle_3\right) \tag{68.7}$$

Total photon states of three photons are:

$$\begin{aligned}|\psi\rangle_{123} = |H\rangle_1 \otimes |V\rangle_{23} = \frac{1}{\sqrt{2}}(&a|H\rangle_1|H\rangle_2|V\rangle_3 - a|H\rangle_1|V\rangle_2|H\rangle_3 \\ &+ b|V\rangle_1|H\rangle_2|V\rangle_3 - b|V\rangle_1|V\rangle_2|H\rangle_3)\end{aligned} \tag{68.8}$$

wherein photons 2 and 3 are in maximally entangled Bell state, while photons 1 and 2 as well as photons 1 and 3 are not entangled. Bell state measurement of photons 1 and 2 at Alice requires us to expand $|\psi\rangle_{123}$ on the Bell basis formed by photons 1 and 2. The Eigen states corresponding to photons 1 and 2 are shown as follows:

$$\left|\phi^+\right\rangle_{12} = \frac{1}{\sqrt{2}}\left(|H\rangle_1|V\rangle_2 + |V\rangle_1|H\rangle_2\right) \tag{68.9}$$

$$|\phi^-\rangle_{12} = \frac{1}{\sqrt{2}}\left(|H\rangle_1|V\rangle_2 - |V\rangle_1|H\rangle_2\right) \tag{68.10}$$

$$|\varphi^+\rangle_{12} = \frac{1}{\sqrt{2}}\left(|H\rangle_1|H\rangle_2 + |V\rangle_1|V\rangle_2\right) \tag{68.11}$$

$$|\varphi^-\rangle_{12} = \frac{1}{\sqrt{2}}\left(|H\rangle_1|H\rangle_2 - |V\rangle_1|V\rangle_2\right) \tag{68.12}$$

Then, three-photon state $|\psi\rangle_{123}$ is expanded on the Bell basis, with the result below:

$$|\psi\rangle_{123}=\frac{1}{2}\left[\begin{array}{l}-|\phi^{-}\rangle_{12}\left(a|H\rangle_{3}+b|V\rangle_{3}\right)-\left|\phi^{+}\right\rangle_{12}\left(a|H\rangle_{3}-b|V\rangle_{3}\right)\\+|\varphi^{+}\rangle_{12}\left(a|V\rangle_{3}+b|H\rangle_{3}\right)+|\varphi^{+}\rangle_{12}\left(a|V\rangle_{3}-b|H\rangle_{3}\right)\end{array}\right] \tag{68.13}$$

As shown by the above formula (68.8), if photons 1 and 2 are measured at Alice, when photons 1 and 2 are in a state of $|\phi^{-}\rangle_{12}$, photon 3 sent to Bob which has the original state of photon 1. If photon states of photons 1 and 2 measured at Alice are $\left|\phi^{+}\right\rangle_{12}$, Alice sends the measurement result to Bob via classic channels, Bob turns vertical polarization $|V\rangle_{3}$ of photon 3 by 180 degrees, and then the original state of photon 1 : $|\phi^{-}\rangle_{12}$ can be obtained. If the state measured at Alice is $|\phi^{-}\rangle_{12}$, then Bob makes a unitary transformation of photon 3 and switches vertical polarization $|V\rangle_{3}$ and horizontal polarization $|H\rangle_{3}$ to obtain the original state of photon 1. If the state measured by Alice is $|\varphi^{+}\rangle_{12}$, then Bob makes the above two transformations of photon 3 at the same time to obtain the original state of photon 1. In this way, by measurement, whatever measurement result is obtained by Alice, Bob can utilize the classic information sent from Alice to carry out a unitary transformation of photon 3, providing photon 3 with a state of photon 1, which is the process of teleportation.

The conception of Quantum Internet similar to Internet has been proposed through the process of photon teleportation. Quantum Internet contacts multiple information processors via quantum channels so as to complete the processing and transmission of quantum information. For the purpose of further promoting work and actualizing networked transmission of quantum information, hyper-entangled states, i.e., simultaneously entangled quantum states with multiple independent degrees of freedom in the atomic system, have also been proposed in recent years [6]. At present, hyper-entangled states have been generated experimentally. Sergienko and Jaeger [7] experimentally generated hyper-entangled two-photon states entangled in energy time, polarization, and momentum freedom. The team composed of Pan and Chen [8] from University of Science and Technology of China reported hyper-entangled photon Schrodinger-cat states generating ten quantum bits; Yan et al. [9] reported a scheme of generating non-degenerate hyper-entangled two-photon states; Hu and Zhan [10] presented a scheme of generating 4-quantum-bit hyper-entangled states in the non-interacting atomic system by the use of linear optical technology and sideband transition. These schemes and experimental studies are significant for further research into the properties of hyper-entangled states as well as in quantum communication and quantum computing.

68.5 Conclusion

Entangled states are the most unique states in quantum mechanics. The non-locality of quantum mechanics is centrally embodied in entangled states. The basic starting point of classical mechanics is local realism, free of super-beam transmission of physical quantity. However, quantum mechanics and quantum information deal with quantum state and quantum information, without classical correspondence; quantum-state measurement is random and irreversible; spatial non-locality cannot be interpreted by classical mechanics. Entanglement is unique to quantum mechanics; different from classical physics, it is an unprecedented resource. We can apply quantum entanglement to many aspects of quantum communication. Theoretical and experimental researches into quantum entanglement not only enjoy a broad development prospect, but also can deepen people's understanding of quantum science.

Acknowledgment This work was financially supported by the project under the National Spark Program (2011GA690403) and the project of research on the higher educational reform of Huaiyin Normal University (11GJA0020): The Practice and Innovation Projects for university students of Jiangsu (201310323058X).

References

1. Shixun Z (2009) Quantum mechanics. Higher Education Press, Beijing, pp 141–148 (in Chinese)
2. Zhengwei Z, Guangcan G (2000) Quantum entanglement. Physics 29(11):695–699 (in Chinese)
3. Shan L (2007) Photon entanglement and quantum communication. University of Science and Technology of China, Hefei (in Chinese)
4. Einstein A, Podolsky B, Rosen N (1935) Can quantum-mechanical description of Physical reality be considered complete? Phy Rev 47(10):777–780
5. Zhuo D (2012) Preparation of photon entanglement and its application in quantum communication. Beijing University of Posts and Telecommunications, Beijing (in Chinese)
6. Clifton R, Feldman DV, Michael L, Redhead G, Wilce A (1998) Superentangled states. Phys Rev A 58:135
7. Sergienko A, Jaeger Gregg (2003) Quantum information processing and precise optical measurement with entangled-photon pairs. Contemp Phys 44(4):34l–356
8. Gao W-B, Lu C-Y, Yao X-C, Xu P, Gühne O, Goebel A, Chen Y-A, Peng C-Z, Chen Z-B, Pan J-W (2010) Experimental demonstration of a hyper-entangled ten-qubit Schrödinger cat state. Nature Physics 6:331–335
9. Yan H, Zhang S, Chen JF, Loy MMT, Wong GKL, Du S (2011) Generation of Narrow-Band Hyperentangled Nondegenerate Paired Photons. Phys Rev Lett 106:33601
10. Hu BL, Zhan Y-B (2010) Generation of hyperentangled states between remote noninteracting atomic ions. Phys Rev A 82:54301

Chapter 69
Thickness Evaluation and Rebar Recognition of Railway Tunnel Lining Based on GPR

Jin Ma and Weixiang Xu

Abstract In this paper, using ground penetrating radar (GPR), we propose a novel method to evaluate the thickness of second lining and the number of rebar for tunnel detection. Based on the propagation characteristics of electromagnetic wave, we propose a method to evaluate the thickness of second lining by using the Fresnel reflection coefficient and attenuation coefficient. We propose a symmetric algorithm to recognize and determine the number of rebar which is recognized according to the characteristics of hyperbolic echo signal from the rebar. In addition, under the condition of symmetry-based algorithm, the Hough transform algorithm is applied to fit hyperbola of rebar. Then, the aforementioned methods are applied to the actual GPR images and obtain better results.

Keywords GPR · Fresnel coefficient · Attenuation coefficient · Hyperbola · Symmetric-based algorithm · Hough transform

69.1 Introduction

At present, among the tunnel detection technologies, GPR with the advantages of nondestructive, high resolution, efficient, the smaller impact on tunnel operation is widely applied in tunnel detection. But most cases of GPR image interpretation depend on the experienced experts, which are not only time-consuming, but also limits capacity. This article mainly solves two problems of tunnel detection, including the thickness evaluation of the second lining, the position identification, and quantity determination of rebar. For the thickness evaluation of second lining, we use a method based on the Fresnel reflection coefficient and the attenuation

J. Ma (✉) · W. Xu (✉)
School of Traffic and Transportation, Beijing Jiaotong University,
No. 3 Shang Yuan Cun, Hai Dian District, Beijing, China
e-mail: 14120864@bjtu.edu.cn

W. Xu
e-mail: wxxu@bjtu.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_69

coefficient to automatically identify the thickness of second lining. For the position and the number of rebar, reflection waves will form the hyperbola, and the vertex of hyperbola corresponds to the position of the rebar. Based on the characteristics of hyperbolic symmetry, we propose a symmetry-based algorithm and Hough transform algorithm to automatically recognize the position and the number of hyperbola. Following, we will do a brief introduction to the above algorithm.

69.2 Thickness Identification of Second Lining

69.2.1 Fresnel Coefficient and Attenuation Coefficient

The electromagnetic wave will reflect when it transmitted in different interfaces of medium, according to Fresnel coefficient [1]:

$$k = \alpha + j\beta \tag{69.1}$$

In Eq. (69.1), the real part α is the attenuation constant, the imaginary part β is the phase constant, ω is the angular frequency of the electromagnetic wave, ε is the permittivity of the medium, σ is the conductivity of the medium, and μ is the conductivity of medium.

Where,

$$\begin{cases} \alpha = \omega \sqrt{\frac{\mu\varepsilon}{2}\left[\sqrt{1+\left(\frac{\sigma}{\omega\varepsilon}\right)^2}-1\right]} \\ \beta = \omega \sqrt{\frac{\mu\varepsilon}{2}\left[\sqrt{1+\left(\frac{\sigma}{\omega\varepsilon}\right)^2}+1\right]} \end{cases} \tag{69.2}$$

The same medium for different frequencies of the electromagnetic wave will show a different conductivity and dielectric properties. There are two kinds of medium, i.e., low loss medium and good conductor medium. The main method is the low loss medium in our paper. For the low loss medium, we approximately consider that:

$$\sqrt{1+\left(\frac{\sigma}{\omega\varepsilon}\right)^2} \approx 1 + \frac{1}{2}\left(\frac{\sigma}{\omega\varepsilon}\right)^2 \tag{69.3}$$

Therefore,

$$\alpha \approx \frac{1}{2}\sigma\sqrt{\frac{\mu}{\varepsilon}} \tag{69.4}$$

$$\beta \approx \omega\sqrt{\mu\varepsilon} \tag{69.5}$$

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \tag{69.6}$$

where η is the wave impedance. In the concrete, relative permeability $\mu \approx 1$, so $\eta = 1/\sqrt{\varepsilon}, \eta$ is proportional to the permittivity ε of the medium. There is a difference about the wave impedance between the adjacent two layers. We expressed it by the reflection coefficient R [8], i.e.,

$$R_{12} = \frac{\cos\theta_1 - \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_2}}{\cos\theta_1 + \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_2}} \tag{69.7}$$

where θ_1 and θ_2 are the incident angle and reflection angle, respectively; ε_1 and ε_2 are the permittivity of two kinds of media interface. In most cases, the transmitting antenna and the receiving antenna are very close, which almost vertical incidence and reflection, at this time $\theta_1 = \theta_2 \approx 0$, substituted into Eq. (69.7).

$$R_{12} \approx \frac{1 - \sqrt{\frac{\varepsilon_2}{\varepsilon_1}}}{1 + \sqrt{\frac{\varepsilon_2}{\varepsilon_1}}} = \frac{\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \tag{69.8}$$

Electromagnetic wave propagates in the medium, and the energy attenuates constantly. The higher frequency of the antenna is the lower depth of its detection. The attenuation coefficient has a close relationship with the medium, when the high-frequency electromagnetic wave propagates in the air layer, which almost without energy attenuation, and amplitude keeps stability when the electromagnetic wave continues to spread; the amplitude will decay slowly, and in a moment the echo signal will change significantly. At this moment, the electromagnetic wave propagates from the air layer into the second lining. In terms of Eq. (69.8), the reflection coefficient is negative ($\varepsilon_{\text{air}} < \varepsilon_{\text{second}}$), because the dielectric constant varies greatly between the interface of air layer and the second lining layer, and the amplitude fluctuation is strong. However, the electromagnetic wave propagates from the second lining to the first lining, due to the differences in dielectric properties, which similarly lead to the change of reflection coefficient.

69.2.2 Time-Domain Diagram of GPR

When the GPR collects data go along the tunnels, a period of time, that is, the time of high-frequency radar pulse signal from transmit to receive, GPR scans this point; we called it as one-dimensional time-domain diagram (A-Scan). When the GPR goes along the exploration line to move a certain interval, all of A-Scan diagrams are mapped into a two-dimensional time-domain diagram (B-Scan). In Fig. 69.1, on

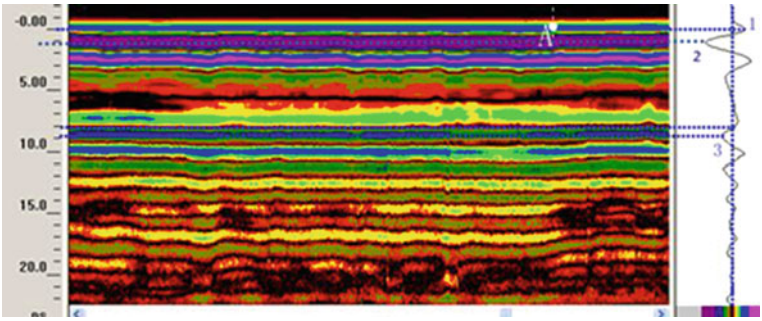


Fig. 69.1 A-scan and B-scan of GPR

the right side is an A-Scan from the left B-Scan of A. Based on the theory of 69.2.1, we can judge the interface position according to the changes in the inflection point of amplitude. At the positions of 1, 2, and 3, the amplitude changes obviously, and we can judge that they are the positions of different interfaces.

Based on the characteristics of the Fresnel coefficient and the attenuation coefficient, we can determine the thickness of second lining by inversion calculation. The inversion of thickness in tunnel lining is based on the propagation time Δt_i of the electromagnetic wave in each lining layers and the propagation velocity v_i of the electromagnetic wave at each layer.

$$\Delta t_i = t_{i+1} - t_i \tag{69.9}$$

$$v_i = \frac{c}{\sqrt{\varepsilon_i}} \tag{69.10}$$

Among them, t_i is the two-way travel time of the electromagnetic wave arriving the interface, c is the propagation velocity of electromagnetic wave in the air, and ε_i is the permittivity of the layer. So, the thickness of the second lining D can be written as follows:

$$D = \frac{c}{2\sqrt{\varepsilon}}(t_2 - t_1) \tag{69.11}$$

In Eq. (69.11), t_2 is the two-way travel time of electromagnetic wave reaching the interface of the second lining and the first lining, and t_1 is the two-way travel time of electromagnetic wave reaching the interface of air layer and the second lining layer.

69.3 Algorithm Introduction

69.3.1 Symmetry-Based Algorithm

The regular distribution of the rebar in the second lining leads to a certain regular distribution of the GPR images. In Fig. 69.2, when the GPR detector collects the data of tunnel, firstly, the pulse signal is transmitted from the horizontal position x, and when the electromagnetic wave encounters the rebar, the reflection occurs and the echo signal is received by the receiving antenna, while at the same time the receiving signal equipment feed backs the two-way travel time t. When the antenna reaches the top of target object, the two-way travel time of electromagnetic waves is the shortest, which is marked with t_0, and when the GPR antenna continues to move forward in another place x to detect the target object, the two-way travel time is equal to the time of frontal place x. In Fig. 69.2, the hyperbolic image of the rebar in the GPR is about the vertex symmetry, and the strong reflection of the vertex is monotonic decreasing.

The calculation formula of the symmetry about hyperbola in the literature [2] is given as:

$$S(x) = \sum_{x=1}^{x\text{size}} \sum_{w=1}^{k} [abs(I(x, y-w) - I(x, y+w)] \tag{69.12}$$

where $I(x, y)$ is the pixels in the two-dimensional image of GPR, k is the maximum calculation column width of the symmetry, $y \in [k+1, y\,\text{size} - k]$, $x\,\text{size}$ is the number of columns width of the image, $y\,\text{size}$ is the number of columns height of the image. Through the calculation of the symmetry, we can determine the local minimum point. However, the minimal value point of hyperbola also emerged in the intersection of hyperbolica. In order to determine the position of rebar, we need to carry out further processing. The vertex and intersection of the hyperbolic curve occurs alternately. When the first minimum point is determined, the following points will be determined. On the other hand, the first point is identified by the GPR

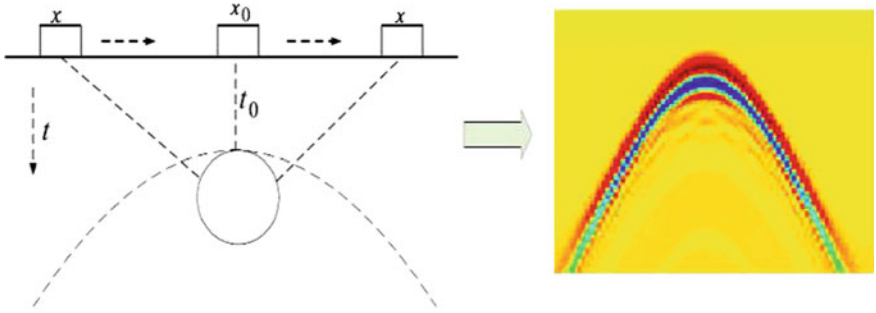


Fig. 69.2 Schematic of the hyperbolic formation

profile [3]. If N_0 is the number of minimum point, then the number of rebar is $(N_0 + 1)/2$.

69.3.2 Hough Transform

The hyperbolic model of the GPR image can be expressed by the following formula:

$$\frac{(t + 2r/v)^2}{(t_0 + 2r/v)^2} - \frac{(x - x_0)^2}{(v \cdot t_0/2 + r)^2} = 1 \tag{69.13}$$

where x is the propagation distance of electromagnetic wave, t is the two-way travel time, x_0 and t_0 are the distance and two-way travel time of the antenna at the top of bar, respectively, v is the propagation velocity of electromagnetic wave, and r is the radius of bar. We use m and n to represent the real semi-axis and imaginary semi-axis of the hyperbola, respectively:

$$m = t_0 + 2r/v \tag{69.14}$$

$$n = v \cdot t/2 + r \tag{69.15}$$

From (69.14) and (69.15), the following can be obtained:

$$r = n(m - t_0)/m \tag{69.16}$$

From Eq. (69.13), we must obtain a parameter group Q if we want to fit the imaging form of hyperbolic of rebar, where $Q = (r, v, t_0)$. Based on the frontal symmetric algorithm, we have determined the vertex positions of the hyperbolic echo signal, namely x_0 and t_0, are given. On the other hand, the relative dielectric constant of the medium is also given, and here, we assume homogeneous isotropic of the lining corresponding to the electromagnetic wave propagation velocity v is known. Therefore, if m and n are determined, then we can fit the hyperbola of steel rebar. In this paper, we use the voting algorithm of the Hough transform to vote on the m and n; Hough transform [4] is a feature extraction techniques used for image analysis, and the formula of the Hough voting is as follows [7]:

$$\hat{T} = \arg \max_{T} H(m_T, n_T) \tag{69.17}$$

$$H(m_T, n_T) = \sum_{i=1}^{N} \sum_{j=1}^{N} F(m_i, n_j) \tag{69.18}$$

where $m_i \in (t_0, m_{\max})$, $n_j \in (t_0 \cdot v/2, n_{\max})$, $(m_{\max} - t_0)/N$ is the step of Computation, $F(m_i, n_j)$ is the spatial voting of m_i and n_j under the condition of (69.14) and (69.15), $H(m_T, n_T)$ is the matrix of accumulation after the vote, $\hat{T}$ is the sum matrix of maximum value, m_T and n_T can be obtained by the maximum of the cumulative matrix of the vote. By the previous vertex and the value of real semi-axis and the imaginary semi-axis, we can fit the hyperbolic curve of rebar; simultaneously, the radius of the rebar can be obtained:

$$r = n_T(m_T - t_0)/m_T \tag{69.19}$$

69.4 Experimental Results

In Fig. 69.3, we use the GPR data, which are formed by the GprMax electromagnetic simulator, and GprMax is based on the finite time-domain difference method (FDTD) to form the GPR data which is set by users [5, 6]. We choose the antenna with 1.5 GHz, scanning step is 0.02 m, which forms the forward modeling of rebar. The number of rebar is set to 7 in the process of forward modeling. By using symmetric algorithm, we identified that the number of minimum point is 13 (The vertices are marked with red "+," and the crossing points are red "○," the same as follows). In addition, we fit the hyperbolic form by Hough transform, and the results are consistent with the default value. So, the algorithm is effective and feasible. Simultaneously, we apply the algorithm to the actual GPR image. During the experiment, we selected the part date of GPR, which are collected from Yan Zipo and Zhuangping tunnel.

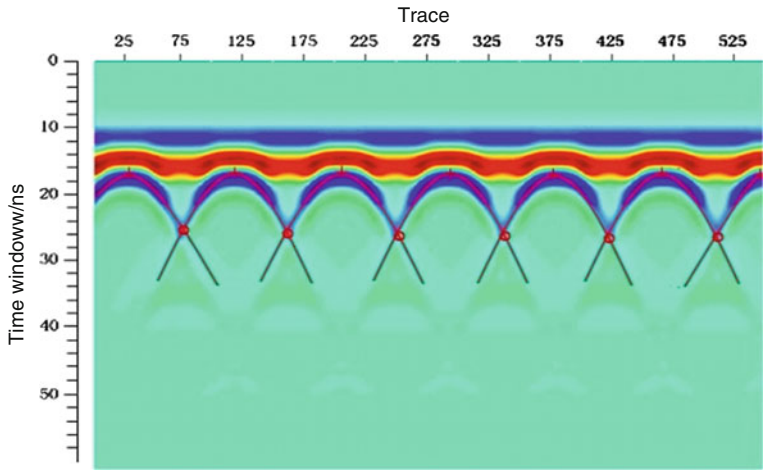


Fig. 69.3 The forward modeling of rebar

69.4.1 Lining Thickness Assessment

Figure 69.4 shows the part of GPR image of the vault (B-Scan), which is the Yan Zipo tunnel that is located in Neikun line from the pile DK179 + 260 to DK179 + 275, and the compound lining was applied to this tunnel. Data acquisition instrument that we select the GSSI SIR-20 of American, with the air-coupled antenna and the center frequency, is 400 MHz, and the relevant parameters are shown in Table 69.1. Based on the application of Fresnel coefficient and attenuation coefficient, Fig. 69.4 shows the interface of lining directly. According to Eq. (69.11), we can calculate the corresponding thickness of the lining, i.e., $\varepsilon = 8$ of the concrete, and Fig. 69.4 also shows the comparison diagram of corresponding design thickness (Red dotted line) and the actual thickness (Black curve) of the tunnel; the design thickness of the Yan Zipo tunnel is 50 cm, and the qualified rate from DK179 + 260 to DK179 + 275 was 100 % (The thickness is greater or equal to the design thickness, which is considered qualified). The qualified rate of the total tunnel after assessment was 94.6 %, and the construction of the tunnel meets the basic quality requirements.

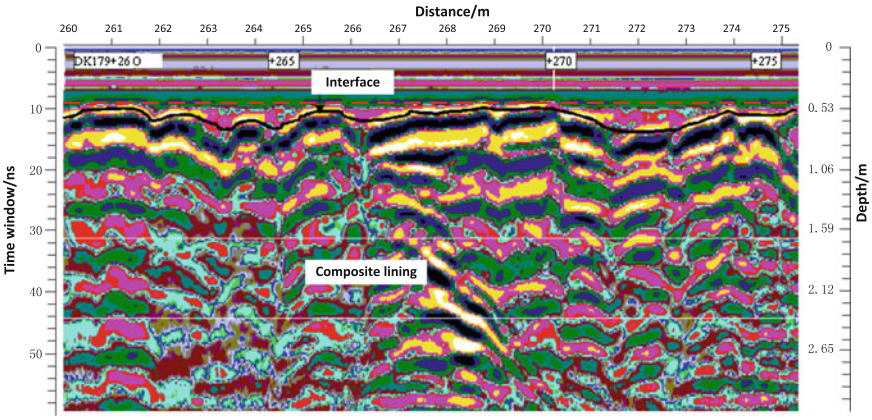


Fig. 69.4 The comparison diagram of the corresponding design thickness and the actual thickness in DK179 + 275 ~ DK179 + 260

Table 69.1 Parameters of ground penetrating radar and Yan Zipo tunnel

Center frequency (MHz)	Total time window (ns)	Sampling frequency	Scan step (m)	Relative permittivity of second lining ε	Design thickness (cm)
400	70	512/trace	0.02	8	50

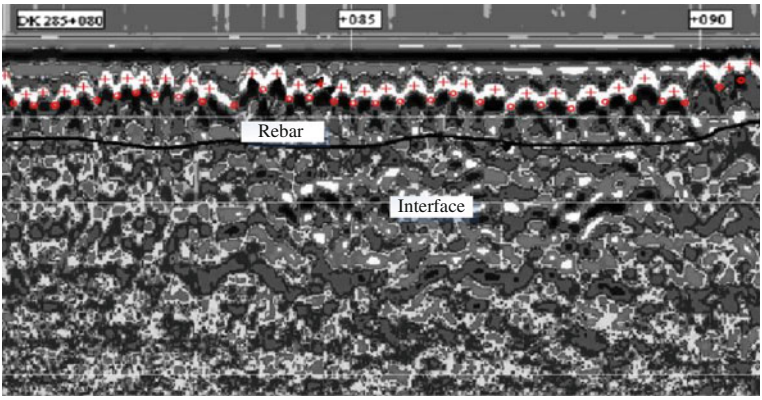


Fig. 69.5 The vertex and intersection of the hyperbola detected by the symmetry algorithm

69.4.2 Position Identification and Quantity Evaluation of Rebar

Figure 69.5 shows a part of gray image of GPR, which is the Zhuangping tunnel that is located in Bei Tong Pufu line from the pile DK285 + 080 to DK285 + 090. We can recognize the position of vertex and intersection of hyperbola by the symmetry algorithm.

In the pile of the DK285 + 080 $\sim$ DK285 + 090, a total number of minimum value is $N_0 = 61$ (31 vertex positions, 30 intersection points). So, the number of rebar is $(N_0 + 1)/2 = 31$; in order to further clarify the position of the rebar, Fig. 69.6 shows the form of a hyperbola fitting after Hough transform, from that we can more clearly distinguish the distribution of rebar. Table 69.2 shows the data of

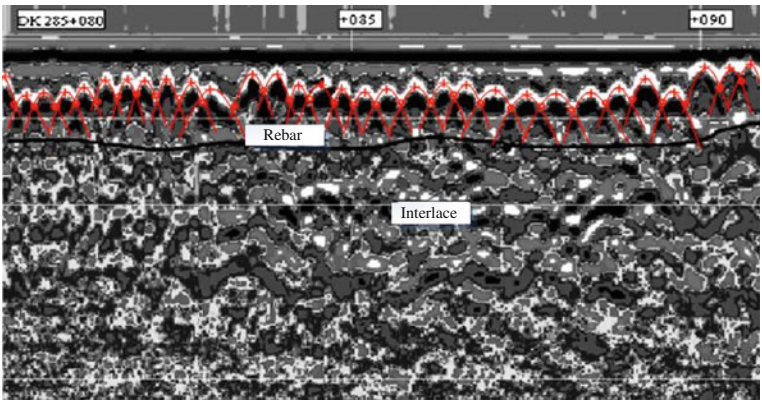


Fig. 69.6 The output images which are fitted by Hough transform

Table 69.2 The comparison between automatic detection date and design data of rebar

Mileage	Length	Design value		Actual value	
		Spacing (cm)	Number	Average spacing (cm)	Number
DK285 + 080 ~ +090	10	34	30	32.3	31

detection and design data in the part of DK285 + 080 ~ DK285 + 090 of the Zhuangping tunnel, which we can see that the rebar distribution of the tunnel meets the quality requirements.

69.5 Conclusion

This article mainly aims at evaluating the thickness tunnel lining, locating the rebar and identifying the number of rebar. Based on the characteristics of the Fresnel coefficient in the interface and the propagation characteristics of attenuation coefficient in the medium, the thickness of second lining can be evaluated. Then, we obtain the synthetic image of GPR lining interface, and with the relevant parameters, the corresponding thickness of lining is obtained. Simultaneously, a visual display of the lining thickness is present. To determine the number and location of the rebar, due to echo signal of rebar product the hyperbola, a symmetric-based algorithm is proposed to automatically identify and determine the position and the number of rebar. Then, the Hough transform algorithm is proposed and applied to fitting hyperbolic equation, by applying Hough voting to determine the values of the real semi-axis and the imaginary semi-axis of hyperbola, which aimed at fitting the hyperbola that is produced by rebar. At the same time, we applied the above methods to the actual GPR image to obtain better experiment results. So, the algorithm mentioned above is feasible for evaluating the thickness of tunnel lining and identifying the position of rebar as well as for determining the number of rebar.

Acknowledgment This work was supported in part by the China Academy of Railway Sciences Support Program under Grant No. 2008G017-A and the National Natural Science Foundation of China under Grant No. 61272029.

References

1. Li C, Li M, Zhao Y, Liu H, Wan Z, Jiangchun X, Xiaoping X, Chen Y, Wang B (2011) Layer recognition and thickness evaluation of tunnel lining based on ground penetrating radar measurements. J Appl Geophys 73:45–48. doi:10.1016/j.jappgeo.2010.11.004
2. Xiang L, Zhou H, Shu Z, Tan S, Liang G, Zhu J (2013) GPR evaluation of the Damaoshan highway tunnel: a case study. NDT&E Int 59:68–76. doi:10.1016/j.ndteint.2013.05.004

3. Maas C, Schmalzl J (2013) Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar. Comput Geosci 58:116–125. doi:10.1016/j.cageo.2013.04.012
4. Illingworth J, Kittler J (1988) A survey of the hough transform. Comput Vis Gr Image Process 44(1):87–116
5. Xie X, Qin H, Chao Y, Liu L (2013) An automatic recognition algorithm for GPR images of RC structure voids. J Appl Geophys 99:125–134
6. Chen H, Cohn AG (2010) Probabilistic robust hyperbola mixture model for interpreting ground penetrating radar data. Neural Netw (IJCNN), IEEE 1–8. doi:10.1109/IJCNN.2010.5596298
7. Zhao Q, Li W, Zhou H (2012) Based on the modified Hough model of hyperbolic fitting and target size inversion. J PLA Univ Sci Technol (Nat Sci Ed) 13(4) (in Chinese)
8. Yang F, Peng S (2010) Study on the principle and method of geological radar detection. Science Press, Beijing, pp 10–22 (in Chinese)

Chapter 70
Outlier Detection for Time-Evolving Complex Networks

Hong Zhang, Changzhen Hu and Xiaojun Wang

Abstract Complex systems have features such as numerous nodes and edges, complicated and hierarchical relations and evolving with time. During their running time, complex systems are influenced by the internal and external factors which can lead to abnormal states. Finding out the outliers can effectively supervise the whole system. Here, we study a real-world complex dynamic complex system, observe the abnormal pattern based on entropy, and find out nodes which will lead to the system collapse by GROD algorithm.

Keywords Data mining · Complex system · Outlier detection · Dynamic graph

70.1 Introduction

Being a difficult problem to solve, outlier detection has attracted wide interest, including complex systems. Complex systems, such as society network [1], biological system [2], information communication network [3], and power grid, occasionally occur some abnormalities, for example, the sudden drop in performance. The breakdown of power grid can cost millions of thousands of economic loss, and the food chain fracture may lead to ecosystem breakdown which will be difficult to recover in several years [4]. All the phenomena show the brittleness, which is a feature of complex systems and can be induced by the system architecture in configuration phase and the abnormalities in the running phase. The notion of the brittleness was introduced in [5], but in this paper, we will focus on the abnormalities (it may be called outliers in some other papers) which may lead to the system collapse.

It is difficult to find the outliers because of its rare occurrence. You can image that in the communication network, it is not an easy way to spot outliers in numerous of normal data; also comparing with the normal financial transactions, the fraud behaviors are low.

H. Zhang (✉) · C. Hu · X. Wang
Beijing Institute of Technology, Beijing, China
e-mail: GraceZXKL@126.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_70

Our contributions mainly lie in the following aspects:

- use the dynamic network graphs to find out the outliers,
- choose the entropy as the feature of the whole system, and
- detect outliers in a heuristic way.

70.2 Related Work

Some important features, such as degree power laws, "small-world phenomenon," or "six degrees of separation," have been studied explicitly in recent years. Outlier detection in complex systems also gained much attention lately. One of the most important aspects of the outlier detection is the change point [6], which features the abrupt change during the operation. The change point is also called shatter point, or the brittleness point.

Many real networks, such as cooperation networks, power grid in infrastructure, biosystem, and social network, have already been studied a lot. Gao et al. [3] found the nodes which do not well belong to the community they reside, and they treated the abnormal nodes as community outliers. Shetty detected the abnormal nodes in the Enron email network, where the entropy of the system increased when the abnormal nodes being removed; [7] gave the abnormal event detection on the evolving graph with the nodes having F features. It selected the top ten abnormal values by comparing the current data with the last w window and SVD (singular value decomposition), using only one feature during one specific abnormal detection.

70.3 Method

First, we give the framework of our work in Fig. 70.1. On the top of the figure are the sequential graphs G_t which change with time, some graphs during the system evolving are not displayed on it. We can find out the graphs which change significantly during the operation of the system based on the entropy. The abnormal graphs are separated from the former normal ones by the vertical dotted line. We put the abnormal graphs in a set. After getting the abnormal graphs set, we will detect the outliers in each abnormal graph by adopting GROD algorithm.

70.3.1 System Input

The system has a series of consecutive states during the running time. We call that a graph is a temporal graph if the graph changes with time and it can be defined as

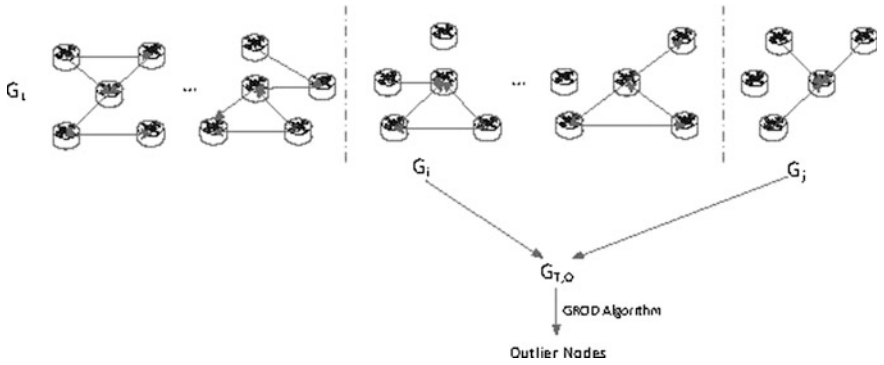


Fig. 70.1 Framework of outlier detection for time-evolving complex networks

$G_t = \{V^t, E_t, W_t\}$, $(t = 1, 2, \ldots, n)$ where V_t, E_t, W_t $(t = 1, 2, \ldots, n)$ are nodes set, edges set, and weights set, respectively, at time t.

70.3.2 Feature Extraction

A lot of research work has been done on the study of the static graphs; however, little attention is paid on the dynamic graphs. As for the outlier detection in dynamic graphs, the most important task is to extract the main overall feature of the whole system.

We all know some features of the networks. In random networks, most of the nodes have a uniform degree, so we call the homogeneous networks are ordered networks. On the other side, many real-world networks have a large number of nodes and complicated scale-free topology, we call these networks are heterogeneous networks, i.e., few nodes have large degrees and all the others have small degrees. Also we call these networks with "long tail" degree distribution are disordered ones. Entropy is understood as a measure of disorder even though the definition of which are not precisely defined. We adapt the concept of entropy in [8], in this paper which equals to Korner's definition on complete graphs, and give only one of the definitions in Eq. 70.1.

$$H(G, P) = \min_{x \in \text{StableSet}(G)} \sum_{v \in V(G)} p_v \log(p_v) \tag{70.1}$$

StableSet (G) in the above equation is called a stable set which is based on the vertices of G, and stable means the set of the vertices do not have any edges. Even though it has many applications in different fields, finding the stable set is a NP-hard problem. There may be some approximate methods to get the stable set,

and it is not so easy in practice. Considering the above factors, we decided to choose the Eq. 70.2 as below.

$$H(G,P) = \sum_{i=1}^{|V|} p(v_i) \log\left(1/p(v_i)\right) \tag{70.2}$$

where $G = (V, E)$ represents graph G consisting of vertex set V and edge set E, and P is the probability distribution of vertex set.

70.3.3 Finding Abnormal Graphs

After getting all the time series graphs of the complex system, we can compute each of the entropy which can be used as a measure of the disorder of the whole system. All the graph entropy are put into a sequence which is used to analyze the change point of the graph in terms of the significant change of the entropy.

70.3.4 Finding Abnormal Nodes

The brittleness of the complex system is the characteristic of the collapse induced by some of the abnormality of the system, which can be measured by the degree centrality. Degree centrality can be viewed as the importance of the node in complex networks. As for the undirected graphs, the degree centrality of the node can be measured easily just by the node degree, yet as for the directed graphs, we chose the in-degree or out-degree depends on the specific application. The pseudocode of algorithm GROD is shown in Table 70.1.

70.4 Real-World Network Analysis

To some sense, the topology of Internet can be viewed as groups of routers. The router and its lower-level routers form a sequence of autonomous system, and all the routers can exchange information with the neighboring nodes or the peer nodes through IGP or BGP. We can construct a communication network by looking through the log of the BGP. The network being analyzed in the paper is CAIDA AS, obtained from Stanford, which is a large autonomous system evolving from Jan. 5, 2004, to Nov. 12, 2007. In each of the graph, the relationship between nodes is recustomers, providers, peers, and siblings, represented by −1, 1, 0, and 2, respectively.

Table 70.1 GROD algorithm

GROD Algorithm: detecting outlier nodes in abnormal graphs
Input: $G_{T,O}$: the abnormal graph set
Output: TopOutlierNodesSet
For i=1: length($G_{T,O}$)
[j,k]=Sort(Degree(:,3),'decend'); [1]
For m=1: length(j);
G_{left}=Remove($G_{T,O}$, k);
$C_{T,k}$=CountLeftEntropy(G_{left})/TotalEntropy(G);
Sort($C_{T,k}$);
end m;
end i;

[1] Degree is weight matrix with 3 columns, the 1 st is NodeId, the 2 nd InDegree and 3 rd OutDegree;

70.4.1 System Input

CAIDA AS comprises 122 evolving graphs, represented by $G_t = \{V_t, E_t, W_t\}$, $(t = 1, 2, \ldots, 122)$ among which $V_t = \{V_t^i | i \in V_t\}$, $E_t = \{E_t^{i,j} | i, j \in V_t\}$ and $W_t = \{W_t^{i,j} | i, j \in V_t\}$ are, respectively, the node set, edge set, and weight set at time t. The relationship between nodes and edges of CAIDA AS is shown in Fig. 70.2. From Fig. 70.2, we can see the edges increased gradually with the increment of the nodes at the early period, but the trend of the increment does not last long. Nov. 13, 2006, is the first change point where the number of edges decreased sharply, and the

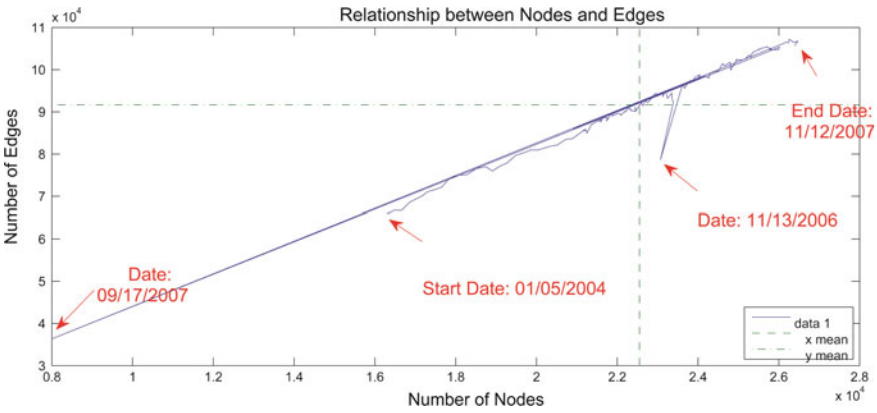


Fig. 70.2 Relationship between nodes and edge

second time is at Sept. 17, 2007, which is another change point. The change points during system running show that the system is not in an consistent state, so we should give the system a further analysis.

70.4.2 Finding Out Abnormal Graphs

After analyzing, we chose entropy as the measurement of the system. Figure 70.3 shows the entropy of the system during its running. We can see that with time elapsing, the entropy has an overall increment, but there also exist some time points, marked with number from 1 to 5, where the entropy exceeds others', and the corresponding date was Feb. 2, 2004, Apr. 5, 2004, Nov. 13, 2006, Sept. 10, 2007, and Sept. 17, 2007. The horizontal line in the figure is the mean value of the entropy of the system.

70.4.3 Finding Out Abnormal Nodes

We can give an analysis to the abnormal graphs found above in terms of the entropy of the system by using the GROD algorithm through gradually removing the nodes sorted by out-degree. Figures 70.4 and 70.5 show the effect on the largest cluster and edges by the nodes which cause a significant change of the system entropy, and the removed node id is 185, with the date Feb. 2, 2004. Other four abnormal node ids are 184, 188, 193, and 48, with the dates Apr. 5, 2004, Nov. 13, 2006, Sept. 10, 2007, and Sept. 17, 2007, whose figures are similar to the one in Feb. 2, 2004.

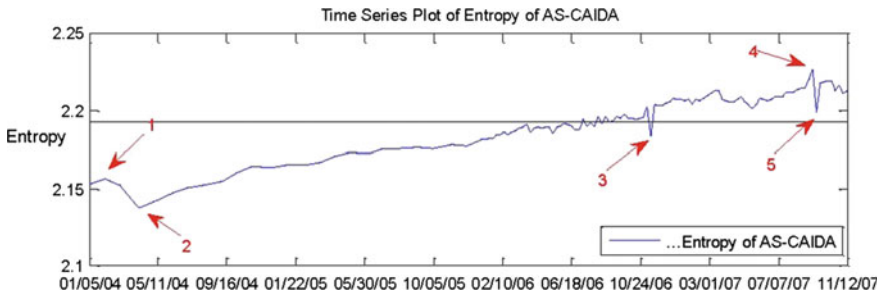


Fig. 70.3 Entropy of the system changes with time

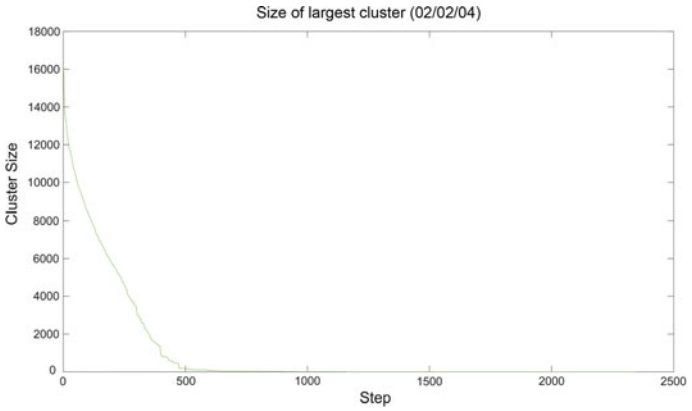


Fig. 70.4 Changes of the largest cluster after removing the abnormal nodes

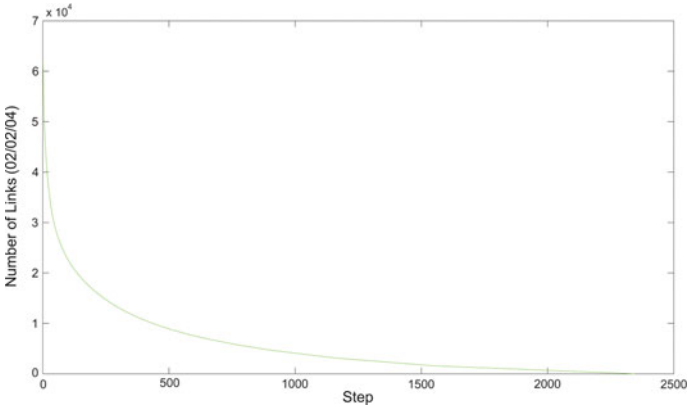


Fig. 70.5 Changes of the edges after removing the abnormal nodes

70.5 Conclusion and Future Work

- We should not rest on just finding out the abnormal nodes, that is to say, we should further discover the inner rules of the abnormal nodes, e.g., the time series of the abnormal nodes, and the regular pattern which can be used to monitor the abnormal behavior of the system.
- Abnormal behavior is a new way to find outliers although there are not many papers in this field. Building the normal pattern of the normal behavior through observing the behavior of the nodes of the system helps to find the abnormal behavior, which can predict the future behaviors [9].

Whether the abnormal nodes with different scores have correlations? This problem has a significant effect on the operation of the system in a long run [10], i.e., if the node with the highest score was removed because of disorder, whether the second one continues in the similar way as the first one?

References

1. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE et al (2010) Identifying influential spreaders in complex networks. http://arxiv.org/abs/1001.5285. Cited 18 Aug 2010
2. Sah P, Singh LO, Clauset A, Bansal S (2014) Exploring community structure in biological networks with random graphs. BMC Bioinf 15:220. doi:10.1186/1471-2105-15-220
3. Gao J, Liang F, Fan W, Wang C, Sun Y, Han J (2010) On community outliers and their efficient detection in information networks. KDD'10, pp 813–822
4. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. Nature 464(7291):1025–1028
5. Perry DE, Wolf AL (1992) Foundations for the study of software architecture. ACM SIGSOFT Softw Eng Notes 17(4):40–52
6. Peel L, Clauset A (2014) Detecting change points in the large-scale structure of evolving networks. arXiv:1403.0989[cs.SI]
7. Akoglu L (2012) Mining and modeling real-world networks: patterns, anomalies, and tools. Carnegie Mellon University, Pittsburgh, PA 15213
8. Korner J (1986) Bounds and information theory. SIAM J Algorithms Discrete Math 560–570
9. Tong H (2014) Query complex graph patterns: tools and applications. In: Proceedings of the companion publication of the 23rd international conference on world wide web companion. International World Wide Web Conferences Steering Committee, pp 1195–1196
10. Gao J, Liu Y-Y, D'Souza RM, Barabási A-L (2014) Target control of complex networks. Nat Commun 5

Chapter 71
Target Pose Measurement Based on Large Baseline Images

Wei Zhu and Yuhan Liu

Abstract In the test range, it is a practical significance to use photoelectric theodolite to obtain the 3D pose of flying object by tracking and recording flying target in image sequences. In this paper, the image is obtained by using the two long-distance theodolite; namely, the target pose measurement is attained by dealing with the large baseline target images. Because the large baseline target is small, weak and details are not prominent. Image is enhanced by using the method of frequency-domain image enhancement. The camera self-calibration is based on the method of fundamental matrix; the 3D pose is calculated by feature point DLT algorithm and beam adjustment method. Experimental results showed that the measuring error of the attitude angle was less 0.1°, when the axis of the plane imaging was more than 500 pixels. According to the test results of the precision comparison, it was concluded that the mathematical model is correct and the algorithm is reasonable in improving effectively the measurement precision of the target pose.

Keywords Target pose measurement · Binocular vision · Camera self-calibration · DLT algorithm · Fundamental matrix

71.1 Introduction

With the development of science and technology, range measurement equipment is more and more advanced, data obtained more and more abundant, including exterior ballistic parameter, interior ballistic parameters, target characteristic parameters. The target characteristic parameters include the attitude parameters. The

W. Zhu (✉) · Y. Liu
College of Communication and Information, Changchun University of Science and Technology, Changchun, China
e-mail: cczhaolirong@163.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_71

photoelectric theodolite is regularly used optical measuring equipment in the range. Due to the development of electronic devices, the use of high-speed camera, medium-wave and long-wave infrared camera, and the photoelectric theodolite has made new progress in the field of measurement [1].

Over the past several decades, the monocular vision and the binocular vision pose measurement have become a scientific research topic, which has been widely concerned by scientists and has been applied in many fields. For example, the electronic equipment, robot control, virtual reality control, etc., all need to measure the target's attitude by the monocular vision and the binocular vision method.

The method of the monocular vision measurement is divided into two categories [2]: (a) feature-based methods and (b) appearance-based methods. The training samples based on contour matching method are usually used to obtain the template to measure the target attitude. A 3D pose of the target is obtained by a target with the same part in two images based on features approach. There are two processing methods usually to deal with the data: (a) linear methods and (b) nonlinear methods. The nonlinear method is usually robust to noise, but it is very large to compute. It must have a good initial value; otherwise, it will not converge, but the linear method needs the computation quantity to be small, because the computation accuracy is not high and interference rejection is not strong, the output result is poor.

In the past few years, the method of multicamera vision pose measurement has been favored by many scientists [3, 4]. Methods include SFM, PnP, and BLUE. SFM (structure-from-motion) is a method of using two-dimensional planar target motion picture to restore the 3D structure of the target (3D shape and 3D motion state). PnP (perspective-n-point) is a classical problem in computer vision, photogrammetry, and geometric relations. In the condition of camera and internal parameter coefficients, the image of a space N feature points is captured, and the image points of each feature point are extracted by image analysis, solving equations to obtain the target pose. BLUE (best linear unbiased estimation) is the optimal linear unbiased estimation.

In this paper, the attitude measurement method is completed by dealing the target image of the photoelectric theodolite by the algorithm based on the image features of the large baseline. The experimental results show that the method is correct and reasonable.

71.2 Frequency-Domain Image Enhancement

The frequency-domain image enhancement has intuition characteristic, so it is usually used to transform the image from the spatial domain to the frequency domain by the Fourier transform, and then, the image is transformed into the spatial image by the inverse transformation of Fourier. That person and machine can

directly identify the image [5]. Basic steps of frequency-domain filtering are as follows: (a) Calculate DFT of the original image $f(x, y)$ to get the spectrum $F(u, v)$. (b) Shift zero frequency point of $F(u, v)$ to the center of the spectrum. (c) Calculation $G(u, v)$, $G(u, v)$ is the product of the filter function and the product. (d) The zero frequency point of the spectrum $G(u, v)$ is shifted to the upper left corner of the spectrum. (e) The Fourier inverse transform is used to calculate the results of (d) step, the result is $g(x, y)$, and the real part of $g(x, y)$ is taken as a result of the final filtering images.

It is known that the results of the frequency-domain image enhancement are determined by the frequency-domain filtering function $H(u, v)$ in the method of frequency-domain filtering. It filters out some of the frequency of the spectrum in the filter, while the other frequency is not affected. Because the image target features of this paper are small, we hope that the image can highlight the details of the target. Because of the stronger response of the Laplace operator to details, the Laplace filter is used in this paper.

The derivation of the Laplace operator in the frequency domain is as follows:

By Fourier transform nature:

$$FFT\left[\frac{\mathrm{d}^n f(x)}{\mathrm{d}x^n}\right] = (ju)^n F(u) \tag{71.1}$$

Laplace transform of the Fourier operator is calculated as follows:

$$FFT\left[\frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2}\right] = (ju)^2 F(u,v) + (jv)^2 F(u,v) = -(u^2+v^2)F(u,v) \tag{71.2}$$

So there are the following:

$$FFT\left[\nabla^2 f(x,y)\right] = -(u^2+v^2)F(u,v) \tag{71.3}$$

Namely, the Laplace filter in the frequency domain is:

$$H(u,v) = -(u^2+v^2) \tag{71.4}$$

According to the shift of the origin of frequency domain, formula (71.4) is changed into:

$$H(u,v) = -\left(\left(u-\frac{M}{2}\right)^2 + \left(v-\frac{N}{2}\right)^2\right) \tag{71.5}$$

71.3 Camera Self-Calibration Method Based on Fundamental Matrix

A camera self-calibration method based on fundamental matrix does not need to do the projection reconstruction of image sequences; instead, the equation is established between two images. This method has some advantages than step-by-step calibration method in the condition that it is difficult to unify all the images in the same projective frame. But the cost is unable to guarantee the consistency of the infinite plane in the projective space which is determined by all the images. The self-calibration method based on Kruppa equation may be unstable when the image sequence is longer, and its robustness depends on the given initial value [6]. The method of camera self-calibration based on fundamental matrix is simple and easy to implement. The specific method is as follows:

The formula for describing the relationship between the two image matching points is shown in (71.6). From the formula, we can know that the basic matrix can be determined only according to the matching points between the two images.

$$p'Fp = 0 \tag{71.6}$$

Method of solving the basic matrix by using linear method (Analytic Method) [7]. In the formula (71.6), $p = (x, y, 1)^T$, $p' = (x', y', 1)$ and the basic matrix is the:

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \tag{71.7}$$

Each pair of matched points (p, p') provides a linear equation for the unknown elements of the fundamental matrix F. Coefficient-specific forms are as follows:

$$(x', y', 1)\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

The 9-dimensional vector, which is composed of vectors, and arranged in a row. The formula can be expressed as the inner product of two vectors:

$$(x'x \quad x'y \quad x' \quad y'x \quad y'y \quad y' \quad x \quad y \quad 1)f = 0 \tag{71.8}$$

If there is a pair of matching points, there is a linear equation.Linear equations:

$$Af = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} f = 0 \tag{71.9}$$

When the number of matching points is greater than or equal to 8, the singular value decomposition (SVD) of the coefficient matrix can be obtained by the least square solution of the fundamental matrix.

71.4 Target Attitude Measurement Method

71.4.1 DLT Algorithm Based on Minimal Algebraic Error

Basic ideas of DLT algorithm are based on minimal algebraic error [8]: multiplication cross eliminate homogeneous vector proportion factor; three equations for each image point are given about the homogeneous coordinates of space points; two of them are linearly independent. Projection formula derived for solving the process.

$$x_{ij} \times (P_i X_j) = 0 \tag{71.10}$$

Formula (71.10) expanded:

$$\begin{aligned} x_{ij}(p_i^3 X_j) - (p_i^1 X_j) &= (x_{ij} p_i^3 - p_i^1) X_j = 0 \\ y_{ij}(p_i^3 X_j) - (p_i^2 X_j) &= (y_{ij} p_i^3 - p_i^2) X_j = 0 \\ x_{ij}(p_i^2 X_j) - y_{ij}(p_i^1 X_j) &= (x_{ij} p_i^2 - y_{ij} p_i^1) X_j = 0 \end{aligned}$$

In it, $p_i^l (l = 1, 2, 3)$ express the first l row of the projection matrix P_i, two equations above are arbitrarily choice, so $2k \times 4$ order linear equations can be composed of k points about X. Unit singular vector corresponding to matrix minimum singular value is as the optimal attitude solution.

71.4.2 Beam Adjustment Method

Beam adjustment method is usually based on the results of the solution of the video camera parameters and the position of the space point. The image point coordinates of the spatial points are calculated according to the imaging model. Space points are reprojected. The deviation between the results of the reprojection and the actual image points is as the optimization objective function. First, the adjustment calculation is listed collinear equation, that is the observation equation, also it is listed the constraint equations with the additional constraints and the initial value of the parameters of the given adjustment. The observation equation and constraint equation are linearization by Taylor expansion. The correction of each adjustment parameter is worked out by the stepwise iteration. When the correction amount is small to a given threshold, the imaging relationship meets the accuracy of the

calculation results. The target pose is obtained by solving the exact position of the target's space feature points.

The beam adjustment method [9] is as follows. The difference parameter is expressed by S. Observation equation:

$$\phi(S) = L \tag{71.11}$$

In which:

$$\phi(S) = \begin{bmatrix} F_x \dfrac{\bar{X}}{\bar{Z}} + C_x + \delta_x \\ F_y \dfrac{\bar{Y}}{\bar{Z}} + C_y + \delta_y \end{bmatrix} \quad L = \begin{bmatrix} x \\ y \end{bmatrix}$$

Conditional equation: $\varphi(S) = 0$. The collinear equation and condition adjustment equation in the initial value of the first-order Taylor are expanded:

$$\phi(S) = \phi(S_0) + B\delta_S = L + \delta_L, \varphi(S) = \varphi(S_0) + C\delta_S = 0$$

In which:

δ_L Correction of observation value;
δ_S Correction parameter adjustment;
B In the colinearity equation, the first-order partial derivative of the adjustment parameter S is carried out at the initial adjustment parameter S_0;
C In the condition equation, the first-order partial derivative of the adjustment parameter S is carried out at the initial adjustment parameter S_0.

That is:

$$B = \left.\frac{\partial \phi}{\partial S}\right|_{S_0} \quad C = \left.\frac{\partial \varphi}{\partial S}\right|_{S_0}$$

To minimize δ_L is regarded the objective function; the correction adjustment δ_S is obtained by using the least squares method in the initial value of the constraint conditions; the adjustment parameter S is corrected in the initial value basis, and the result S is stable until δ_S is less than a given threshold.

71.5 Test Result

In the field experiment, the focal length of the camera is 800 mm, the pixel size is 12 μm, and the resolution is 1280 × 1024. The distance between target and two stations was, respectively, 1310 and 923 m.

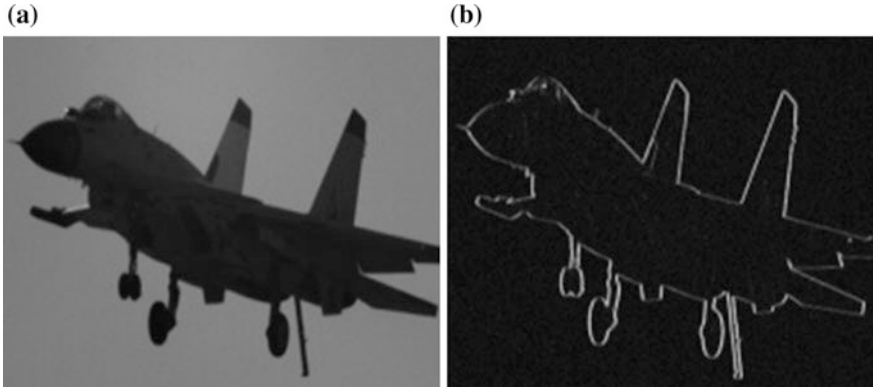


Fig. 71.1 Image of the *left* test station. **a** Original image. **b** Image after enhanced

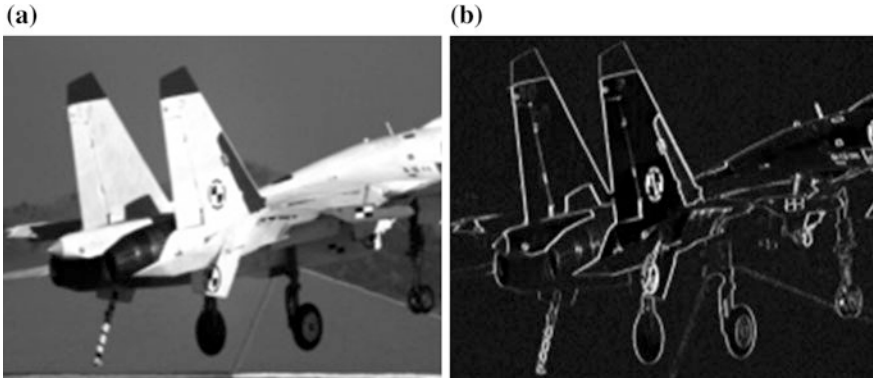


Fig. 71.2 Image of the *right* test station. **a** Original image. **b** Image after enhanced

Figure 71.1 is an image of the left test theodolite obtained; Fig. 71.2 is an image of the right test theodolite obtained. Target imaging axis direction is greater than 500 pixels.

Two station sequence image to be selected at the same time two frame image is calculated. The true values are provided by the aircraft GPS. The yaw angle, the pitch angle, and the roll angle are as follows: $\alpha' = 10.73°$, $\beta' = 5.24°$, $\gamma' = 2.92°$. Table 71.1 shows the attitude measurement results of DLT algorithm and beam adjustment method.

Table 71.1 The attitude measurement results of DLT algorithm and beam adjustment method

Attitude measurement method	The yaw angle(°)	The pitch angle(°)	The roll angle(°)	Mean error (°)
DLT algorithm	10.82	5.29	2.88	0.06
Beam adjustment method	10.79	5.32	3.01	0.08

71.6 Conclusion

Target image is small, weak, and not outstanding details based on large baseline theodolites. Image enhancement processing bases on the method of frequency-domain image enhancement. The camera self-calibration bases on the method of fundamental matrix. Attitude measurement method bases on feature points of DLT algorithm and beam adjustment method. Experimental results showed that the measuring error of the attitude angle was less 0.1°, when the axis of the plane imaging was more than 500 pixels. According to the test results of the precision comparison, it was concluded that the mathematical model is correct and the algorithm is reasonable in improving effectively the measurement precision of the target pose.

References

1. Zhu W, Zhao LR et al (2008) The extraction method for missile bullets' multiple targets of optical-electronic theodolite. Opt Precis Eng 16(11):2140–2144 (in Chinese)
2. Kahl F (2005) Multiple view geometry and the l-infinity norm. In: Proceedings of the international conference on computer vision
3. Agarwal S, Chandrakar M, Kahl F, Kriegman D, Belongie S (2006) Practical global optimization for multi-view geometry. In: European conference on computer vision, pp 592–605
4. Wang Y, Liu YH, Tao LM, Xu GY (2006) Real-time multi-view face detection and pose estimation in video stream. In: Proceedings of the 18th international conference on pattern recognition, vol 4, pp 354–357
5. Zhang Z, Wang YP (2010) Digital image processing and machine vision. People's Posts and Telecommunications Press, Beijing 4 (in Chinese)
6. Qu XJ, Zhang L (2011) Camera calibration and 3D reconstruction based on new imaging model. J Instrum Meter 32(8):1830–1836 (in Chinese)
7. Hartley RI (1992) Estimation of relative camera position for un-calibrated cameras. In: European conference on computer vision, pp 579–587
8. Qiu ML, Ma SD (2000) A survey of camera calibration in computer vision. Autom J 1 26 (1):43–55 (in Chinese)
9. Zhao LR, Zhu W et al (2012) Improvement of measurement precision of plane attitude by constructing optimization functions. Opt Precis Eng 20(6):1325–1333 (in Chinese)
10. Yu QF, Shang Y (2009) Principle and application of photogrammetry. Science Press, Beijing 3 (in Chinese)

Chapter 72
Reasonable Operation Distance on Mid–long Transport of High-Speed Railway

Yu Yang, Yun Fu, Xiaoxun Gao and Tao Niu

Abstract In China, the length of high-speed railway (HSR) line has sustained growth since 2007. There is more than 16,000 km HSR in operation and another 10,000 km under construction. According to the national strategy "Mid- and Long-Term Railway Network Plan," a national HSR network has appeared which was composed of "four north–south and four west–east" HSR corridors. The high-speed trains can reach more and more cities; mid–long-distance high-speed trains will be a new option for passengers. On the basis of a passenger transportation market survey, this article analyzes the passenger's transportation demands, identifies the market positioning of HSR for mid–long-distance service, and investigates the reasonable operation distance of high-speed trains which is suitable for Chinese situations. The results can serve as a reference for the operation plan of mid–long-distance high-speed trains; it also gives full play to HSR transportation advantages, meets the social demands, and improves transport efficiency and benefit.

Keywords High-speed railway (HSR) · Mid–long-distance transport · Passenger demand · Operation distance

72.1 Introduction

Chinese High-Speed Railway (HSR) has evolved to be the largest system in the world. By the end of 2014, the length of HSR in China had reached 16,000 km, accounting for more than 60 % of the total length in the world.

Y. Yang (✉) · Y. Fu · X. Gao
China Academy of Railway Sciences, No. 2 Daliushu Road, Haidian District, Beijing, China
e-mail: yuer162@163.com

T. Niu
China Railway Communication and Signal Survey and Design (Beijing) Co. Ltd., Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_72

By the end of this year, the length of HSR in China is expected to reach 18,000 km. According to the national strategy "Mid- and Long-Term Railway Network Plan," a nationwide HSR network has appeared which is composed of "four north–south and four west–east" HSR corridors. The network will cover the cities with more than 500 thousand residents [1]. Except for the eastern developed area of China, the HSR network has also extend to the less developed regions in central and western China.

HSR is playing a significant role in supporting and leading the overall economic and social development, while providing extreme convenience for the general public and even bringing the revolutionary change of the concept of traveling. The HSR gives a great push to the economic growth in promoting trade and economic exchange, accelerating cultural exchange and spread, coordinating regional economic development, speeding up urbanization process, and promoting transformation of economic structure.

With the rapid growth of HSR, how to make full use of the railway network and provide services matching the market demands that are major concerns of Chinese railway enterprises while approaching the market. The high-speed trains can reach more and more cities, and mid–long-distance high-speed trains (in this article, mid–long distance refers the distance above 1500 km) will be a new option for travelers. In recent years, China Railway has operated high-speed sleeper trains on the mid–long-distance high-speed lines, such as Beijing west to Guangzhou south (Shenzhen north), Shanghai Hongqiao to Guangzhou south (Shenzhen north), for which the longest running distance is less than 2400 km [2]. Based on the technical and economic characteristic of electrical multiple units (EMU) trains, it is very important to find the most cost-effective distance on mid- and long-distance high-speed railways.

On the basis of a passenger transportation market survey, this article analyzes the passenger's demand, determines the target market of HSR for mid–long-distance transportation, and investigates the reasonable operation distance of high-speed trains adaptive to Chinese situations. The results can provide technical support for the operation plan of mid–long-distance high-speed trains. A reasonable transport distance can give full play the advantages of HSR transport, meet the social demand, and improve transport efficiency and benefit.

72.2 Demand on Mid–long Transport of HSR

72.2.1 Market Segmentation of HSR

The transportation demand of passenger refers to the demand of passengers that they have the ability to pay for their movement in the social activities and in a certain period of time or certain economic conditions [3]. The demand for HSR transportation is influenced by many factors, such as economic development, people's income, urbanization, fare, and the alternatives of transportation mode.

Table 72.1 Passenger classification of HSR

Passenger classification	Sensitivity			Note
	Time	Fare	Service	
VIP passenger	**O**	**X**	**O**	Have very high requirements on the conditions of the facilities, service level and travel time, punctuality
Business passenger	o	**X**	o	The demand of comfortable, convenient and temporary changes of plan is higher
Casual passenger	x	o	x	Have plans for travel, visit, study, or work, requiring less change of itinerary
Hybrid passenger	o	o	o	Considering the economic factors, comfort, and convenient
Potential tourists	/	/	/	May change from the original transportation mode to high-speed rail

O—Very sensitive, o—Sensitive, **X**—Very insensitive, x—Insensitive, /—Not sure

Market segmentation theory can be used in analyzing the traveler's behavior of HSR. Based on a reasonable passenger classification, the railway enterprises can provide customized transportation products and improve the benefit. In view of differences of passengers' income, social status, and purpose of the trip, the HSR passengers also have different requirement and expectation for fare, service, and the traveling time.

As a public transport mode, HSR is a comparatively expensive traveling mode while it has competitive advantages with faster speed. Therefore, the travelers can mainly be classified by fare prices, services level, time sensitivity, and so on. This article will divide the HSR passenger into 5 groups: VIP passenger, business traveler, casual passenger, hybrid passenger, and potential tourist [4] (see Table 72.1).

The railway enterprises can make the market segmentation based on passenger's willing to pay, service demand, preference, etc. By conducting passenger questionnaire survey, the enterprises can get the passenger real demand under different distances and set reasonable operation distance, formulate the different marketing strategy, improve the transportation service quality, etc. Passengers then have different choices and it also provide the support for railway enterprises to achieve maximum benefit.

72.2.2 Market Surveys on Mid–long Transport of HSR

Based on the mid–long passenger traveling demand, a questionnaire is made targeting different groups of passengers from different transport industries such as air, railway, and highway. In order to know the passengers' traveling preference, we made 520 questionnaires and took back 518, accounting for 99 % of the total. A statistical analysis is done for the mid- and long-distance market mainly based on

Table 72.2 Distribution of travel mode choice in different income groups %

Income (RMB)	Travel mode					
	Civil aviation	Semi-cushioned seats	Conventional sleeping berth	HSR seats	HSR sleeping berth	Highway bus
≤2000	1.4	2.2	2.6	3.6	0.8	1.0
2001–4000	4.0	3.4	4.8	5.2	1.6	1.0
4001–6000	6.0	4.4	3.8	8.1	4.6	0.6
6001–10,000	10.3	1.4	2.4	8.7	5.8	0.0
≥10,001	7.7	1.0	0.2	1.8	1.6	0.0
Total	29.4	12.4	13.8	27.4	14.4	2.6

different preference of passengers who have different income, social status, purposes of traveling, and choices of traveling modes.

In terms of passengers' income structure, the mid–long-distance passengers with an income of 2001–10,000 RMB account for 76 % of the entire survey group. Those people selecting HSR trains (sleeping berth, first-class, and second-class seats) and airplanes mostly have an income of 4001–10,000 RMB, those selecting conventional trains (seats and sleeping berth) mostly have an income of 2001–6000 RMB and people selecting highway buses mostly have an income below 4000 RMB. See Table 72.2 for more details.

In terms of the passengers' professional structure, the main passenger flow includes ordinary company employees, managers, civil servants, and students, who account for 91 % of the entire survey group. Passengers selecting air are mainly enterprise employees, managers, and civil servants; those selecting HSR trains (first-class and second-class seats, and sleeping berth) and conventional trains (seats and sleeping berth) are mainly enterprise employees, managers, and students; and those selecting highway buses are mainly enterprise employees, students, individual businessman, workers, and farmers. See Table 72.3 for more details.

The main purposes for traveling include business, working, tourism, studying, visiting, and trade, which can be simply classified as business travel and non-business travel. For business trip, the journey time is a critical factor to be concerned; therefore, the travelers usually select the fastest traveling mode in line with the expense standard without considering too much the cost of the trip. However, non-business travelers are not so much concerned about the journey time and would consider more about their income and consumption level while selecting the traveling mode [5]. It is found in this survey that the purposes for mid–long-distance traveling are mainly tourism, business, visiting, and studying, which take a 97 % proportion of the entire group.

For the passengers' preference for traveling modes, the survey has shown that for mid–long-distance traveling, about 30 % of people would travel by air and 42 % would select HSR (first- or second-class seats or sleeping berth). HSR is getting more and more popular in the mid–long-distance transport market and becoming a strong competitor of air.

Table 72.3 Distribution of travel mode choice in different careers %

Travel mode	Careers					
	Enterprise ordinary employees	Enterprise managers	Civil servants	Individual business workers and farmers	Students	Others
Civil aviation	11.6	12.2	2.5	1.0	1.2	1.4
Semi-cushioned seats	3.7	4.5	0.6	1.2	1.9	0.4
Conventional sleeping berth	7.7	2.7	0.2	0.6	1.6	1.2
HSR seats	13.0	8.5	1.2	1.2	3.5	0.4
HSR sleeping berth	3.9	7.9	0.4	0.8	0.8	0.2
Highway bus	0.6	0.0	0.0	0.4	0.6	0.4
Total	40.5	35.8	4.9	5.2	9.6	4.0

72.2.3 Market Orientation on Mid–long Transport of HSR

HSR has the advantages of large capacity, high speed, outstanding safety, low pollution, low exterior transport cost, and better social and economical return. Nowadays, Chinese HSR has shared traffic from the conventional railways. Moreover, for its high speed and low fare, it has gradually become a strong competitor of air in the mid- and long-distance transport market.

Based on the above market survey analysis, mid–long-distance HSR service mainly target those passengers with an income of roughly 4001–10,000 RMB, including managers, enterprise employees, students, and civil servants who mainly travel for tourism, business and visiting, etc. The main competitor is air.

With rapid development of Chinese economy, the passengers have a higher expectation of transport services, which requires the railway companies to provide higher quality and diversified services to address their need. In order to improve the competitiveness of HSR, one of the measures is to provide innovative HSR service products. In addition to the existing HSR trains, which operated offering first- or second-class seats, China Railway started to operate night EMU sleeping trains from 2008, which mainly serve those passengers preferring to travel at night. For this new service, high-speed trains normally travel for mid–long distance at night, which saves the passengers a lot of time. The train also features good operation stability, for it does not suffer from external factors such as bad weather or air traffic control. Moreover, EMU sleeping train is more comfortable comparing with the limited seating space of airplane.

China has preliminarily established a HSR network covering major large- and medium-sized cities, which provides more convenience for the passengers' traveling. In order to make HSR more attractive in the mid- and long-distance transport

market, technical and economic factors shall be jointly taken into account to figure out a reasonable operation distance of high-speed trains. By learning experience of existing HSR sleeping trains, innovative mid- and long-distance HSR services shall be designed by fully exerting the advantage of HSR to address the passenger's needs and improve transportation efficiency and benefits.

72.3 Reasonable Operation Distance on Mid–long Transport of HSR

72.3.1 Operation Distance Defined by Considering the Comfort of Passengers

The travel on the train is a consumption process of physical strength. The passenger will feel tired after a certain period of time onboard and become more and more uncomfortable as time goes by despite that modern amenities are provided on the train. The rest of the people on the transportation means are just a kind of forced rest which cannot completely eliminate the feeling of fatigue [6]. The fatigue degree can be calculated by the following formulas:

$$F = F_0\left[1 + a_{\mathrm{f}}\max\left(0, t_{\mathrm{f}} - t'_{\mathrm{f}}\right)\right] \tag{72.1}$$

$$t_{\mathrm{f}} = l/v \tag{72.2}$$

$$F_0 = l/(S + l/c_{\mathrm{a}}) \tag{72.3}$$

In which: "F" refers to "Fatigue index"; "F_0" refers to the basic fatigue factor (dimensionless) relating to the comfort level; "S" refers to per occupied space on the vehicle, m^3; "c_{a}" refers to the average vibration acceleration during operation of the vehicle, ms^{-2}; "a_{f}" refers to the increase coefficient of fatigue level with time, h^{-1}; "t_{f}" refers to traveling time, h; "l" refers to the distance of travel, km; "v" refers to traveling speed; and "t'_{f}" refers to the traveling time that people begin to feel the fatigue because of a long period of riding, h. By formula (72.3), the fatigue level, calculation parameters, and indicators at different speeds can be calculated and the results are shown in Table 72.4.

The passenger market survey has shown that the time that passengers in developed countries start to feel the fatigue is 3–4 h. According to a survey made by China Academy of Railway Sciences on the Beijing–Guangzhou HSR market [7], passengers feeling fatigue in 6 h account for approximately 70 % of the total, with 4.6 h on average, indicating the longest riding time for EMU first-class and second-class seats is expected to be 4–5 h. This article assumes that the majority of mid–long-distance travelers choose EMU sleeping train so that the feeling of fatigue will be delayed. Assume that the time that passengers start to feel fatigue on a sleeping train is 4.5 h and the fatigue increases by 0.15 for each hour, by formula

Table 72.4 Fatigue level factor in EMU sleeping trains at different traveling speeds

Traveling speed (km/h)	Per capita space (S)	Average vibration acceleration (c_a)	Fatigue level factor (F_0)
250	5	2.5	0.1851
300	5	3.2	0.1882
350	5	3.7	0.1897

Table 72.5 Fatigue level of passengers in EMU sleeping train at different traveling speeds

Traveling distance (km)	Fatigue level at different traveling speeds (km/h)		
	250	300	350
1500	0.2129	0.2023	0.1897
1800	0.2601	0.2305	0.2080
2000	0.2823	0.2494	0.2243
2200	0.3045	0.2682	0.2405
2400	0.3267	0.2870	0.2568
2600	0.3489	0.3058	0.2730
2800	0.3711	0.3246	0.2893
3000	0.3933	0.3435	0.3056
3200	0.4155	0.3623	0.3218
3400	0.4378	0.3811	0.3381
3600	0.4600	0.3999	0.3543
3800	0.4822	0.4187	0.3706
4000	0.5044	0.4376	0.3869

(72.1) the fatigue level of different travel distances can be calculated. See Table 72.5 for the calculation results.

The study showed that when the fatigue level is greater than 0.4, the time for recovery will increase dramatically, which means that people need extra time to recover to normal physical and mental status, i.e., additional recovery time, equivalent to an increase of traveling time. Table 72.3 shows that the EMU trains should not run for more than 3000–3600 km when the speed is 250–300 km/h and not more than 3600–4000 km when the speed is 300–350 km/h.

In addition, the passenger survey also shows that 78 % of the passengers will feel the fatigue in 16 h while traveling by long-distance sleeping trains, in 12–14 h on average, as shown in Table 72.6. Therefore, from the perspective of riding

Table 72.6 Distribution of fatigue limit when passengers on long-distance sleeping trains %

5–8 h	8–12 h	12–16 h	16–20 h	>20 h
17.6 %	36.3 %	23.8 %	14.4 %	8.27 %

comfort, assuming that the maximum tolerable time of passengers on EMU sleeping train is 14 h; then, the reasonable operation distance for mid- and long-distance HSR trains is 3500 km at most (assuming that the average speed of EMU sleeping trains is 250 km/h).

72.3.2 Operation Distance Based on Maintenance of EMU

China railway carries out a planned maintenance for EMU, including service maintenance and regular maintenance. There are 5 levels of maintenance work, for which the first and second levels are service maintenance, and the third, fourth, and fifth levels are regular maintenance. Level 1 maintenance work is carried out for the CRH1, CRH2, CRH3, and CRH5 EMUs for a running distance of every 4000 km [8]. As the HSR network is formed, the EMU routing will be longer and the operation mode will be optimized. An EMU train may travel up to 3000 km for a long-distance trip and that for short-distance operation may only come back to the workshop for maintenance after a few round trips. Therefore, the operation distance of EMUs running on HSR shall not exceed the specified level 1 maintenance distance, namely 4000 km.

72.3.3 Operation Distance in Line with Time Efficiency

In the mid–long transportation area, HSR mainly competes with air. If we consider the critical distance to be l_{th} when HSR and aviation have the same transportation time, then we can calculate according to the following formulas:

$$l_{\text{r}}/v_{\text{r}} + t_{\text{ra}} = l_{\text{a}}/v_{\text{a}} + t_{\text{aa}} \tag{72.4}$$

$$l_{\text{r}} = l_{\text{a}} = l_{\text{th}} \tag{72.5}$$

$$t_{\text{ra}} = 2t_{\text{s}} + t_{\text{ru}} + t_{\text{rd}} \tag{72.6}$$

$$t_{\text{aa}} = 2t_{\text{p}} + t_{\text{au}} + t_{\text{ad}} \tag{72.7}$$

In which: "v_{r}," "v_{a}" refer to the operating speed of HSR and air. In China, the average operating speed of air is about 700–800 km/h, and the average operating speed of HSR is about 250–300 km/h; "t_{s}," "t_{p}" refer to the average time for passengers to arrive at the HSR stations and airports from home; here, we set "t_{s}" to be 1 h and "t_{p}" to be 1.5 h; "t_{ru}," "t_{rd}" refer to the average time for passengers to get on and off HSR trains; here, we set "t_{ru}" to be 0.5 h and "t_{rd}" to be 0.5 h; "t_{au}," "t_{ad}" refer to the average time for passengers to get on and off the plane, here we set "t_{au}" to be 1.2 h and "t_{ad}" to be 0.8 h. According to the formulas, we can work out that

the critical distance of l_{th} is approximately 800–1100 km. That means with this range the HSR takes precedence of the air for its service quality and short journey time. When the distance exceeds 800–1100 km, HSR has to compete with the air on fare and service. The average fare level of Beijing–Guangzhou EMU sleeping trains is about 0.35–0.45 RMB per passenger-kilometer and the average economy class fare of air is about 0.75 RMB per passenger-kilometer [9], by which it can be concluded that for a distance of 1500–2400 km, the competiveness of HSR is equal to that of the air.

72.4 Conclusion

Based on a market survey and taking into account the actual demands of passengers, this article defines the market positioning of mid–long-distance HSR service and proposes a reasonable operation distance of high-speed trains. By taking the EMU sleeping trains, for example, major findings of the analysis are summarized below.

First, mid–long-distance HSR trains mainly serve those passengers with an income of roughly 4001–10,000 RMB, dominantly managers, ordinary company employees, students, and civil servants traveling for tourism, business, and family visiting, etc. The competitor of high-speed railway is mainly air.

Secondly, considering factors like passenger's fatigue during traveling and maintenance of EMU, the reasonable operation distance for mid–long-distance HSR service should not exceed 3500 km given the current speed of EMU sleeping trains. Whereas, considering the competition existing with the air in terms of fare and journey time, the optimal operation distance of mid–long-distance HSR trains is 1500–2400 km.

Finally, EMU sleeping trains are still a new mid–long-distance HSR service product. If the railway enterprises plan to operate EMU sleeping trains with even longer trip, it is very essential to analyze the demand of passengers who are used to night trains, control the total journey time and have a reasonable distribution of stations. Meanwhile, the railway enterprises should learn the fare strategy of airlines to implement the revenue management and flexible pricing systems to improve the HSR service quality.

Acknowledgment This work was supported by the Fundamental Research Funds from Railway Science and Technology Research and Development Center of China Academy of Railway Sciences (NO. J2014X004).

References

1. Guang Zu S (2015) Develop Chinese railway by innovation, adapt to the new normal, and create a new reform situation. Report on the work of China Railway, pp 1–12 (in Chinese)
2. China Railway (2014) Announcement about the sunset departure and sunrise arrival EMU sleeping trains of HSR. http://www.12306.cn/mormhweb/zxdt/201505/t20150529_3000.html. Accessed 15 May 2015 (in Chinese)
3. Chaohe R (2008) Transport economics. Economic Science Press, Beijing (in Chinese)
4. Yin Z, Xiaoman S, Yu Y, Huaixiang W (2015) Fare strategy based on revenue management of high-speed railway. In: The 17th annual meeting of China association for science and technology, pp 345–349 (in Chinese)
5. Xiaoyan L, Wei B (2012) High-speed railway service and competition. Social Sciences Academic Press, Beijing (in Chinese)
6. Zhihui X, Junxia G, Chunqing L (2008) Study on the reasonable distance of electrical multiple units. J Lanzhou Jiaotong Univ 27(1):83–85 (in Chinese)
7. Transportation and Economics Research Institute, China Academy of Railway Sciences (2014) Research on the train operation plan of Beijing-Guangzhou high-speed railway under the conditions of inter-day maintenance window: 18–19 (in Chinese)
8. Ministry of Railways of the People's Republic of China (2007) Service and maintenance protocol of electrical multiple units. China Railway Publish House, Beijing (in Chinese)
9. Guangyuan Z, Dongsheng Z, Wei W (2005) Tariff management. China Market Press, Beijing (in Chinese)

Chapter 73
Modeling and Research for Underground Engineering Safety of Urban Rail Transit Based on Multiphase Flow

Yudong Tian, Guangxun Wang and Jianhong Gao

Abstract The underground engineering safety of urban rail transits is an important research direction in the research field of urban rail transit safety, and its state estimation reflects the characteristics of urban rail transit underground engineering, and then becomes a research difficulty and a key technology. Aiming at the problem, at first the underground engineering safety requirements and characteristics of urban rail transits were analyzed. Then, a multiphase flow model of the urban rail transit underground engineering safety state estimation was designed, and the relation of the state estimation and pressure testing location of the urban rail transit underground engineering safety were advanced. Moreover, the modeling structure, algorithm, simulation, and experiment of the urban rail transit underground engineering safety state estimation characteristics applied multiphase flow were particularly presented in detail. Finally, the experimental results prove that it is fast and accurate and provide the state estimation characteristic curve of the urban rail transit underground engineering safety.

Keywords Urban rail transit · Underground engineering · Safety engineering · Multiphase flow

73.1 Introduction

With the rapid development of economy and the acceleration of urbanization, urban rail transit is an important transportation for solving the traffic congestion of developed areas and the important symbol of modern city in the world. In recent years, China has vigorously developed urban rail transit in many large cities; thus, the total traffic mileage of urban rail transit is more than three thousand kilometers, and the average daily passenger flow has exceeded five million of its traffic in the

Y. Tian (✉) · G. Wang · J. Gao
Mechanical Engineering School, Suzhou University of Science and Technology, Suzhou 215009, Jiangsu, China
e-mail: tyd@mail.usts.edu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_73

mega cities such as Beijing, Shanghai, and Guangzhou [1]. Nowadays, urban rail transit has become the preferred mode of transport for people to travel in large and medium cities because of its safe, fast, and accurate, especially in the commuting rush hours and holiday peak times relative to the ground transportation.

In the growing urban traffic, urban rail transit not only promotes the city development, but also produces a certain impact on the environment and architecture, which brings new problems and challenges to the city ecological environment and security [2]. At present, the underground construction safety problems of urban rail transit, such as ground subsidence, structural collapse, building cracks, and other geological disasters and their secondary disasters, must be measured, computed, modelled and controlled in order to get a good result.. As one of social civilization construction, ecological civilization construction is becoming more and more important in China. In fact, since the birth of the subway, the subway construction detection and repair began, and the subway damage has been no interruption to the environment and the building, thus fully exposed that urban rail transit underground construction environment safety and control measures are not enough. Therefore, research on the underground engineering construction environment safety technology of urban rail transit, such as the subway building safety test way, the subway environmental safety control method, the subway geological environment change characteristics, and constantly emerging geological disasters and so on, has important practical significance.

73.2 Requirements and Characteristics of Urban Rail Transit Underground Engineering Safety

As an organic component of the urban public transport network system, urban rail transit is the main use of urban underground narrow channel space [3]. Underground construction of urban rail transit is mainly acted through two processes that are the excavation and the shield, and formed as a steel hybrid supporting structure, which need for a long-term operation after the construction.

According to the national standard of urban rail transit, underground railway engineering, and railway tunnel engineering, the useful life of the engineering design of underground building structure is 100 years in the urban rail transit [4]. Its safety requirements include the local environment condition, construction influence, geological disasters, and maintainability demand. Summarizing the domestic and foreign literature data, the urban rail transit underground construction safety can be described as the normal operation of the basic structure function.

From the structure, urban rail transit underground construction is mainly constituted of subway stations and subway tunnel, which belongs to the typical tunnel node structure [5]. The goal of the underground construction safety detection of urban rail transit is to meet the above-stated requirements by measuring and evaluating the underground construction building structure safety of urban rail transit.

The underground engineering safety of urban rail transit is a complex nonlinear system of the multi-input and multi-output that has uncertain factor and random disturbance. There are extremely complicated nonlinear relations among the temperature, humidity, pressure, flux, load, and so on. The internal environment of the subway is closed, the ventilation is complex, and the change of the flux is fast, so that the environmental safety control is a complex system control, and the process has the characteristics of the multi-volatility. Otherwise, the external geological structure of the subway is relatively stable, and the change in the water, soil, and air in the underground environment is slow, so that the environmental safety control is a civil system control, and the process has the characteristics of the large delay, uncertainty, and nonlinearity. Thus, the temperature, humidity, and pressure have important influence on the safety control of the underground construction environment in the urban rail transit. When the environmental condition changes, its dynamic characteristics are obvious, and it also has the characteristics of distributed parameters and multi-disturbance.

73.3 Underground Engineering Safety Modeling of Urban Rail Transit Based on Multiphase Flow

In the Yangtze River Delta region of China, the underground construction of urban rail transit is mostly the shield project and the underground construction geological and environmental conditions of urban rail transit is complex, so its long-term security threat is based on the seepage water and the sedimentation [6]. Here, the method of multiphase flow is used to research the water pressure parameters. At the same time, considering the actual water pressure sensor, the measuring point is generally arranged in the supporting connection of the tunnel steel-mixed block.

73.3.1 Unit Model Selection

The underground construction building of urban rail transit is basically composed of several same structures of the platform and the tunnel. Now one tunnel and two adjacent half-platform are selected as the basic unit model. Its structure is shown in Fig. 73.1.

The unit interior is the sealing structure, the vent is set in the platform, which passes into the atmosphere; the ventilation of the single-hole circular tunnel mainly relies on the piston wind when the subway cars are running. The temperature of the unit internal air is generally higher than the temperature of the unit underground building structure at the 5–10 °C. The unit geological is the mud soil-attached water with the uniform porosity.

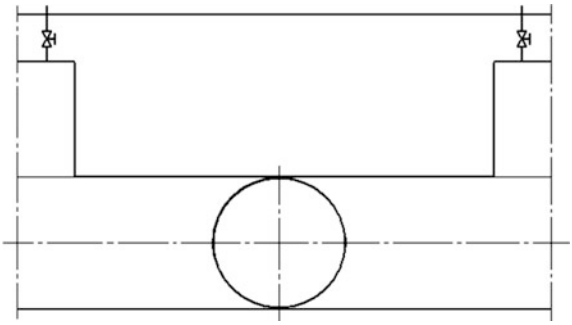


Fig. 73.1 Schematic diagram of the urban rail transit underground structure unit model

73.3.2 Modeling Assumption

For ease of modeling, the following assumptions are made:

1. The geological soil layer is a homogeneous porous medium with constant temperature, and the porosity is uniform, and the attached water is liquid.
2. The liquid water flows as laminar flow in the unconfined and confined aquifers.
3. The underground construction sealing structure completely isolates the water vapor infiltration of the geological stratum and the subway inner space.
4. The temperature distribution and pressure distribution are uniform.
5. The solubility of gas in liquid water is negligible.

73.3.3 Multiphase Flow Modeling

The change process of the underground environment is a distributed parameter system. The state estimation includes the model equation, the measurement equation, and the noise equation.

According to the energy conservation law and the multiphase flow pressure equation [7], the system dynamic model equation is given as:

$$\frac{\partial P}{\partial t} = F(P, P_x, P_{xx}, O, t) + N(x, t) \tag{73.1}$$

where P is the system state value of the pressure parameter, t is the time, F is the model equation, P_x is the first-order derivative of P to x, P_{xx} is the second-order derivative of P to x, O is the operation variables, N is the model noise (it meets the statistical characteristics of zero mean), and x is the dimensionless axial position.

Its boundary condition is given as follows:

$$\begin{aligned} x = 0, \quad & B_0(P, P_x, O_0, t) + N_0(t) = 0 \\ x = 1, \quad & B_1(P, P_x, O_1, t) + N_1(t) = 0 \end{aligned} \tag{73.2}$$

where B_0 and B_1 are the boundary condition equation, O_0 and O_1 are the boundary condition operation variables, and N_0 and N_1 are the boundary condition noise (which are all satisfied with the statistical properties of zero mean).

Let x_i^* for the position of the ith water pressure parameter measuring point (i is the natural numbers, $i = 1, 2, \ldots, n$). According to the characteristics of the distributed parameter system, the system state measurement value is given as follows:

$$m = M\left(P\left(x_1^*, x_2^*, \cdots, x_n^*\right), t\right) + N_{\mathrm{m}}(t) \tag{73.3}$$

where M is the measurement equation, and N_{m} is the measurement noise (it is satisfied with the statistical properties of zero mean).

Therefore, the dynamic equation of the system state estimation is given as:

$$\frac{\partial \hat{P}}{\partial t} = F\left(\hat{P}, \hat{P}_x, \hat{P}_{xx}, O, t\right) + C(x, t)\left(m(t) - M\left(\hat{P}(x^*, t)\right)\right) \tag{73.4}$$

where $\hat{P}$ is the optimal value of the pressure parameter system state, $\hat{P}_x$ is the first-order derivative of $\hat{P}$ to x, $\hat{P}_{xx}$ is the second-order derivative of $\hat{P}$ to x, and C is the state estimation gain matrix.

Because the gain matrix C describes the weights of the model and measurements in the state estimation, it can be expressed by the differential sensitivity matrix S of the measurement point position x_i^* as follows:

$$C(x, t) = \sum_{i=1}^{n} S\left(x, x_i^*, t\right)\left(\frac{\partial M\left(\hat{P}\left(x_i^*, t\right)\right)}{\partial \hat{P}}\right)^{\mathrm{T}} R_{\mathrm{m}}^{-1}(t) \tag{73.5}$$

where R_{m} is the covariance matrix of N_{m}.

Therefore, according to the Riccati equation [8] in the concentric method of distributed parameter system, the partial differential equation of the differential sensitivity matrix S is given as:

$$\begin{aligned} \frac{\partial S(x, x', t)}{\partial t} = {} & \frac{\partial F}{\partial \hat{P}} S + \frac{\partial F}{\partial \hat{P}_x}\frac{\partial S}{\partial x} + \frac{\partial F}{\partial \hat{P}_{xx}}\frac{\partial^2 S}{\partial x^2} + \frac{\partial F^{\mathrm{T}}}{\partial \hat{P}} S + \frac{\partial F^{\mathrm{T}}}{\partial \hat{P}_{x'}}\frac{\partial S}{\partial x'} + \frac{\partial F^{\mathrm{T}}}{\partial \hat{P}_{x'x'}}\frac{\partial^2 S}{\partial x'^2} \\ & + \sum_{i=1}^{n}\sum_{j=1}^{n} S\left(x, x_i^*, t\right)\frac{\partial\left(\partial M / \partial \hat{P}\left(x_i^*, t\right) R_m^{-1}(t)\right)\left(m(t) - M\left(\hat{P}(x^*, t)\right)\right)}{\partial \hat{P}} S\left(x_j^*, x, t\right) + R(x, x', t) \end{aligned} \tag{73.6}$$

where x' is x the corresponding dimensionless virtual axial position, $\hat{P}_{x'}$ is the first-order derivative of $\hat{P}$ to x', $\hat{P}_{x'x'}$ is the second-order derivative of $\hat{P}$ to x', and R is the covariance matrix of N.

According to the calculation formula [8] of capillary pressure p of gas water two-phase, the capillary pressure is given as:

$$p = \chi\left(\frac{\varepsilon}{K}\right)^{\frac{1}{2}} J(s) \tag{73.7}$$

where χ is the component mole fraction, ε is the porosity, J is the diffusion flux, and s is the water saturation.

73.3.4 Calculation Method

Considering that the above model is the distributed parameter system, the partial differential equation is firstly discretized into the ordinary differential equation, and then the common Runge–Kutta algorithm is used to solve the problem.

Under the relative stability of the underground geological environment, the differential sensitivity matrix S is a nonlinear function of the process state in the distributed parameter system, and its variation range is small. So it can be used to solve the specified value of a certain state, which can greatly reduce the computation quantity and can guarantee the engineering precision.

In addition, the differential sensitivity matrix S is the same as the water pressure parameter, which has the distribution characteristic. Compared with the whole monitoring time history, the integral of S variance, i.e., $S(x, x)$, must be the smallest in the domain, if the system state is estimated to be the best in space, which means that the best measuring point position is obtained. Thus, the target function INTS is given as:

$$\min(\text{INTS}) = \int_0^1 S(x, x)\mathrm{d}x \tag{73.8}$$

73.4 Analysis of Urban Rail Transit Underground Engineering Safety Multiphase Flow Modeling

The model is calculated by MATLAB on PC. While the measurement point $i = 1, 2, 3, \ldots$, the change of the differential sensitivity matrix variance $S(x, x)$ is obtained. When using a single measurement point, the change of the target function

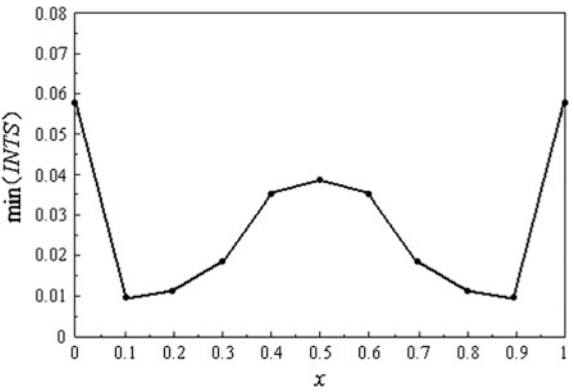


Fig. 73.2 The simulation result of the model with the measuring point location

INTS of different measurement positions is getting out. The result is shown in Fig. 73.2.

Through the above modeling and simulation experiment, the results show that the optimal measurement point location of the water pressure distribution parameters is most affected by the underground station and the underground construction defect in the case of stable underground construction geological environment of urban rail transit, where the optimal measurement point is located in the influence of the 10–20 % location.

73.5 Conclusions

Building a safe and efficient underground engineering safety of urban rail transit is helpful to save resources and promote social harmony. The state estimation of the underground engineering safety of urban rail transit reflects the characteristics of urban rail transit underground engineering. On the basis of analyzing the safety requirements and characteristics of underground engineering in urban rail transit, a mathematical model of water pressure distribution parameter state estimation based on multiphase flow is established in the paper, and the modeling structure, algorithm, simulation and experiment of the urban rail transit underground engineering safety state estimation characteristics applied multiphase flow are particularly presented in detail. The modeling has important significance for safety and monitoring of urban rail transit underground engineering.

Acknowledgment The paper is supported by the 2013 science and technology project of China Construction Ministry (Grant No.: 2013-K5-34).

References

1. Qian Q (2012) Challenges faced by underground projects construction safety and countermeasures. Chin J Rock Mech Eng 31(10):1945–1956 (in Chinese)
2. Scaman ME, Economikos L (2014) Computer vision for automatic inspection of complex metal patterns on multichip modules. IEEE Trans MCM-D 37:675–684
3. Zhang K (2014) Urban rail transit safety. Science Press, Beijing (in Chinese)
4. Tian Y (2015) Research for safety detection system of underground engineering in urban rail transits. Pro IEEE ICCSIP: 322–326
5. Yang CC, Marefat MM, Ciarallo FW (2008) Error analysis and planning accuracy for dimensional measurement in active vision inspection. IEEE Trans RA 24:476–487
6. Ying YB, Wang JP, Jiang HY (2012) Inspecting diameter and defect area of fruit with machine vision. Trans CSA 28:216–220
7. Angel I, Ruben C, Luigi O (2014) Modelling transport and real-estate values interactions in urban systems. J Transp Geogr 24(6):370–382
8. Tian Y, Wang G, Gao J (2015) Intelligent test system platform of underground building safety for urban rail transit. J Urban Rail Transit Res

Chapter 74
Relationship Between Set Covering Location and Maximal Covering Location Problems in Facility Location Application

Huaijun Peng, Yong Qin and Yanfang Yang

Abstract Covering problem is one of the classical facility location problems; set covering location model and maximal covering location model are two kinds of basic covering location models. After introducing those models, analyzing and expounding the relationship in application between the two kinds of model, the paper draws a conclusion that maximal covering location model needs to have qualification within the application, which is that its limited number of facilities is not greater than the minimum number of facilities obtained by set covering location model. Finally, the paper makes data experiment of traffic police allocation to verify the existence of relationship between the two models.

Keywords Relationship · Maximal covering location model · Set covering location model · Simplex method · Limited number of facilities

74.1 Introduction

In the production and management of government or enterprises, facility location is using scientific methods to plan the best physical location of material substance, such as facilities, equipment, factories, so that the purpose of business management can be achieved effectively and economically [1]. As one classical problem of facility location, covering problem is a series of problems about how to choose the

H. Peng · Y. Qin (✉) · Y. Yang
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Hai Dian District, Beijing, China
e-mail: qinyong2146@126.com

H. Peng · Y. Qin · Y. Yang
School of Traffic and Transportation, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Hai Dian District, Beijing, China

H. Peng · Y. Qin · Y. Yang
Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Hai Dian District, Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_74

facilities to satisfy customer's demand within a certain range [2]. In covering problem, it depends on the distance between the customer and the facility whether customers' demand is met. When the distance is less than or equal to specified value, customer's demand is covered by facility. The predefined distance is called covering distance or covering radius. Location covering model (LCM) is a mathematical model, which is abstracted out from solving the covering problem of facility location. Many problems can be solved by applying LCM, such as planning the number and location of schools, police stations, libraries, hospitals, post offices, radar sets, bank branches, shopping centers [3]. The problems to be solved are summarized in four aspects: what's the number of facilities, how to locate the facilities, what's the facilities' service capability, and how to divide the facility service.

In 1995, Daskin divided LCM into set covering location model (SCLM), maximum covering location model (MCLM) and extensions of these models [4]. SCLM is to find the minimum cost of locating facilities to satisfy all the demand. SCLM is firstly put forward by Roth [5] and Toregas [6], which is used to solve the location of emergency public service facilities such as fire control center and ambulances. MCLM is finding the facility set to meet demands as much as possible on condition of limited facility resources. MCLM is firstly presented by Church and ReVelle [7]. For many years, these two kinds of basic covering problem attract interest of researchers, spawns a variety of improved optimization models [8], which is widely used in many fields.

From a large amount of literatures that have been published over the years [9–15], SCLM and MCLM solve the problems with two different conditions and objectives, which are belong to different branches of research. In the practical application of covering problem, we usually choose which kind of model according to the goal of the problem and whether there are limited resources or not. When Church and ReVelle put forward the MCLM [7], the cost-effectiveness curve under different limited number of facilities were given from the experiment data, and they pointed out that the covering results is important to policymakers decided what was the number of facilities. Through qualitative analysis, this paper reveals the rule of MCLM, establishes a relationship with SCLM, and expounds how to cross the application of two models mentioned above in the facility location problems. After that, one experiment is applied by field data to verify the existence of the relationship. Finally, the application scene and significance of the relationship found in this paper are discussed.

74.2 Basic Covering Models

74.2.1 Notation Declaration

Assume that I is the demand set and J is the facility site set, and $i \in I$, $j \in J$. Let D_c be the distance coved by facility, and d_{ij} be the shortest distance between demand node i and facility site j. If d_{ij} is less than or equal to D_c, then demand node i is

covered by facility site j. To state the models, the variables and parameters are defined as follows:

a_{ij} 1, 0; it is 1 if facility site j can cove demand node i, and 0 otherwise;
c_j Cost of locating facility at site j;
P Total number of facilities to locate;
w_i The number of demands at node i;
X_j 1, 0; it is 1 if a facility locate at site j, and 0 otherwise;
Z_i 1, 0; it is 1 if demand node i is covered, and 0 otherwise.

74.2.2 Set Covering Location Model

SCLM is to find a minimum cost of locating facilities in a finite set of facility sites so that each demand node is covered by one facility at least. It can be formulated as follows:

$$\text{Min} \sum_j c_j X_j \tag{74.1}$$

s.t.

$$\sum_j a_{ij} X_j \geq 1 \tag{74.2}$$

The objective function (74.1) is found a facility site set to minimize the total cost. Constraint (74.2) indicates that each demand node i must be covered by at least one facility. Regardless of the costs of locating facility, we can set $c_j = 1$, and the objective function (74.1) can be rewritten as follows:

$$\text{Min} \sum_j X_j \tag{74.3}$$

74.2.3 Maximum Covering Location Model

SCLM treats all demand nodes identically while ignoring demand distribution inhomogeneity. Because SCLM covers all demands and the cover radius is the same, the place with small demand needs build facilities with same density, which

makes the construction efficiency low. In many practical applications, the objective is to maximize total satisfied demands on condition that maximum number of facilities is limited. So was born MCLM. MCLM can be formulated as follows:

$$\text{Max} \sum_i w_i Z_i \tag{74.4}$$

s.t.

$$\sum_j a_{ij} X_j \geq Z_i \quad \forall i \tag{74.5}$$

$$\sum_j X_j = P \tag{74.6}$$

The objective function (74.4) maximizes the number of covered demands. Constraint (74.5) indicates that demand at node i can be covered on condition that at least one of the selected facility sites cover node i. The lefthand side of constraint (74.5) $\sum_j a_{ij} X_j$ represents the times that node i is covered. Constraint (74.6) states that the number of facilities is no more than P.

74.3 Relationship Between SCLM and MCLM

Through analyzing the two models, we can reveal the relationship between SCLM and MCLM. First, we rewrite the formula of MCLM. Defining variable $Y_i = 1 - Z_i$, the value of variable Y_i is 1 if demand node i is not covered, and 0 otherwise. Constraint (74.5) of MCLM is equivalent to:

$$\sum_j a_{ij} X_j + Y_i \geq 1 \quad \forall i \tag{74.7}$$

From the above formula (74.7), we can deduce:

$$\sum_j a_{ij} X_j \geq 1 \text{ or } Y_i = 1 \quad \forall i \tag{74.8}$$

Replace variable Z_i by Y_i in the objective function (74.4):

$$\text{Max} \left(\sum_i w_i + \sum_i -w_i Y_i \right) \tag{74.9}$$

The first part of the formula is constant term, and negative maximum can be converted into positive minimum in second part, so the objective function of MCLM can be written as follows:

$$\text{Min} \sum_i w_i Y_i \tag{74.10}$$

Next, it is analyzed whether P has any constraints or not, which is the limited number of facilities to locate in MCLM. Assume that P is incremental variable. When P exceeds a certain value P' that demand of all nodes is covered, there is $Y_i = 0$ on condition $\forall i$, while the objective function (74.4) is equal to 0. That is to say, when $P \geq P'$, the value of objective function is minimum 0 that make the model not working properly. Therefore, the condition while MCLM work properly in practical application is the limited number P of facilities to locate must match the condition:

$$P \geq P' \tag{74.11}$$

In MCLM, when $P \geq P'$, there is $Z_i = 1$ on condition $\forall i$, so the objective function (74.4) can be rewritten as:

$$\text{Max} \sum_i w_i Z_i = \sum_i w_i \tag{74.12}$$

Because w_i is constant, the objective function is fixed value. When $P \geq P'$ and $Z_i = 1$, constraint (74.5) of MCLM is equivalent to constraint (74.2) of SCLM. Since P is assumptive variable, constraint (74.6) is equivalent to the objective function of SCLM where cost of locating facility is 1. Thus, it can be seen that P' is minimum number of facilities covering the demand of all node, which is the result of SCLM on condition of $c_j = 1$.P' can be written as:

$$P' = \text{Min} \sum_j X_j \tag{74.13}$$

SCLM involves cost of locating facility, so if MCLM also limits the locating cost rather than the number of facilities, the two models have the consistency with each other. Thus, constraint (74.6) is converted into following form, which restrains that the total cost of locating facilities is C.

$$\sum_j c_j X_j = C \tag{74.14}$$

From the above, the objective functions and constraints of SCLM and MCLM are corresponding to each other, and the two models solve different problems on different conditions. In practical application of MCLM, the limited number of facilities to locate is not greater than minimum number of facilities obtained by

SCLM. Considering cost of locating facility, the total cost of locating facility in MCLM is not greater than minimum cost of locating facility obtained by SCLM.

74.4 Cases Study

74.4.1 Experimental Method and Environment

In order to verify the relationship, this paper applies simplex method of linear programming to solve the two models by the same experimental data. As the objective function and constraints of SCLM and MCLM are linear, and the variables are 0–1 integer, those two kinds of models are belonging to 0–1 integer linear programming problem. As one basic method to solve the linear programming problem, the simplex method can get global optimal solution and can solve problems with small variable scale although its speed is slow.

In the experiment, SCLM and MCLM are applied to solve the problem of traffic police deployment. SCLM is to locate the minimum number of the traffic police to cover all demand of the area, while MCLM is to locate the limited number of traffic police to meet the maximum demand of covering area. The purpose of the experiment is to verify the existence of relationship between the two models, so many practical factors are ignored to simplify the problem, including demands, locating cost, traffic condition, distribution of traffic accidents, urban-rural effects.

The algorithm of the two kinds of models are implemented in MATLAB 7.10, which is running on the operating system with Intel core i3 350 M, CPU 2.26 GHz, 4G memory, Windows 7 64-bit.

74.4.2 Experimental Data

The field data of the experiment is a part of road network in Beijing, including main intersections and roads from the Second Ring to the Fourth Ring in Haidian District and Xicheng District. The data include 106 intersection nodes and 181 edges (the side length is the actual path length). This part of road network belongs to the core of the urban road, where the intersections are equally distributed and there is no very long road segment, so it is suited for police deployment only in accordance with the reach time of single condition.

In the experiment, we suppose that the traffic police should arrive in three minutes after the accident happens. According to the average traffic condition of Beijing traffic road, assume that the traffic police vehicle speed is 30 km/h, so the

traffic police's effective coverage radius is 1.5 km. To simplify the model calculation, it is assumed that the locations of accident and police deployed are all in the intersections; thus, all of the demands and police locations are on nodes in the road networks.

74.4.3 Experimental Results

Ignoring the number of demand, the objective function of MCLM is to get the maximum coverage node number. In the experiment, intersection nodes coverage, which is equal to the number of covered nodes divided by total number of all nodes, is used for reflecting the covering degree of the police. According to the experiment, change limited numbers of police P from 1 to 20 and repeat running the program of MCLM, and we get the cost-effectiveness curve about intersection node coverage and limited number of police that is showed in Fig. 74.1. It shows that the coverage is increasing and the increasing trend is slowing down along with the increasing of P. When P is equal or greater than 17, the coverage reaches the maximum value 100 %.

The experiment applies simplex method to solve SCLM, and we get the number of intersection nodes is 17, which is the limited number of police resources critical value when the coverage of MCLM reaches 100 %. So the experimental result is consistent with the formula (74.13).

The contrastive results of SCLM and MCLM (P = 17) are shown in Table 74.1. There are four same located nodes with same covering scope, while the rest of the located nodes are close to each other with similar covering scope.

In order to contrast the results more intuitively, the paper applies graphic method to show the results of MCLM (P = 17) and SCLM, as shown in Fig. 74.2. In the figure, location intersections label with intersection number, covering that intersection scope is distinguished by Voronoi diagram. Voronoi diagram is also called

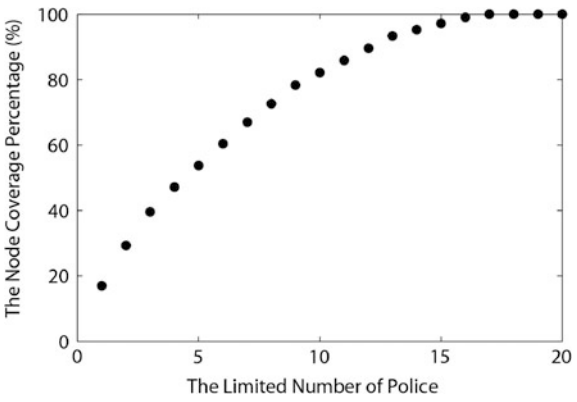


Fig. 74.1 Cost-effectiveness curve for MCLM under different limited numbers of police

Table 74.1 The contrast of MCLM (P = 17) and SCLM's calculating results

MCLM			SCLM		
Located node	Covering node	Covering number	Located node	Covering node	Covering number
2	1, 2, 37, 38	4	37	1, 2, 8, 9, 37, 38	6
13	12, 13, 14, 42, 85, 86	6	42	12, 13, 14, 42, 43, 45	6
15	15, 16, 43, 45, 87, 88, 89	7	88	15, 16, 86, 87, 88, 89	6
17	17, 18, 19, 49, 52, 56, 60, 90, 91, 93	10	60	17, 18, 19, 49, 50, 52, 53, 56, 57, 60, 61, 64, 65	13
23	22, 23, 24, 99, 101	5	22	22, 23, 78, 99, 101	5
25	25, 81, 105	3	25	25, 81, 105	3
26	26, 27, 28, 82, 83, 106	6	106	26, 27, 106	3
29	29, 30, 84	3	84	28, 29, 82, 83, 84, 85	6
32	31, 32, 33	3	31	30, 31, 32, 33	4
34	34, 35, 92, 94	4	92	34, 35, 90, 91, 92, 93, 94	7
40	10, 11, 39, 40, 41, 44, 47	7	40	11, 39, 40, 41, 44, 46, 47	7
54	3, 4, 46, 48, 50, 51, 53, 54, 55, 57, 58, 59, 61, 62, 63	15	4	3, 4, 5, 48, 51, 54, 55, 58, 59, 62, 63, 66, 67	13
69	5, 6, 7, 65, 66, 67, 68, 69, 70, 73, 76, 77	12	77	6, 7, 68, 69, 70, 72, 73, 75, 76, 77	10
71	20, 21, 64, 71, 72, 74, 75	7	21	20, 21, 71, 74	4
79	8, 9, 78, 79, 80	5	80	10, 24, 79, 80	4
98	36, 95, 96, 97, 98	5	98	36, 95, 96, 97, 98	5
103	100, 102, 103, 104	4	103	100, 102, 103, 104	4

Thiessen polygon or Dirichlet diagram, which is composed of a set of connection of two adjacent points straight line perpendicular bisector of continuous polygon. It divides the plane domain with Voronoi diagram, which is suitable for the principle of the shortest distance covering in covering model.

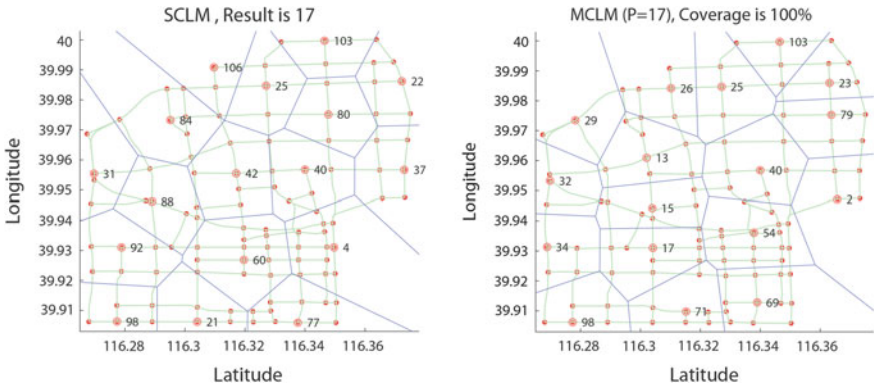


Fig. 74.2 The results of SCLM and MCLM (P = 17)

74.4.4 Experimental Analysis

It can be seen from experimental result that when the limited number of police of MCLM increases to a certain number, all intersection nodes can be covered and the result is consistent with the result of SCLM. It cannot improve the coverage of nodes adding the police.

The rule of the experimental results is same when we change the covering radius to do more experiments. When the covering radius is 1 or 2 km, the limited number of police that make the coverage of MCLM reached 100 % is 35 or 9, which is consistent with the results of SCLM. Through these experimental results, we can verify that the limited number of police of MCLM should not be more than the number of police obtained by SCLM.

In the practical application of traffic police deployment, it is limited of a city's police resources, but it does not mean that we can apply MCLM directly. First, we should calculate the minimum number of the traffic police to cover all demand with SCLM, which can be compared with existing police resources that can be deployed. If the police are enough, we can send extra police to other tasks, whereas if the police are not enough we should apply MCLM to solve the problem. It is also completely applicable to other actual covering problems.

74.5 Conclusion

SCLM and MCLM are two basic models of facility location covering problem. This paper verifies the relationship between the two models through analysis and experiment: The limited number of facilities in MCLM should not be more than the minimum number of facilities obtained by SCLM. If the cost of locating facility is considered, then the total cost of locating facility in MCLM should not be more

than minimum cost of locating facility obtained by SCLM. In the practical application of MCLM, we should firstly apply SCLM to obtain the minimum number of facilities when the demand is fully covered, which can be used to compare with the limited number of resources given. If the limited number is less, we can apply MCLM to solve; otherwise, it is unnecessary to apply all resources. This relationship has a very practical significance in the application.

References

1. Owen S, Daskin M (1998) Strategic facility location: a review. Eur J Oper Res 111:423–447
2. Fallah H, Sadigh AN, Aslanzadeh M (2009) Facility location: concepts, models, algorithms and case studies. Springer, Berlin Heidelberg, pp 145–175
3. Francis R, White J (1974) Facility layout and location: an analytical approach. Prentice Hall, Englewood Cliffs, NJ
4. Daskin M (1995) Network and discrete location: models, algorithms, and applications. A Wiley-Interscience Publication, Canada, pp 92–135
5. Roth R (1969) Computer solutions to minimum cover problems. Oper Res 17:455–465
6. Toregas C, Swain R, ReVelle C, Bergman L (1971) The location of emergency service facilities. Oper Res 19:1363–1373
7. Church R, ReVelle C (1974) The maximal covering location problem. Pap Reg Sci 32:101–118
8. Farahani RZ, Asgari N, Heidari N, Hosseininia M, Goh M (2012) Covering problems in facility location: a review. Comput Ind Eng 62:368–407
9. Current JR, Storbeck JE (1988) Capacitated covering models. Environ Plan 15:153–163
10. Boffeya B, Narula S (1998) Models for multi-path covering-routing problems. Ann Oper Res 82:331–342
11. Murray AT, Tong D, Kim K (2010) Enhancing classic coverage location models. Int Reg Sci Rev 33:115–133
12. Berman O, Krass D (2002) The generalized maximal covering location problem. Comput Oper Res 29:563–581
13. Church R (1984) The planar maximal covering location problem. J Reg Sci 2:185–201
14. ReVelle C, Hogan K (1989) The maximum availability location problem. Transp Sci 23:192–200
15. Jia H, Ordonez F, Dessouky MM (2007) Solution approaches for facility location of medical supplies for large-scale emergencies. Comput Oper Res 52:257–276

Chapter 75
Harmony Assessment of Network Transportation Capacity in Urban Rail Transit

Man Jiang, Haiying Li and Lingyun Meng

Abstract This paper is based on the state of the art in urban rail transit network transportation capacity and its demand for harmony of network planning and operation. It clarifies the concepts of urban rail transit network transportation capacity and its harmony. Then, this paper builds up an index system and analytic hierarchy process (AHP) is used. Data of urban rail transit network are used to calculate all the related coefficients and the final result, which verify the validity of this method.

Keywords Urban rail transit · Network transportation capacity · Harmony · Assessment index system

75.1 Introductions

Nowadays, both narrow urban construction land and high people density cause limitation of urban traffic development. Rail transit is a perfect mode to relieve urban traffic pressure. It has advantages of large volume and high speed. Also it is more safe, reliable, punctual, and comfortable for passengers. Large-scale construction of urban rail transit will be a trend in Chinese urban public traffic and urban rail transit

M. Jiang · H. Li (✉) · L. Meng
State Key Laboratory of Rail Traffic Control & Safety,
Beijing Jiaotong University, Beijing 100044, China
e-mail: hyli@bjtu.edu.cn

L. Meng
e-mail: 13120855@bjtu.edu.cn

M. Jiang · H. Li · L. Meng
School of Traffic and Transportation, Beijing Jiaotong University,
Beijing 100044, China

M. Jiang · H. Li · L. Meng
Beijing Research Center of Urban Traffic Information Sensing and Service Technologies,
Beijing Jiaotong University, Beijing 100044, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_75

will be characterized with networking. There exist transfers among rail transit lines. And thus, harmony research of urban rail transit network transportation capacity is an urgent subject to develop urban traffic and relieve urban traffic pressure.

There are many researches on capacity calculation. Ruegg [1] reviewed and summarized all kinds of assessment methods in definitions, restraints, and application scopes and then applied the methods to numerical examples. Widodo [2] developed a capacity prediction system in VB. The system has advantages of both optimization method and simulation method. Ming [3] divided harmony into different aspects, developed an assessment index system of rail transit network operation, and applied fuzzy comprehensive assessment method to a numerical example.

This paper researches on harmony of urban rail transit network transportation capacity and an assessment index system in Sect. 75.2. A method of AHP is used to analyze the system, and data of urban rail transit network in a city are used to calculate all related coefficients and the assessment result in Sect. 75.3. In Sect. 75.4, some concluding remarks are made.

75.2 Harmony Assessment of Urban Rail Transit Transportation Capacity

75.2.1 Structure of Assessment Index System

There are many researches on harmony in urban rail transit network transportation capacity [4]. In this paper, it is divided into three aspects. They are harmony between delivery capacities of two different lines, harmony between capacity of lines and collection and distribution capacity of stations and harmony between network transportation capacity and passenger demand. Referring to related researches [5, 6], the structure of assessment index system is as follows (Fig. 75.1).

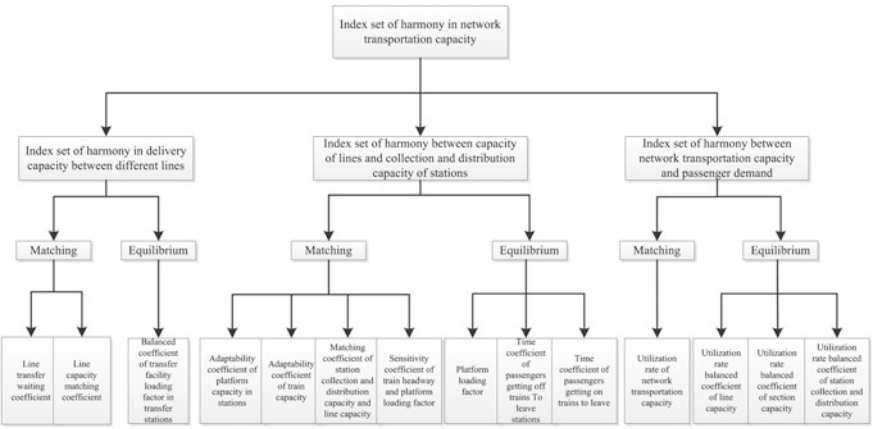


Fig. 75.1 Index system of harmony in urban rail transit network transportation capacity

75.2.2 Assessment Index Set of Harmony in Urban Rail Transit Network Transportation Capacity

75.2.2.1 Assessment Index Set of Harmony in Delivery Capacities Between Different Lines

Indexes of matching

(1) Line transfer waiting coefficient (C_{tw})

$$C_{\text{tw}} = \frac{T_{\text{tw-ave}}}{T_{\text{t-ave}}} \tag{75.1}$$

where $T_{\text{tw-ave}}$ is line average transfer waiting time and $T_{\text{t-ave}}$ is average transfer time.

Illustration: "Transfer waiting time" is the time for passengers to take from beginning to wait on platforms to getting on trains (in minutes). "Transfer time" is the time for passengers to take from getting off a train to getting on another train after transfer, including transfer walking time and transfer waiting time (in minutes).

(2) Line capacity matching coefficient ($C_{\text{dc-mat}}$)

$$C_{\text{dc-mat}} = \frac{V_{\text{tA-B}}}{\text{DC}_{\text{avail-B}}} \tag{75.2}$$

where $V_{\text{tA-B}}$ is transfer volume from line A to B and $\text{DC}_{\text{avail-B}}$ is available capacity of line B.

Illustration: This coefficient is the number of people who transfer from line A to line B divided by the number of people who can be delivered at most within the scope of train capacity on line B.

Indexes of equilibrium

(1) Balanced coefficient of transfer facility loading factor in transfer stations ($C_{\text{DoL-tf}}$)

$$C_{\text{DoL-tf}} = \frac{\min\{\text{DoL}_{\text{tf-A}}, \text{DoL}_{\text{tf-B}}\}}{\max\{\text{DoL}_{\text{tf-A}}, \text{DoL}_{\text{tf-B}}\}} \tag{75.3}$$

where $\text{DoL}_{\text{tf-A}}$ is transfer facility loading factor on line A and $\text{DoL}_{\text{tf-B}}$ is transfer facility loading factor on line B.

Illustration: A transfer station connects line A and line B. Transfer facility loading factor in a station is ratio of the number of transfer people in transfer facility divided by the number of people who can be delivered at most in it.

75.2.2.2 Assessment Index Set of Harmony Between Capacity of Lines and Collection and Distribution Capacity of Stations

Indexes of matching

(1) Adaptability coefficient of platform capacity in stations ($C_{\text{phc-sta}}$)

$$C_{\text{phc-sta}} = \frac{\sum \text{MDC}_{\text{train}}}{\text{PHC}} \tag{75.4}$$

where $\text{MDC}_{\text{train}}$ is maximum train capacity and PHC is platform capacity.
Illustration: "Maximum train capacity" means number of passengers who can be delivered by trains at full-load ratio (in persons).

(2) Adaptability coefficient of train capacity ($C_{\text{dc-train}}$)

$$C_{\text{dc-train}} = \frac{\text{PHC}}{\sum \text{MDC}_{\text{train}}} \tag{75.5}$$

(3) Matching coefficient of station collection and distribution capacity and line capacity ($C_{\text{sta-line}}$)

$$C_{\text{sta-line}} = \frac{\max\{\text{CDC}_{\text{sta}}, \text{DC}_{\text{line}}\} - \min\{\text{CDC}_{\text{sta}}, \text{DC}_{\text{line}}\}}{\min\{\text{CDC}_{\text{sta}}, \text{DC}_{\text{line}}\}} \tag{75.6}$$

where CDC_{sta} is station collection and distribution capacity and DC_{line} is line capacity
Illustration: We need to compare different capacity matching coefficients of adjacent stations and sections and then choose the minimum coefficient of the whole line.

(4) Sensitivity coefficient of train headway and platform loading factor ($C_{\text{h-Dol}}$)

$$C_{\text{h-Dol}} = \frac{R_{\text{DoL}}}{R_{\text{h}}} \tag{75.7}$$

where R_{DoL} is change rate of platform loading factor and R_{h} is change rate of train headway.

Indexes of equilibrium

(1) Platform loading factor (DoL)

$$\text{DoL} = \frac{\text{FV}_{\text{plat}}}{\text{PHC}} \tag{75.8}$$

where FV_{plat} is flow volume in the waiting area of platforms and PHC is platform capacity.

(2) Time coefficient of passengers getting off trains to leave stations (TC_{off})

$$TC_{off}=\frac{T_{off\text{-}ave}-T_{off\text{-}min}}{T_{off\text{-}max}-T_{off\text{-}min}} \tag{75.9}$$

where $T_{off\text{-}ave}$ is average time for passengers to get off trains to leave the station, $T_{off\text{-}max}$ is maximum time for passengers to get off trains to leave the station, and $T_{off\text{-}min}$ is minimum time for passengers to get off trains to leave the station.

Illustration: "Minimum time of passengers to get off trains to leave the station" is the time for passengers to leave the station in a free condition. "Maximum time of passengers to get off trains to leave the station" is the time for passengers to leave the station when the station is in the most crowded condition.

(3) Time coefficient of passengers getting on trains to leave (TC_{on})

$$TC_{on}=\frac{T_{on\text{-}ave}-T_{on\text{-}min}}{T_{on\text{-}max}-T_{on\text{-}min}} \tag{75.10}$$

where $T_{on\text{-}ave}$ is average time for passengers to get on trains to leave, $T_{on\text{-}max}$ is maximum time for passengers to get on trains to leave, and $T_{on\text{-}min}$ is minimum time for passengers to get on trains to leave.

Illustration: "Maximum time of passengers to get on trains" is in the condition of the longest headway. "Minimum time of passengers to get on trains" is in the condition of the shortest headway.

75.2.2.3 Assessment Index Set of Harmony Between Network Transportation Capacity and Passenger Demand

Indexes of matching

(1) Utilization rate of network transportation capacity ($UT_{tc\text{-}net}$)

$$UT_{tc\text{-}net}=\frac{TV_{net}}{TC_{net}} \tag{75.11}$$

where TV_{net} is network traffic volume and TC_{net} is network transportation capacity.

Indexes of equilibrium

(1) Utilization rate balanced coefficient of line capacity ($URC_{dc\text{-}line}$)

$$URC_{dc\text{-}line}=\frac{\sum L_{ur>x-line}}{TL_{line}} \tag{75.12}$$

where $L_{\text{ur>x-line}}$ is length of lines whose utilization rate of line capacity is over X in network and TL_{line} is total length of lines in network.
Illustration: Utilization rate of line capacity X can be appointed by assessment conditions.

(2) Utilization rate balanced coefficient of section capacity ($\text{URC}_{\text{cc-sect}}$)

$$\text{URC}_{\text{cc-sect}} = \frac{\sum N_{\text{ur} > \text{x-sect}}}{\text{TN}_{\text{sect}}} \tag{75.13}$$

where $N_{\text{ur>x-sect}}$ is number of sections whose capacity utilization rate is over X in network and TN_{sect} is total number of sections in network.
Illustration: Utilization rate of section capacity X can be appointed by assessment conditions.

(3) Utilization rate balanced coefficient of station collection and distribution capacity ($\text{URC}_{\text{cd-stat}}$)

$$\text{URC}_{\text{cd-stat}} = \frac{\sum N_{\text{ur} > \text{x-stat}}}{\text{TN}_{\text{stat}}} \tag{75.14}$$

where $N_{\text{ur>x-stat}}$ is number of stations whose utilization rate of collection and distribution capacity is over X in network and TN_{stat} is total number of stations in network.
Illustration: Utilization rate of station collection and distribution capacity X can be appointed by assessment conditions.

75.3 An Assessment Method on Urban Rail Transit Network Transportation Capacity

Of assessment methods [7, 8], AHP is a decision-making method. It divides factors which are always related to decision making into different levels, such as targets, criteria, and projects. In this paper, this method will be applied in the following four steps.

(1) We choose assessment index system of harmony in urban rail transit network transportation capacity as the hierarchy structure model.
(2) We choose expert scoring method. As to factors in the same level which are belong to a factor in an upper level, we use 1–9 to score the importance of each two indexes, and larger the score is, more important one index to the other is. Then, we will get a paired comparison matrix.
(3) As to each paired comparison matrix, we will calculate the largest eigenvalues and corresponding eigenvectors. Then, we will do the consistency tests with consistency indexes, random consistency indexes, and consistency ratio. If the

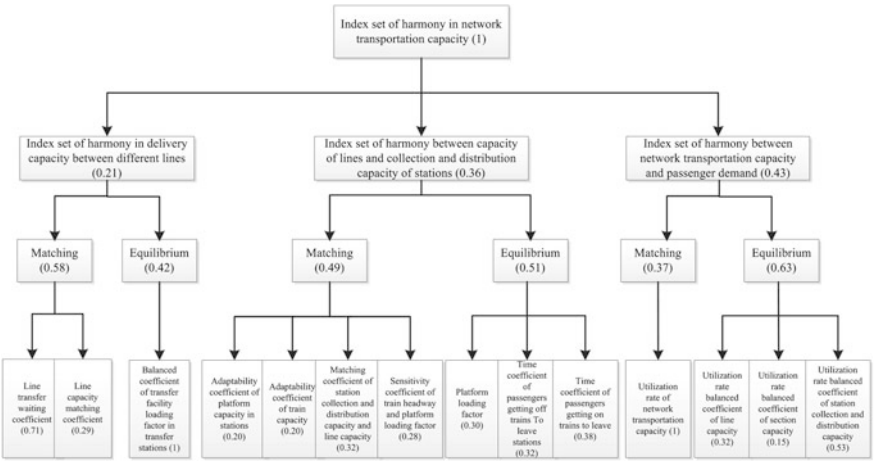


Fig. 75.2 Expert scoring weight results

tests are all passed, eigenvectors after normalization are just the weight vectors of indexes in the same level.

(4) We will calculate the combined weight vectors and do combined consistency tests. If the tests are all passed, we can get assessment value of harmony in urban rail transit network transportation capacity according to results of combined weight vectors.

And a numerical example will be used to verify the validity of this method. This paper uses data of urban rail transit network in Beijing and expert scoring weight results as the following Fig. 75.2. The final assessment result of the city is 0.69.

75.4 Concluding Remarks

We proposed an index system including fourteen coefficients and presented a numerical example to verify its validity in this paper. While along with successive construction of urban rail transit network, more new characteristics will turn up. The index system in this paper must be updated continuously with the characteristics so that theory research on harmony can adapt to practical demands and promote development of urban rail transit. Our future research aims to classify harmony assessment results into different levels. As to each level, there is analysis on what capacity is not enough and what measures we can take to improve the capacity.

Acknowledgments The work was jointly supported by State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (No. RCS2015ZZ002). Authors are of course responsible for all results and opinions expressed in this paper.

References

1. Ruegg R, Jordan G (2007) Overview of evaluation methods for R&D programs. In: A directory of evaluation methods relevant to technology development programs, prepared for US Department of Energy, Office of Energy Efficiency and Renewable Energy
2. Widodo A, Hasegawa T (2009) A study on the effective road capacity and the modified road capacity: new evaluation parameters for traffic flow evaluation. In: Intelligent transportation systems, pp 617–621
3. Ming Z (2010) Fuzzy assessment of harmony in urban rail transit network operation. Transp Inf Saf 28(3):38–41 (in Chinese)
4. Ming Z, Ruihua X, Ke Y (2009) Research of harmony in urban rail transit network operation and organization. Urban Railway Transp Res 11:44–48 (in Chinese)
5. Shuangji Z (2008) Influencing factor analysis of urban road network loading capacity. Transp Syst Eng Inf 8(l1):92–97 (in Chinese)
6. Kasikitwiwat P, Chen A (2005) Analysis of transportation network capacity related to different system capacity concepts. J East Asia Soc Transp Stud 6:1439–1454
7. Confessore G, Liotta G, Cicini P et al (2009) A simulation-based approach for estimating the commercial capacity of railways. In: Proceedings of the 2009 winter simulation conference (WSC), pp 2542–2552
8. Abril M, Barber F, Ingolotti L et al (2008) An assessment of railway capacity. Transp Res Part E: Logistics Transp Rev 44(5):774–806

Chapter 76
Research on Framework and Key Technologies of Urban Rail Intelligent Transportation System

Yong Qin, Baojun Yuan and Si Pi

Abstract This paper introduces the framework of urban rail intelligent transportation system (URITS), and four key technologies are described: dispatch command and operations management, communication and signals, safety monitoring and operation, and passenger services. Finally, the development trends of URITS are summarized.

Keywords Urban rail intelligent transportation system · System framework · Active safety assurance · Fully automatic operation · Interoperability

76.1 System Framework of URITS

URITS is the new generation urban rail transportation system, which integrates advanced information processing technology, communication technology, control and system technology, computational intelligence, and decision support technology. Based on realizing function of the state perception, information seamless transmission, intelligent information processing and sharing, the new transportation

Y. Qin (✉) · S. Pi
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing 100044, China
e-mail: yqin@bjtu.edu.cn

S. Pi
e-mail: spi@bjtu.edu.cn

Y. Qin · S. Pi
Beijing Engineering Research Center of Urban
Traffic Information Intelligent Sensing and Service Technologies,
Beijing Jiaotong University, Beijing 100044, China

B. Yuan
China Railway Construction Investment Co., Ltd., Beijing 100055, China
e-mail: yuanbaojun@vip.sina.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_76

system with more efficiency, safety, service quality and environment-friendly can be achieved, and its typical characteristics are safe, efficient, green, comfortable and sustainable [1].

The composition of URITS is a typical hierarchical structure. It is divided into five layers, namely perception layer, communication layer, integration layer, operation layer, and service layer, as shown in Fig. 76.1. The perception layer can capture and store real-time data of the mobile equipment, the fixed infrastructure, and the environment in railway system by taking use of advanced sensor technology and equipment. According to the high-capacity communication, the communication layer can transfer the data which are collected by the perception layer among the vehicle and the ground equipment, and provide data resources for fusion layer. With the support of the communication layer and the capabilities of distributed computing, data integration, and data visualization, the integration layer can be used to integrate process and display data, so as to provide support for management and decision of the operation layer. The operation layer corresponds with the functions of business organization and management of the urban rail transportation system. Based on the information provided by the integration layer, it can realize the function of systems such as communication signals, dispatch command, integrated supervision, and operations management. And then provides

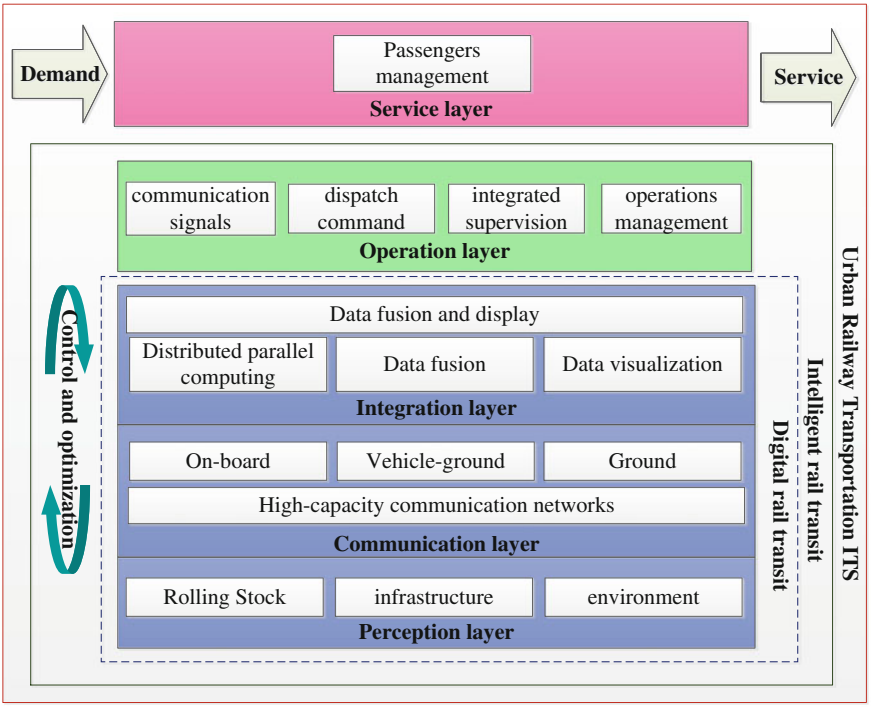


Fig. 76.1 Framework of URITS

support to the service layer. The service layer's main function lays in users' need-based, comprehensively considering the current rail transit operation management actual situation and providing passenger service.

76.2 Key Technologies

76.2.1 Dispatch Command and Operation Management Technology

Traffic control center (TCC) undertakes responsibility for commanding and coordinating rail transportation network, clearing, accounting and operations management of the automatic fare collection system, the command and control of the line. Command center is composed by functional facilities such as OCC, TCC, ACC, and matched office management.

Operating control center (OCC), a single line operation control center, is the management mode of a line a center. The subway control center is the "central point" of dispatching, commanding, and monitoring the whole subway operations such as the whole train operation, electric power supplying, station equipment operation, fire alarming, environmental monitoring, ticket management, and passenger services. TCC, the traffic control center, is the management mode of a whole city at a center. TCC system is an integrated command platform that is integrated by operation monitoring, data sharing, and emergency command function, which can realize the real-time monitoring and management of information such as network videos, train operation, power supplying, passenger flows, and disasters. Besides, the system builds auxiliary decision-making database, which integrates the relative information such as lines' professional drawings, videos, disasters. Build the emergency commanding platform based on emergency plan management, which can be done through a variety of communications emergency command [2]. Build the auxiliary analysis system, which can assist network operation coordination. So far, the construction of our metropolis rail transportation command center is developing in the direction to the TCC mode.

76.2.2 Communication and Signals Technology

76.2.2.1 Car-Ground Broadband Wireless Communication Technology

At present, new wide-area broadband wireless access technology such as "TDLTE" 4G, WiMAX wireless communication technique has a complete development and is gradually applied to the rail transportation industry. It solves the subway monitoring

transmission problem in the past, not only realizes the passenger information system and the on-board video monitoring network hosting, but also realizes the running train and ground control center of real-time data transmission, strengthens the control center for car interior personnel, train running state of real-time control, enhanced safety operation, and emergency handling abilities.

LTE technology has the characteristics of high-speed data transmission, low delay, and high mobility. LTE system adopts TDD mode, which can also offer broadband uplink business channels and satisfy car HD video monitoring data through wireless real-time demand comes back. Furthermore, the system has perfect quality of service (QOS) mechanism, which can effectively handle the service priorities (minimize the system time delay and packet loss probability in a network congestion condition), and provide both PIS and video monitoring service at the same time. This advanced mechanism not only effectively reduce the network construction investment, but also provides the opportunity to overlay the new wireless services.

LTE technology can provide a solution combined with Tetra, can further satisfy subway operators' PIS, vehicle surveillance, communication-based train control system (CBTC), and broadband cluster-rich metro application requirements, which provide two-way, continuous, highly reliable wireless transmission channel for car between voice, data, video, and other business, and enhances the comprehensive capacity of rail transportation.

76.2.2.2 Full-Automatic Self-driving Technology

Recently, self-driving technology based on fully automatic operation (FAO) mode [3] has been developed in the urban rail transportation industry in China. For example, south island line of Hong Kong Metro network has developed the driverless automatic train based on automatic control system, open driver's cab, derailment inspection system and so on. Although the train is fully automatic, but with driving system, when is needed, the staff will drive the train. Fully automatic system can be more flexible to deploy the train, if the passengers increase suddenly, which can transfer the train to join the service immediately and save the time to mobilize staff.

Self-driving technology makes a breakthrough and marks a higher level of automatic control in our country; vehicle-integrated fault diagnosis level reaches a new stage, which also represents the highest level of the mechanical, electronic, electrical, and control technology applied in rail transit vehicles. Now our country's lines that adopt self-driving are Hong Kong international airport passenger MRT system, Disneyland Resort Line, South Island Line, Beijing airport line, Shanghai's subway Line 10, Guangzhou pearl river new city transit line (APM line), Zhengzhou subway line 1.

76.2.2.3 Interoperability Technology

With the rapid development of regional urban economy in our country, subway, the development trend of urban rail network is integration, networking, and connectivity. The integration and networking of urban rail transportation require the unified standard of communication and signal devices of different lines, connectivity of the system, the integration sharing of resources, synchronous construction of transfer hubs, and collaborative efficiency of management system. Sequentially ensure that network security improves network efficiency and utilizes resources high effectively. The keys of interoperability are the unified signal system standard in the whole rail network and the compatibility of devices of different lines and different signals suppliers. So the unified train operation control system for the whole network can be established on the different kinds of signal equipment.

76.2.3 Safety Monitoring and Operation Technology

With China's urban rail transportation operation mileage and traffic intensity increasing urban rail trains face unprecedented security pressure, it is urgent to adopt advanced safety assurance technology. Recently, the train safety assurance technologies have been developed rapidly including train on-board sensor network, train key equipment fault diagnosis, car-ground wireless data transmission, risk mining, and maintenance plan optimization. The integrated system of train on-board safety warning, emergency disposal, and operation decision supporting and the first intelligent urban train have been completed in China. This will promote the safety assurance mode transit from passive to active prevention, maintenance mode transit from the plan to condition-based maintenance.

Urban rail transportation video monitoring system is the important guaranteed system of urban rail operation public safety, which shows a trend of high-definition digital and intelligent analysis. The current video monitoring technology develops through three development cycle from analog networks, based on the number of DVR to today's digital network. The advantages of digital reflects in the overall system high interference, networking convenience, the openness of the interface or else the advantages of high-definition reflects in more detail and information provide a great possibility and imagination space for visual management, and intelligent analysis opens the infinite space for the deep application of video data. Video monitoring system provides basic video resources for SCADA, environmental control, shielded gate, fire control, safeguard, else it can offer intelligent analysis information of passenger-flow analysis and equipment exception for operations management system, which will bring in safe city and intelligent city's monitoring system.

76.2.4 Passenger Service Technology

The integration and full automation of passenger service is the main trend of the current customer service system development, with the closely connecting of the transportation means such as the urban rail transportation system, urban public transport and suburban railway system, there is a need to establish the integration of passenger service system from the aspects of travel information and ticketing services. The focus lies in the construction of travel information services, transfer of cohesive devices and information service, and cartoonish ticketing services. As the amount and type of passenger service facilities are increasing, fully automated and unmanned has become the main trend of passenger transport service equipment development.

An important point of passenger service is realizing a complete coverage of information service, now with the promotion of the wireless communication technology development, WiFi subway covered the wireless Internet system of subway is becoming a hot spot of development. Subway system can send the subway users WiFi signal and provide the subway operation information and Internet services through the online system. At present, some cities' rail transportation lines are trying free wireless service.

76.3 Development Trends

76.3.1 Digital Urban Rail Transportation

Digital urban rail transportation is an electronic information system that in the rail transportation planning, construction, and operation management, make full use of digital information processing technology and network communication technology, rail traffic information associated with geographic information according to the standard of unity, form 3D digital information, and base on this to manage, analysis, and auxiliary decision. It is helpful to the integration and utilization of urban rail transportation system and outside the city's digital information and the subway information resources, form a unified standards and specifications. The construction of the digital rail transportation has profound social significance and economic benefits to the leap-type development of rail traffic information, the improvement of transport efficiency and scientific management, and the ability of handling in the face of emergency.

76.3.2 Services Based on Cloud Platform

In the rail transportation field, huge data quantity and complex business process logic have brought challenge to the efficient operation of business process. Now

with the depth of the ICT technology integration, cloud computing is more and more prominent to the advantageous ability of dealing with large amount of data and complicated transactions system.

Cloud information platform is a service platform that can provide information exchange, data storage, analysis and calculation, and the resources sharing. According to the construction of urban rail transit cloud platform, it will bring benefits to the urban rail transportation. Centralized business system construction can support integration of rail transit's entire network unified operation. Shared hardware resources and the ability of reducing the additional storage and computing can reduce the IT investment through storage and server virtualization. Information platform of cloud after integration makes information sharing between businesses, improves the business system of seamless collaboration, quick linkage efficiency, provides the powerful analysis tool for the subway operators, realizes real-time acquisition, powerful data mining, and improves the intelligent level of operations.

76.3.3 Active Safety Assurance

Rail transportation operation safety is the core competence of the rail transportation; it also faces with more and more pressure and challenges, traditional security and operational technology cannot satisfy its development demand. Active safety can make a real-time monitoring and analysis of the hidden danger and risk of rail transportation system and operating status of equipment, and actively prevent the accident and equipment fault from happening. It can promote the mode transform from passive response to active prevention, also built the condition-based maintenance mechanism. So far the rapid development of related technology such as the Internet of things and big data promotes the remarkable progress in the field, along with a large-scale use of our country rail transportation safety inspection new technology, monitoring data is growing, It becomes research focus on rail transportation management that how to make full use of the monitoring system of rail transit active safety system is established and then effectively support management of rail transit in terms of operation and maintenance.

76.3.4 Green Rail Transportation and Interoperability

Green rail transportation has obvious advantages than other forms of transportation in the aspect of energy saving, efficiency, safety, total quantity of harmful gas emissions, and noise. But because of its huge, complicated, works cross city area's system engineering, urban ecological environment will be influenced by the construction of track traffic at varying degree during construction period and operation

period, such as noise pollution, vibration pollution, electromagnetic interference. Because of its large capacity characteristics, according to the development of urban rail transit speed of our country at present, the energy consumption of the urban rail transportation will achieve considerable scale, and large energy consumption still has energy-saving potential. In this situation, green transportation was proposed, which aim to maintain the sustainable development of urban traffic system, satisfy the demands of people of the traffic, realize the maximum transportation efficiency, friendly develop with the urban environment. So far there has a rapid development in the aspects of new energy and energy storage-type vehicles, vehicle light weighting, energy-saving driving and control, comprehensive energy management, main passive noise reduction, multi-type of urban rail transit system, etc.

The rapid development of China's regional urban economy will promote the area of subway, light rail, the inter-city rail transportation toward the integration, networking, and interconnectivity. The future research emphasis focuses on the integration of traffic network plan, the interconnectivity of regional rail network, and the interoperability of operation equipment.

References

1. Zourbakhsh M (2010) Role of intelligent transportation systems in urban sustainable development. In: 1st international conference on sustainable urbanization (ICSU), pp 1437–1443
2. D'Acierno L, Gallo M, Montella B, Placido A (2013) The definition of a model framework for managing rail systems in the case of breakdowns. In: 16th international IEEE conference on intelligent transportation systems (ITSC 2013), pp 1059–1064
3. Masahiro F, Ryuji T, Hitoshim T, Naoji U, Junichi S (2005) Full-automatic operation system for Nanakuma line subway in Fukuoka. Hitachi Rev 54(4):185–192

Chapter 77
Framework Design of Urban Traffic Monitoring and Service System

Qian Li, Jizhi Wan and Guanghui Cao

Abstract Urban traffic information construction still exists some systemic problems, such as insufficient coordination, scattered resources, unbalanced development level, etc. The article presents the urban traffic monitoring, and the service system framework. The system is composed of decision support, information services, coordinated and integrated monitoring, analysis, and resource management. The system is used to promote urban transport information resources' exploitation and utilization, and promote transportation system network linkage and collaborative applications.

Keywords Intelligent transportation systems · Transport information service · Demand systems · Functional framework · System architecture

At present, our country is in a crucial period for in-depth development of new towns [1], and urbanization continued to improve in the future, for the construction and development of urban traffic information service system which provides a broad stage. By 2020, the urbanization rate of the resident population of the city may be 60 %, which adds about 100 million urban population. Urban population growth will lead to further growth in demand for passenger and freight traffic, and the promotion of traffic operation efficiency and service level will face a severe test [2, 3].

In recent years, traffic informationization of some major cities (Beijing, Shanghai, Guangzhou, and so on) is based on the construction of network platform and standard system, with a focus on business applications and front-end monitoring infrastructure, business applications, security monitoring, emergency management, and information services, and has achieved remarkable results [4–6]. Although urban traffic informatization construction playing an important support role in nowadays city construction, there are still some overall and systemic problems existed, such as insufficient overall management, scattered resources,

Q. Li · J. Wan · G. Cao (✉)
Henan Provincial Communications Planning, Survey and Design Institute Co., Ltd.,
No.70 Long Hai Road, Er Qi District, Zhengzhou, China
e-mail: guanghui1235@163.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_77

uneven development level, imperfect standard management system, relatively weak management and maintenance power and so on, for example, information island, valuing construction but neglecting service, system coordination problems etc. Therefore, strengthening the interconnectivity among different means of transportation and cross-regional traffic information management by depending on effective running mechanism and modern information technology will become an significant way to help urban traffic system fully create overall benefits.

77.1 Problems in Urban Traffic Information Service

1. Data access is not comprehensive and sophisticated, which has a big gap about dynamic monitoring. It has been basically formed the integrated traffic data systems, which is covering the various modes of transport and the field of related industries. The overall dynamic of urban traffic operation can be basically reflected. But the depth of the uneven access to data, data granularity, and timeliness cannot meet the requirements of traffic dynamic operation monitoring, which still has a great distance between from the requirements of urban transport running dynamic monitoring. There is a certain gap in terms of access to traffic data range and depth of application. Four functional urban traffic information services, "routine traffic operation monitoring, forecasting and early warning, decision support, and information services," are not yet fully realized.
2. The integrity of information systems is not strong, and the showing of comprehensive efficiency is not enough. Existing applications are mostly developed according to the needs of individual application, and also the lack of synergy between systems. A unified working environment for the urban traffic operation monitoring and application is not formed. The support efforts for a wide range of co-regulation and large areas surveillance and monitoring are not enough. Intercity transportation, such as civil aviation, railways, and other local refinement of the continued coordination between the transport linkages, needs to be strengthened. The coordinated support among ground bus, rail transport, and road network running is not enough. The means of daily operation monitoring are weak, and monitoring indicator system is not yet perfect, which cannot effectively meet the city's demand for traffic monitoring, prediction, and early warning. As Traffic monitoring report was edited by hand, efficiency is relatively low.
3. Lack of deeper association mining and integrated applications and customized service are incompetence. Clear and single condition of each mode of transportation businesses area, low realization degree of mutual integration and information sharing, insufficient specifity and pertinence of support services aiming at each business sector, all of the above factors affect the overall benefits. The development and utilization of information resource is not high, which is not deepened application in decision support, business rules, business process reengineering, management, and external services. It is difficult to deliver

customized and targeted assistance decision support. Interdisciplinary association analysis is insufficient to provide adequate support for the transportation planning, strategic analysis, and decision making, which has become an important reason for restricting integrated transport emergency response and rapid development of traffic information services.

4. Standardization of data still need to improve. The construction of standard specification in the construction of existing urban traffic information system industry cannot keep up with traffic, and low standards have been implemented. Intelligent standard system has not formed, which caused the low degree of about traffic running data access, analysis, application, and service systems. Traffic information compatibility between the transport sectors is poor, which cannot achieve effective information sharing.

77.2 Urban Traffic Operation Monitoring and Service Framework

Based on the urban traffic information construction, urban traffic operation monitoring and service provides all kinds of real-time, dynamic and customized transportation information service for government decision-making, industry supervision, enterprise operation, travel and so on by the mean of integrated modern information technology, photonics technology, sensing technology, system control technology, software technology, management and decision support technology [7]. Also, urban traffic operation monitoring and service takes achieving the collection, transmission, processing and sharing of different kind of transportation operation information as the way, and target at achieving integrated transport system to run concurrently traffic fine management and precision of service. It provides all kinds of real-time, dynamic, and customized traffic information service for government decision making, industry regulation, enterprise operation, travel [8, 9].

Experience in domestic and foreign large-scale information systems' development shows that the research and development of system frame are all the starting point and the basis for the construction of complex large-scale systems [10, 11]. Urban transport systems is a typical complex systems composed of multiple subsystems, which contains a lot of supervision, control, management functions, as well as a wide range of various types of high tech and a great deal information of time, space, static, and dynamic. The demand for services of the system is also a wide variety and ubiquitous [12, 13]. Therefore, urban traffic monitoring and service framework are used to identify and describe the functions of full urban traffic information system users must have. Framework built a models of urban traffic information service system logically and physically, and determine the modules and interfaces between modules, which contains the design of all the subsystems to implement the function of user service.

77.3 The Goals and Content of Framework Design

To provide support for achieving urban traffic construction goals of integrated transportation, intelligent transportation, green transportation, and safe transportation, there is an urgent need for framework design of urban traffic monitoring and service system research. Integrated urban transport information can be used to guide and standardize the construction and management of urban traffic information service system.

1. Goal is to fully support the construction and development of information system of urban traffic, which provides a systematic overall framework of urban traffic informatization development.
2. In order to solve the development structure of urban traffic information service, components and relationship of construction work of urban traffic information service application system are provided. Clear planning and construction work, and the overall direction are provided for construction work of urban traffic information system, which is based on demand system, functional framework, and system structure.
3. The target architecture model of information systems and functional architecture are proposed, the development blueprint of urban traffic is made clear.

77.4 Frame Design Methods and Examples

1. Demand system design methodology

First, define the user principals of urban traffic information service. This is the premise and foundation of requirements' analysis and definition of customer service.

Second, the user principal needs user analysis to get user needs. The main tasks of requirement analysis is to describe which kind of system functions and system characteristics is needed by users. User needs are the basis for user services and subservice definition; therefore, the user services' definition stages most important work.

The last, according to the results of the analysis of user needs, user service definition is made. User service is to meet the needs of users which is provided by urban traffic information service system, which is a bridge between user needs and function of system, which is the basis of the logical framework and the physical frame.

According to a summary of system design theories methods the design technology course described is as follows, which is shown in Fig. 77.1.

2. Function frame design example

According to the service definition, the top-level function framework is designed. Function requirements analysis start from the perspective of system

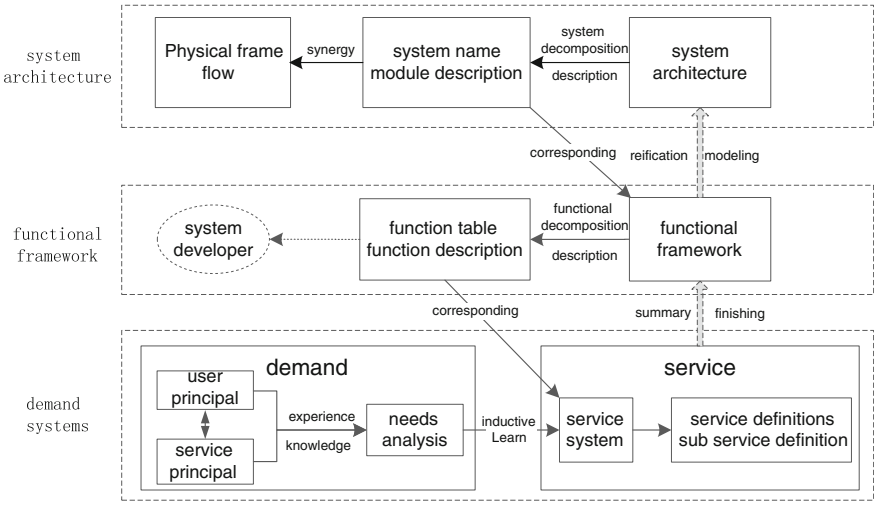


Fig. 77.1 The technology roadmap of framework design

analysis, and get urban traffic information services function. Analysis of user demand for services is the main user, and functional requirements' analysis is for system developers.

Physical frame is a concrete manifestation of the logical framework, which is to convert the information of data flow diagram to the design descriptions of system structure. System architecture is a system of physical, functional entity, modeled in the logical framework, related system functions, and data flows into the system and

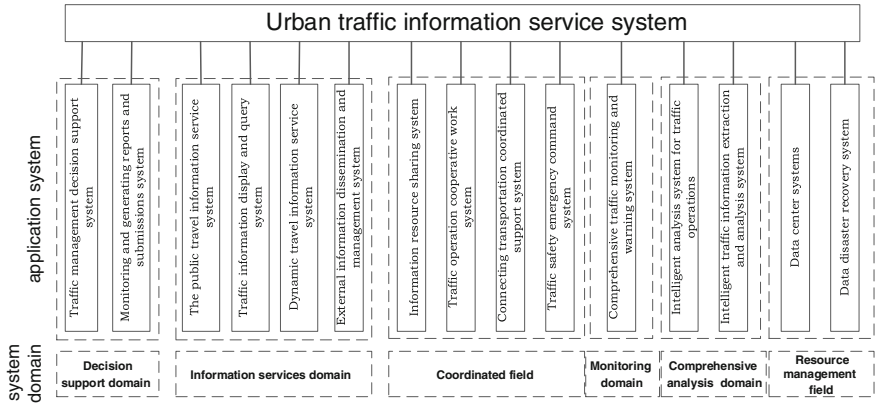


Fig. 77.2 Urban traffic information service system

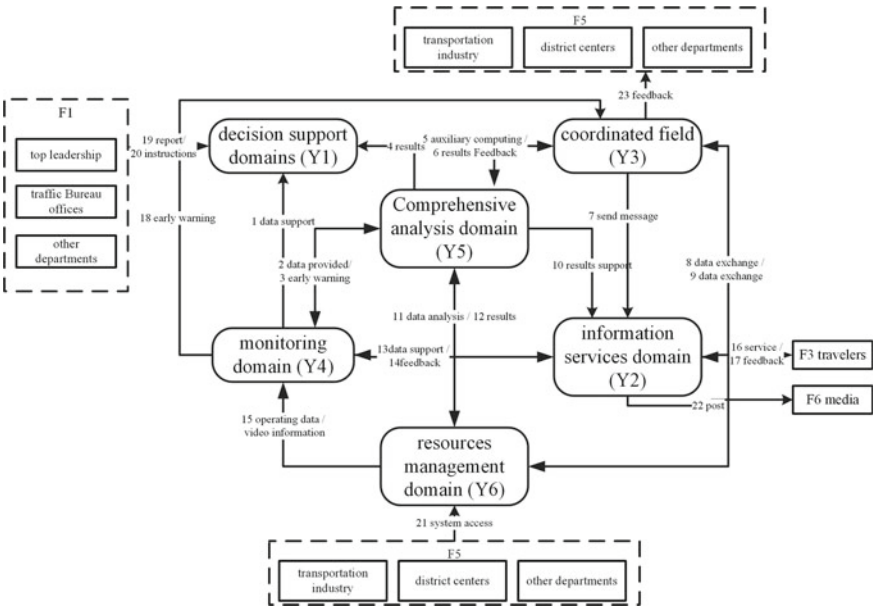


Fig. 77.3 Physical flow of traffic information service

subsystems. Urban traffic information service system architecture is shown in Fig. 77.2.

Based on identifying subsystem of traffic information service system and its functions, service objects and service as the main line, establish cooperative working relationships between the various systems. Relationship between the various systems is shown in Fig. 77.3.

77.5 Summary

The framework make a blueprint for the construction and development of urban transport operational monitoring and service systems and provides a technical framework. The framework is based on the user needs, through the main line of "information resource management, operation analysis, information transmission services," including demand system functional framework and system architecture. Requirements system is faced to the "Government, industry regulation, enterprise operation, travel", the user needs are divided into six main areas, which including decision support, information services and cooperating monitoring, comprehensive analysis, resource management, and so on. Functional framework of urban traffic information service system to determine various user services abstracted their overall function and subfunctions that should be included, and formed the general

functional architecture. Physical schema is a description of the physical implementation of urban traffic information service which will function within the framework of the types of functions defined by integrating modules and data flow that defines urban traffic information service function by physical subsystem, as well as interactive stream of framework and the relationship between these subsystems.

In the application of framework design, the framework can be adjusted according to the city's social and economic characteristic, to improve its applicability. The framework being used in the monitoring and services of urban transport modes in systems development practice, which gets new design requirements and nurture existing system design. The framework progressively apply to monitoring urban traffic and service development, and faced to improving the monitoring of city traffic and the level of service.

References

1. Wu C, Luo Z (2005) Study on urban traffic problem in China. Science and Technology of China Press, Beijing (in Chinese)
2. Niu L (2008) Key techniques research of pure electric bus charging system. PhD, Beijing Jiaotong University, Beijing (in Chinese)
3. Cai W (2003) Strategic conception of the development of intelligent transportation system in China. Transp Eng Inf 3(1):16–22 (in Chinese)
4. The Theme Group on Framework of Intelligent Transportation System in China (2002) Intelligent transportation system frame of China. People's communications press, Beijing (in Chinese)
5. Zhang K, Wang G, Du Y (2007) Study on functional orientation and development pattern of comprehensive traffic information platform. Transp Eng Inf 7(3):30–35 (in Chinese)
6. Guo J, Wen H, Zhang K (2004) Research and demonstration project of Beijing comprehensive traffic information platform construction. Transp Eng Inf 4(3):7–9 (in Chinese)
7. Li X (2008) Public travel information service system of highway construction. Stand 1:34–37 (in Chinese)
8. Zhang K (2005) The overall framework of integrated transportation information platform construction guide. State guidance technical document for standardization (in Chinese)
9. Li M, Lu H, Si Q (2005) Study on the development situation and trend of integrated traffic information platform. Road Traffic Technol 22(4):90–94,143 (in Chinese)
10. Achman JA (2007) A framework for information systems architecture. IBM Syst J 38: 454–1470
11. The Open Group (2009) TOGAF version 9: the open group architecture framework
12. Philippe K (1995) The 4+1 view model of architecture. IEEE Softw 12(6):42–50
13. Such NP (1990) The principles of design. Oxford University Press, New York

Chapter 78
Technical Framework of Cooperative Control for IT Services of Railway

Jicheng Qu and Xiaojun Chen

Abstract According to the current development needs of IT services of railway, the author designed technology framework of cooperative control on IT services of railway based on control objectives for information and related technology and service-level agreement which includes cross-functional architecture commission, service-level agreement management, business value and project tracking and share of IT costs and audit mechanisms. By designing this series of mechanisms, this paper aims to be able to achieve the target classification of different priority target in IT services of railway and reach the purpose of the prevention of risks and opportunistic behavior.

Keywords IT services of railway · Cooperative control · Technical framework

78.1 Introduction

With the improvement of IT service market, the cooperative control of IT service and its application in the industry have quickly become the focus of the industry and academic world [1–6]. But considering the particularity of railway transportation and its derivative services, the relevant research is still at the starting stage [7–9]. General researches think that railway enterprises can reduce their costs and fully meet the requirements of railway passengers by implementing effective collaborative control of IT services. Therefore, how to build a technical framework for cooperative control of railway IT service, how to improve the success rate of the implementation of railway IT service, and how to promote the development of marketable operation of information services in China's railway industry have very important practical significance.

J. Qu (✉) · X. Chen
School of Economic and Management, Beijing Jiaotong University,
No. 3 Shangyuancun, Haidian District, Beijing 100044, People's Republic of China
e-mail: wfbuaa@126.com

X. Chen
e-mail: wangfei54188@163.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_78

78.2 The Construction of Technical Framework for Cooperative Control of Railway IT Service

In view of the inherent technicality of IT services and the complexity involving the interests of many participants, this paper argued that all parties in railway IT service should share the risks. Meanwhile, decision makers of railway enterprises should encourage their IT service teams and sales teams to cooperate with each other, and on this basis, they should also introduce audit mechanisms to improve the technical framework for cooperative control.

According to the structure of technical framework, this paper argued that the governance interactive platform jointly constructed by them will be more conducive to forming a reasonable collaborative governance structure and a control mechanism of railway IT service with perfect information channels, from both the interior of technological structure for cooperative control of railway IT service (mainly including the railway companies and railway IT service suppliers) and the exterior (mainly including government authorities and the railway passengers). But under the collaborative governance structure and control mechanism, all relevant interest bodies play their respective roles and also restrict one another, so that they can create more multiplicative commercial value and social value.

Moreover, the technical framework for cooperative control of railway IT service embodies incentive and constraint multiple orientation, especially setting a senior management team between railway enterprises and railway IT service suppliers to participate in the service control, and in this way, it can alleviate the risk of railway IT service from a wider range and reduce the information asymmetry in the decision-making layer. By the diversified redistribution of rights and obligations of different interest bodies in technical framework, it can effectively supervise the railway IT service and its performance through the relationship of mutual checks and balances. Meanwhile, it can also motivate the initiative of various contract bodies of railway IT service to participating in the governance of railway IT service, so as to improve its governance performance.

It is obvious that the technical framework for cooperative control of railway IT service proceeds from the basic need to meet the demand of interested parties, guided by the rule that interested parties of railway IT service share the interests, it fully mobilizes the initiative of various interested parties, let them become the sharers while creating IT service values, and then it fundamentally controls IT services in railway enterprises and improves the service performance, providing support and protection for railway transportation service as a core business. As the "rational economic man," interested parties may actively involve in the cooperative control of railway IT service and have an incentive to provide the necessary support for the cooperative control of railway IT service only when their interests are met. In addition, decision makers in railway enterprises should consider to introduce COBIT control framework while sharing the interests, that is to control the whole process of railway IT service from investment decision making to the operation and then to the results, and allow the top management teams in railway enterprises to

organize the internal audit directly at railway IT service for the supervision and control. Furthermore, the external sectors (especially government administrative departments of the railway industry) should also employ external auditors to audit the management process of railway IT service irregularly, so as to clear the rights and responsibilities of interested parties of railway IT service (Fig. 78.1).

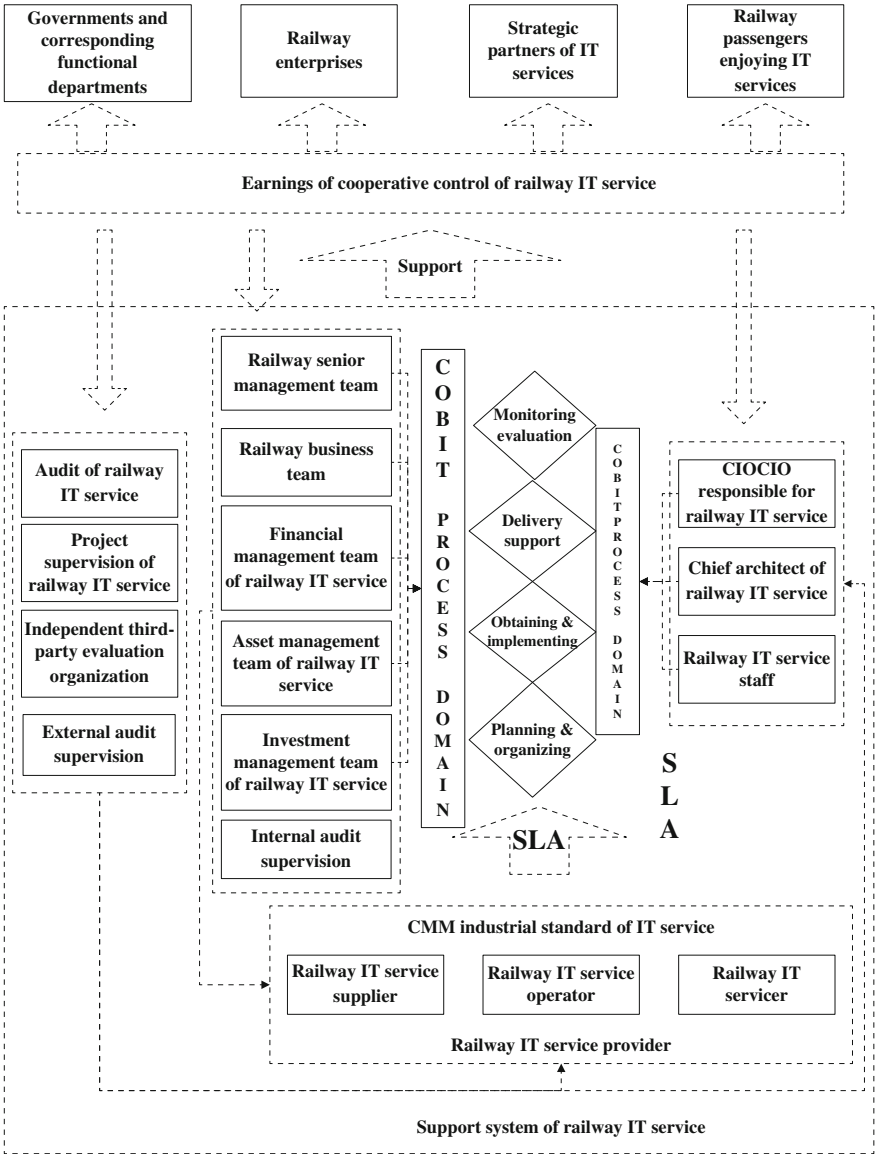


Fig. 78.1 Technical framework for cooperative control of railway IT service

78.3 The Design Content of Technical Framework Based on Cooperative Control

Of course, it is difficult to guarantee that the behavior of interested parties of railway IT service can be consistent with the interests of railway enterprises and passengers only through the interests driving and internal and external audit. Therefore, the thought of management control and incentive control of railway IT service embodied in the technical framework should be used to design the relevant rules and regulations to normalize the stakeholders in the process of railway IT service, so that the interests of stakeholders can be coordinated with the ones of railway enterprises and passengers. Among them, the management control in technical framework is designed according to technical features of railway IT service and from an administrative perspective to control the railway IT service so as to ensure that railway IT service can effectively support the business operation.

While the incentive control is designed by fully using the interests driving stakeholders to mobilize the initiative of stakeholders to make their goals consistent, in order to ensure that all stakeholders of railway IT service can work hard according to the agreed direction, realizing the business objectives of railway enterprises. Below, this paper would elaborate on the design of management control and incentive control in technical framework.

78.3.1 The Content Design of Management Control

In general, railway IT service may involve the hard work of railway IT service provider and the support from non-IT service departments, to let its objective consistent with the strategic objective of railway enterprises, so it is necessary for railway enterprises to establish the corresponding governance mechanisms and supporting measures. This paper thought it includes the following aspects:

1. Cross-functional railway IT service architecture board. Establish railway IT service architecture board with different business departments containing the railway transportation service, with the design of IT service architecture for railway enterprises and the description of plotted course for railway IT service as the main tasks. To be specific, railway IT service architecture board provides the railway IT service with a vision and points out how to integrate IT services into the process of major railway transport services and to realize the cooperation among multiple departments, and on this basis, it re-estimate the value of railway IT service to the operation of railway enterprises.
2. Manage and control it by using Service-Level agreement (SLA). With the universal use of IT Infrastructure Library (ITIL), SLA has also attracted more and more attention from railway IT service. In SLA, by signing higher-layer protocols between railway IT service suppliers and railway-related business

managers, it can show the optional services, optional quality, and related costs. And the biggest characteristic of it is to set the IT departments under the railway IT service suppliers as the external service providers, and in this way, it can make more cost-effectiveness in its internal services, or make some decisions related to railway IT service outsourcing in any case of getting expected outcomes.

3. The track mechanism for railway IT service. In addition to the above methods, to establish and perfect the track mechanism for railway IT service is an essential link in the government process. And its main purpose is to alert the management layer in railway enterprises to pay attention to the potential problems at the early stage of IT service supply and to take corresponding actions to prevent the delay of service or cost overruns. For instance, it can ensure the management in highly standardized process of railway IT service by introducing Capability Maturity Model Integration (CMMI) certification. And all the above-mentioned introduction of management methods and means are to quickly prevent, identify, and deal with the correspondence problems, undoubtedly, which will greatly increase the implementation success rate of railway IT service.
4. Integrate audit accountability mechanism in railway IT service. Railway IT audit is an important mean and also an important tool to check and balance and realize the power, duty, and benefit distribution related to IT management department in railway enterprises. Its main task is to perfectly and effectively inspect and evaluate the planning and development, change management, use and maintenance, system security and other relevant activities and outputs of railway IT service system by professional IT auditors based on objective criterion, so as to find out internal control design flaws and whether the monitor instruction is effectively implemented and other problems in the process of railway IT service, providing warning mechanisms and means of prevention for the operation of railway IT service. To play the full roles of railway IT auditing, railway enterprises should introduce the third-party IT audit on the basis of the internal audit.

In summary, when railway enterprises manage and control the railway IT service, they should make full use of the above several mechanisms to achieve the goal of governance. Among them, cross-functional railway IT service architecture board, as a guider in the management control system, should be the constitutor of IT service programs in railway enterprises as well as an internal auditor for railway IT service. SLA is the joint of mutual cooperation and communication among various interests in the process of management control, and its operating log can also provide the basis for railway IT payment and audit. The track mechanism is the basic protection orderly implemented by railway IT service according to the SLA agreement. IT audit accountability mechanism is an important supervision mechanism, which will help to clear the boundary of power, duty, and benefit distribution in railway IT service, providing a transparent coordination environment for various stakeholders.

The demonstrative control framework for railway IT service proposed in this paper is a technical control process designed mainly for four control domain objectives in railway IT service. But considering the particularity of railway IT service, the author would improve COBIT-based framework, which is to isolate two control flows of different priorities. Among them, the higher-priority control objective in railway IT service should be arranged to be implemented in the first stage, while the lower-priority control objective should be implemented in the second stage (Fig. 78.2).

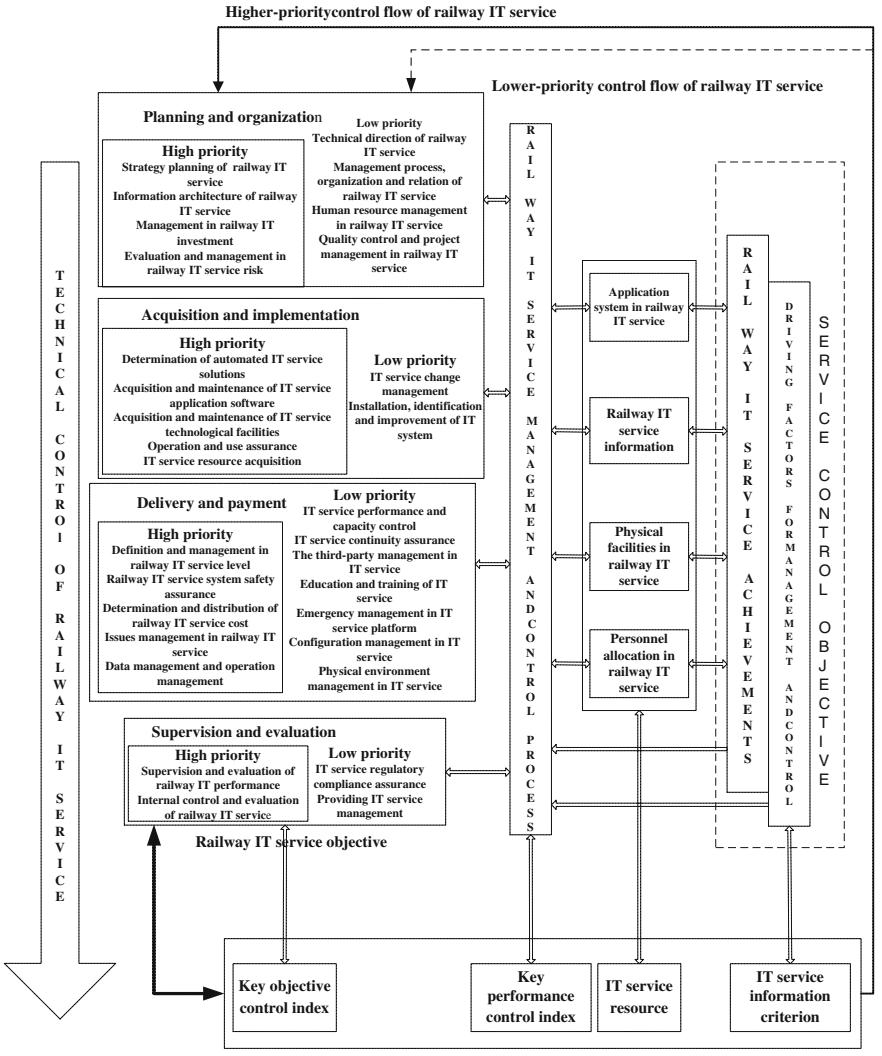


Fig. 78.2 COBIT-based control framework for railway IT service

78.3.2 The Content Design of Incentive Control

It can be seen from the above analysis that to ensure the smooth implementation of cooperative control technology of railway IT service, besides the above mechanisms, we should also adopt the risk allocation mechanism to mobilize the initiative of all stakeholders or contract bodies to cooperative control of railway IT service, thus improving its control performance. In the research, the design of incentive control related to railway IT service can include the following aspects:

1. The cost allocation mechanism. It mainly comes from the internal control methods in accounting system, which is to allocate the railway IT service cost (such as the equipment purchasing and installation charge) to the accounting system of related business departments. Of course, this is mainly because many railway IT service costs are fixed, and changing needs of internal users for railway IT service cause the complication of its cost calculation. And the fundamental purpose of the use of effective cost calculation mechanism is allocation cost, so that the d business departments related to railway IT service cost can feedback the use of shared services, so as to improve the application efficiency of railway IT service. In this paper, the specific process of cost allocation mechanism is shown in Fig. 78.3, among which railway IT service cost accumulation includes the direct cost (such as the railway IT service's equipment cost, software cost, labor cost) and indirect cost (such as site cost, transfer cost, budget costing). As a functional department, the architecture board is responsible for recording the collection of direct and indirect cost objects and handing to finance department for accounting calculation. The most important part of cost allocation in railway IT service is the formulation of the allocation standards. When designing the cost allocation mechanism, the railway IT service architecture board will handle the direct and indirect costs separately. For the

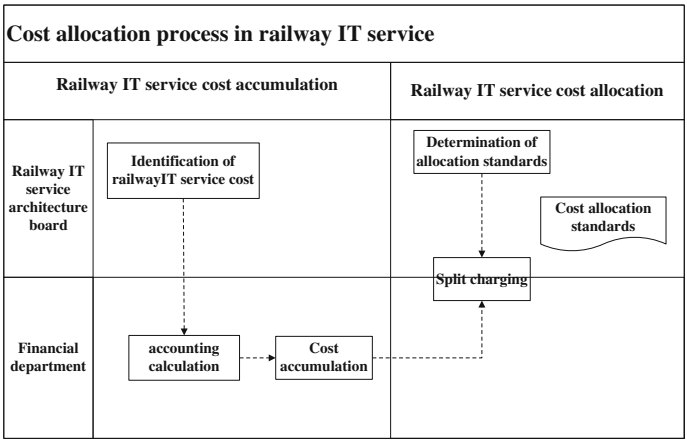


Fig. 78.3 Split charging

direct cost, it has a very clear production subject, so it will directly allocate the cost according to the using condition of subjects of the cost; but because it is hard for the indirect costing to find the appropriate allocation index, it can consider to be credited into office expenses after the discussion in architecture board, which then will be allocated to the relevant objects who bear the fees together with office expenses.

2. The formal tracking of railway IT service value. In the process of cooperative control of railway IT service, a big problem is to estimate the value brought by railway IT service. The formal tracking of railway IT service value can better solve this problem, and this is mainly because the tracking of railway IT service value helps the top competent departments to understand the source of value created by the IT service and make a more accurate assessment on the planned benefits generated by the service. In the process of formal tracking, railway IT service value can generally be divided into tangible value and intangible value. Among them, the tangible value is easier to track and be recognized by all the parties, while the intangible value is more difficult to measure. Based on this, this paper argued that the formal tracking of railway IT service value can be divided into the following two stages according to the classification of value: The first stage is the tracking of tangible value, and the second stage is the tracking of intangible value. The tracking of tangible value can form an organizational structure in charge by the architecture board (responsible for the object recognition of railway IT service value) and participated in and cooperated by finance department and IT functional office (responsible for the accounting of corresponding subject value), so as to track the tangible value. The tracking of intangible value should be established on the basis of the first stage (this is mainly to provide a support for intangible value), and from the perspective of strategic adaptability of benefits, the intangible value is tracked by means of value chain evaluation, cost–benefit analysis, and added value analysis.

In summary, railway enterprises should combine the cost allocation mechanism and service value tracking mechanism to clear the sharing situation and beneficiaries of railway IT service value and also pass the risk to all stakeholders to push them to response to the cooperative control of railway IT service. Among them, the cost allocation in railway IT service should be implemented according to the SLA and other contracts to ensure the fairness. And the use of service value tracking mechanism can fully display the value created by railway IT service to let stakeholders get a clear look at the benefits of IT service and then to actively participate in the cooperative control of railway IT service. In the meanwhile, the value tracking may become a performance standard that railway enterprises motivate railway IT service providers.

78.4 Conclusion

To resolve the risk in the cooperative control of railway IT service, a set of effective cooperative control system of railway IT service should be established. Normalize the operation of railway IT service from the double perspectives of technology and institution, so as to reduce the opportunistic behavior in it caused by the information asymmetry. On this basis, in this paper, a technical framework for cooperative control of railway IT service was designed to alleviate the above problems from the double perspectives of management control and technical control. Specific mechanisms containing in the technical framework for cooperative control and their roles are as follows: (1) It is overall responsible for risk issues in the process of cooperative control of railway IT service by constructing a cross-functional architecture board; (2) it prevents and restrains the opportunistic behavior in cooperative control of railway IT service by using SLA-based contract; (3) it clearly classifies the objectives of different priorities in cooperative control of railway IT service with introduced COBIT-based framework; and (4) it implements a comprehensive IT service value tracking mechanism to motivate all interested parties in cooperative control.

References

1. Finkle A (2005) Relying on information acquired by a principal. Int J Ind Organ 23(3–4):263–278
2. Debreceny RS (2006) Re-engineering IT internal controls: applying capability maturity models to the evaluation of IT controls. In: Proceedings of the 39th annual Hawaii international conference on system sciences
3. Gammelg M, Simonsson MR, Lindstr Å (2007) An IT management assessment framework: evaluating enterprise architecture scenarios. IseB 5(4):415–435
4. Canada J, Sutton Q, Kuhn J (2009) The pervasive nature of IT controls: an examination of material weaknesses in IT controls and audit fees. Int J Account Inf Syst 17(1):106–119
5. Chan L, Jee HL, Qian W (2007) Internal and external influences on IT control governance. Int J Account Inf Syst 8:225–239
6. Dale Stoel M, Muhaxina Waleed A (2011) IT internal control weaknesses and firm performance: an organizational liability lens. Int J Account Inf Syst 12(07):280–304
7. USDOT (2010) ITS strategic research plan 2010–2014
8. Paulsson B, Pointner P, Jaiswal J et al (2009) An overview of wheel–rail interface related research in the European project INNOTRACK including issues in technical and economical validation
9. Ning B, Tang T, Gao Z et al (2006) Intelligent railway systems in China. Intell Syst IEEiE 21 (5):80–83

Chapter 79
Social Network-Based Onboard Passenger Entertainment and Information System for High-Speed Trains

Weifeng Yang, Zheng Wen and Aijun Su

Abstract Providing fruitful passenger entertainment and information service on high-speed trains through smart devices is challenging and promising. In this paper, we firstly review related systems providing the service: the third-generation communication system for railway passengers, the onboard Wi-Fi system, and the in-flight entertainment system. Based on the insight obtained, we propose the system architecture of our social network-based passenger entertainment and information system for high-speed trains, which builds up a social network among onboard passengers and provides dynamic applications to improve passenger travel experience. Additionally, we introduce the function breakdown structure of the system. We finally present the developed system that is ready to use.

Keywords Passenger service · Smart device · Onboard Wi-Fi · High-speed train

79.1 Introduction

The fast growth of the Chinese high-speed train market has greatly improved passengers' travel experience in saving the travel time and offering comfort accommodation. However, the onboard passenger entertainment and information service is far from satisfied. Passengers have few means to spend their spare time during the long journey such as reading printed materials, i.e., magazine and newspaper, listening to the passenger radio, and playing off-line games on their own electronic devices. Unfortunately, they are not interconnected, nor they are connected to the Internet. This situation could have been fundamentally changed by providing a social network-based passenger entertainment and information service in the mobile Internet era. And providing social network-based passenger entertainment and information service also conforms to the trend of onboard multimedia subsystems for railways (OMMSR), as standardized in IEC62580 [1].

W. Yang · Z. Wen (✉) · A. Su
CSR Zhuzhou Institute, Shidai Rd, Zhuzhou 412001, Hunan, China
e-mail: wenzheng@csrzic.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_79

In this paper, we propose our passenger entertainment and information system (PEIS) for high-speed trains which builds up a social network among onboard passengers and provides dynamic services to improve passenger travel experience. In contrast to the traditional passenger information system (PIS) that pushes passenger-related information through LCD screen and radio, PEIS focuses more on the two-way interactive entertainment and information service via smart devices. Our PEIS interconnects all passenger smart devices, i.e., smartphone and tablet, by setting up a wireless local area network on the train. This greatly reduces the deployment complexity and installment cost of the system.

PEIS consists of three parts: the mobile apps installed on passengers' smart devices, the onboard PEIS platform, and the ground back-end system running at the ground data center. Specifically, the PEIS app is a system interface for passengers and it supports various passenger-oriented services, i.e., onboard social network, onboard catering, and video surveillance. The PEIS platform, residing on the onboard servers, facilitates the onboard data communication, passenger information, and entertainment services. The ground back-end system fulfills the general requirement of system management, interfleet data service integration as well as service extension onto the Internet. With the support of the back-end system on the ground, PEIS serves as a single yet large cyber community for all passengers/subscribers; the onboard PEIS services can be transparently extended to all PEIS-available high-speed trains as well as on the Internet.

In the rest part of this paper, we firstly review related systems that give the insights for building the PEIS: the third-generation communication system for railway passengers, the onboard Wi-Fi system, and the in-flight entertainment system. We then present our PEIS in detail, including its system structure, function specification, service instances, and its implementation.

79.2 Related Works

Providing information and entertainment service to railway passengers has long been a hot topic in both industry and academia. Given the sharp increase of passengers' Internet surfing demand during the long-time duration of the high-speed train journey, there has been a lot of discussions and practices in real life to enable the onboard Wi-Fi as part of the passenger information and entertainment service.

The third-generation (3G) communication technology has been adopted at the first place [2]. Unfortunately, 3G communication technology, if not optimized, suffers from the Doppler effect, the signal loss, and the fast handover problem when implemented on the rail sector [3]. As a result, passengers' end devices can hardly connect to the 3G networks outside the train directly. The quality of service is compensated. Passengers may suffer from long buffering time and frequent link break.

An alternative is to provide the onboard Internet surfing through wireless LAN technology, mounting multiple access points along the train to provide local

wireless access for passengers' end devices, while relaying on optimized 3G network, private network, or even satellite connection to provide Internet access services [4, 5]. This series of technical work is transparent to passengers. They can seamlessly access the network following the normal procedures. Obey with this general strategy, are many industry start-ups in the onboard Wi-Fi field in Europe, i.e., Nokia Semens Networks, Orange Business Services, and Corenet Oy [6]. Onboard Wi-Fi service has now been provided in many countries, including the Great Britain, France, Japan, and Hong Kong [6]. Although many efforts have been enforced to improve the onboard Wi-Fi service, the throughput of the network is still low due to the onboard-to-ground bandwidth limitation. In many situations, the throughput only meets the instant messaging and e-mail requirements, surfing the Web site would suffer from long buffering time, and watching online video and downloading bulk data are generally prohibited. In addition, the onboard Wi-Fi is now primarily free of charge [7]. The installing investment and operating cost can only be compensated by the increasing number of passengers. As a result, the onboard Wi-Fi lacks commercial attractiveness and is difficult to motivate the train operators to invest.

In the civil aviation field, in-flight entertainment system (IEF) has been available for years [8]. IEF systems typically include LCD screen and entertainment control system (platform). An IEF system offers many services including music, video-on-demand streaming, game, flight map, and duty-free shopping. Compared to the PIS, IEF system has a certain degree of interactivity, but the service it provides is relatively simple with no mobile Internet concept being integrated yet. This is likely due to the fact that cell phone shall not be used on the plane for safety reasons and the seat-mounted LCD screen has every limited resource for human–machine interaction. Two other demerits of the IEF system are as follows: that the IEF service cannot be extended after alighting the fleet (e.g., it is not possible to watch purchased movie after the journey) and that the IEF system generally provides Internet access via satellite communication which makes it every expensive, if not unaffordable, for passengers.

79.3 The PEIS

Our PEIS not only offers Internet access to the passengers but also builds a social network among passengers no matter they are onboard or alighted. The PEIS emphasizes more on the passenger experience through interactive communication using smart devices, and this largely reduces the installment cost of the system.

The insight of building the social network-based PEIS is that passengers on the same train, which is approximately several hundreds, share common characteristics, i.e., heading to the same destination, caring about the train schedule, and interested in news and advertisements at the destination. And of course, many of them are also interested in surfing the Internet. Moreover, many of them could have no choice but sit in the cabin for several hours before arrival. By building up a social network

onboard, it is very easy to attract passengers to engage. An additional advantage of doing that is to reduce the total bandwidth requirement for the train-to-ground communication, as many of the data service requirements from the passengers can be fulfilled within the onboard network.

In the rest of this section, the system structure of the PEIS is presented first. Its function breakdown as well as its various services is introduced subsequently. We finally introduce the PEIS implementation.

79.3.1 The System Breakdown

The PEIS is composed of the end devices, onboard network, and the ground data center, as depicted in Fig. 79.1.

The end device, i.e., smartphone, tablet, or even mounted touchable screen, is the system terminal for human–machine interaction. The PEIS service is delivered through apps installed on the end devices.

The onboard network adopts Wi-Fi to construct a local area network on the train. Each cabin is equipped with several access points to serve the end devices. Considering the frequent coupling and decoupling requirement of the train, the

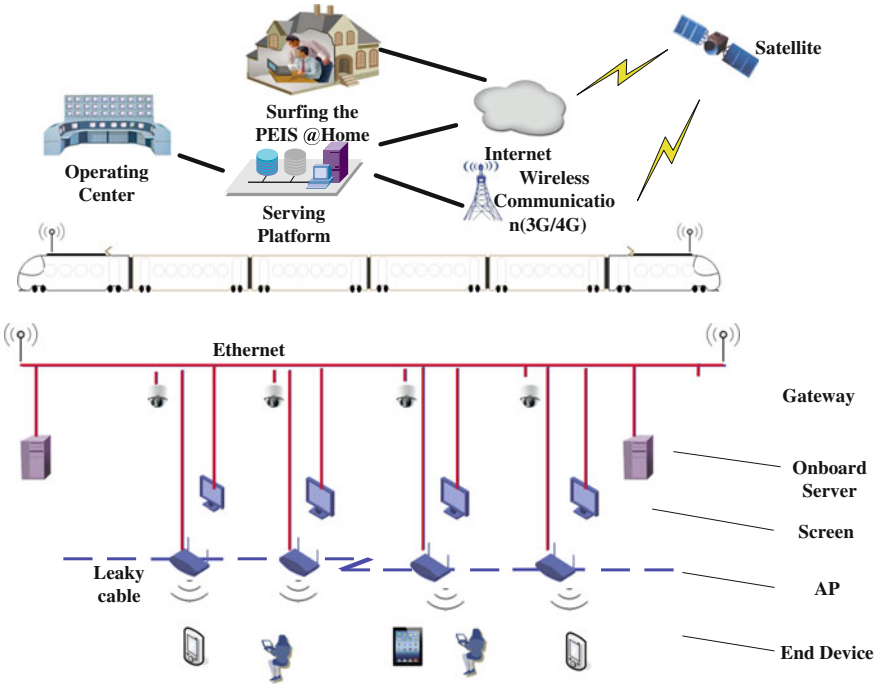


Fig. 79.1 System structure of the PEIS

construction of the onboard network is flexible in the sense that the connection between the access points could be either wired through Ethernet or wireless through bridging. The onboard network also connects onboard servers that facilitate the PEIS service for all passengers and a gateway that is in charge of train-to-ground communication. In order to ensure sufficient outbound bandwidth of the onboard system, the gateway integrates several communication interfaces, including the 3G network, the 4G network, and the satellite communication. The switch among these interfaces is made automatic and transparent to the end devices. It is also feasible to connect other facilities to the onboard network, i.e., cameras for video surveillance and speakers for passengers.

The ground data center serves as the back-end system to support the system management, interfleet service integration as well as the extension of the PEIS service onto the Internet. With the ground data center, PEIS serves as a single yet large community for all passengers/subscribers; the onboard PEIS services can be transparently extended to all PEIS-available high-speed trains as well as the Internet. That is, with the same account, passengers can enjoy their PEIS service whenever they are connected to the system, either through the WLAN onboard or through the Internet after alighting.

79.3.2 The Function Breakdown

The PEIS comprises several function modules to realize the system functionalities, as shown in Fig. 79.2.

For the onboard PEIS platform, the corresponding module and its functionality are listed below:

- Advertisement management module is in charge of the advertisement managing and pushing;
- Content management module manages the storage, dispatch, and copy right issues of all the media content, especially the video, music, and e-books;
- Media streaming module streams the media content to the passengers' end devices; in addition, it also handles the load balance strategy for asynchronous media streaming;
- System security module enforces authentication, encryption, and malicious attack detection measures and enhances the system reliability;
- Account management module handles the onboard PEIS user information and periodically synchronizes it with the ground back-end system;
- Social network service module is in charge of handling the two-way interaction between passengers' end devices and the PEIS. More importantly, it implements all the functionalities of the social network services (to be introduced in the subsequent section);
- Wireless access management module controls the onboard Wi-Fi network as well as its connection to the Internet;

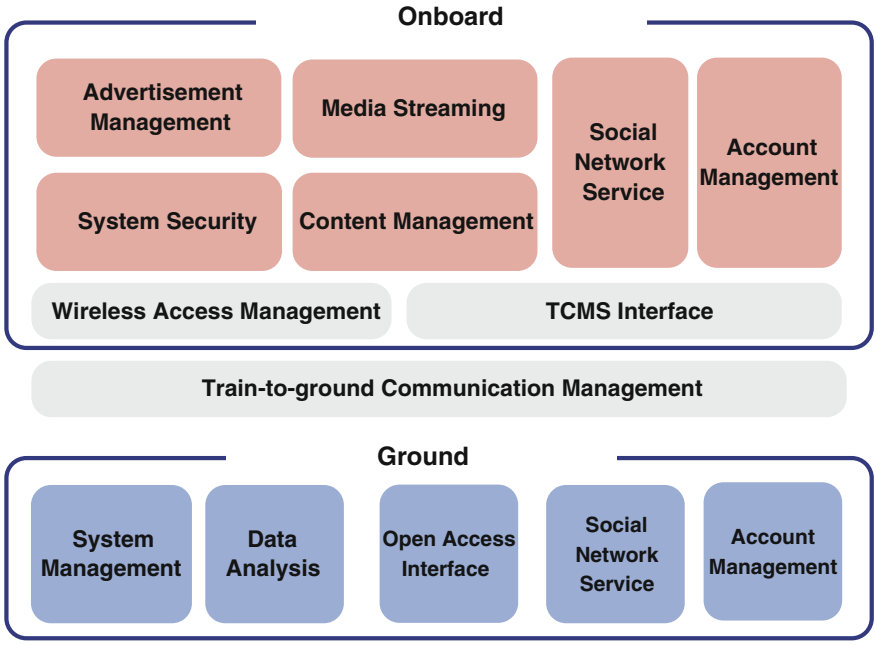


Fig. 79.2 Function module of the PEIS

- TCMS interface module retrieves train operating information, i.e., speed, location, and schedule, for the PEIS services from the train control and management system.

Regarding the ground back-end system, corresponding function modules are mainly for the system management and service extension purpose, including:

- System management module that handles system installation, deployment, daily operation, and management;
- Account management module that handles the user subscription and authentication issue for all PEIS subscribers;
- Data analysis module that analyzes system performance and studies the passenger behavior information in order to improve the passenger experience through customer portrait.
- Social network service module which is the extension of the PEIS service on the ground. It handles service requests from the Internet.
- Open access Interface module that grants the interfaces of the PEIS for third-party service and application development.

One more function module relays in between the onboard PEIS platform and the ground back-end system is the train-to-ground communication management module. It hands the two-way data exchange and manages the resource allocation and priority issue.

79.3.3 The PEIS Service

The PEIS integrates various passenger entertainment and information services, including e-book, music, video streaming, game, map, video surveillance, location-based advertising, catering, and online chatting. These services can be broadly categorized into two parts: traditional ones, as those provided in the IEF system, and advanced ones, as those derived from the mobile Internet social network.

For traditional services, shown in blue in Fig. 79.3, the PEIS succeeds by extending the service from the onboard network to the Internet. As a result, the service is no longer bounded within the cabin but available wherever and whenever the passenger intends to have it, assuming that Internet access is available. For instance, the purchased video playback can be resumed after alighting through either smart devices or computers at home.

For advanced services, shown in green in Fig. 79.3, the PEIS forms a social community by exploiting passengers' common characteristics and demands and brings the social network services that everyone used in the mobile Internet to the onboard community. The train operating information, such as speed, schedule, map, and location, is deeply incorporated in the social community services. Interactive and social collaboration features are highly emphasized in the service design. Instances of the advanced service include supporting online games within the onboard WLAN, location-based/destination-based advertisement, onboard catering, chatting among passengers, drift bottle messaging, etc.

Some novel PEIS services are briefly introduced below:

Fig. 79.3 PEIS service

- Onboard catering: Passengers can retrieve the menu of the onboard restaurant and place the order through their smart devices. On receiving the order, the attendee will deliver the dishes directly, which saves the effort of pushing a trolley along the crowd aisle for hawking and also avoids the long waiting queue outside the catering coach.
- Dynamic wake-up alarm: Passengers can set up a dynamic wake-up alarm based on the live train schedule through their smart devices, i.e., 10 min/2 miles before arrival. This service is extremely useful for passengers on the overnight train, as these trains may run behind the schedule and the fixed train schedule may not be appropriate to guide the alarm setting.
- In-cabin video surveillance: Passengers can retrieve the streaming video from the surveillance cameras mounted within the cabin, i.e., facing toward the luggage storing place or right on top of the door. This service helps to alleviate the passengers' property safety concerns.
- Spotting old friends: Passengers may be able to spot their old friends on the same train but different coaches through contact information or social network ID mapping. This is based on the fact that friends tend to have each other's contact information, i.e., cell phone numbers. By mapping onboard passengers' phone book, it is easy to identify pairs of passengers who are likely to be friends (i.e., they have each other's phone number), though they may not even be aware of the encounters yet.
- Location-based advertising: Passengers may receive dynamic advertisement related to their destination or the intermediate stations the train is currently passing through. Advertisements such as accommodation, souvenir discount, and scenic area guide map may be of great interest to the passengers.

79.3.4 The PEIS Implementation

The PEIS is implemented to prove the design philosophy. We construct the onboard Wi-Fi network and mount two access points per coach. Considering the modernization and installation complexity, all onboard network components are connected using power line Ethernet. A multiband antenna installed at the roof of a train communicates with track-side base stations supporting different technologies, i.e., WCDMA and TD-SCDMA.

The onboard PEIS platform is developed using spring framework, and MySQL is adopted for database usage. Hibernate connects the two by mapping from Java classes to database tables and providing data query and retrieval facilities. The back-end system is deployed on public cloud platform. We have implemented the system management, account management, and data analysis module at the first stage. The app is also developed supporting Android 4.0 and above. Figure 79.4 demonstrates some snapshots of the app and the back-end system interface.

Fig. 79.4 Management system and APP interface

79.4 Conclusions

Providing fruitful passenger entertainment and information services on the high-speed train is challenging. In this paper, we propose our PEIS that is able to provide continuous Internet connectivity for onboard passengers and, more importantly, to form an onboard social network that attracts passengers to engage and experience novel services that are closely tied to their journey. The PEIS service is made available both onboard and through the Internet. It is believed that the PEIS will greatly improve passenger experience on the high-speed train.

Acknowledgment This work was supported in part by the China National Science and Technology Support Program under the grant (2015BAG14B00).

References

1. IEC 62580-1:2015 Electronic railway equipment—on-board multimedia and telematic subsystems for railways—part 1: General architecture. International Electrotechnical Commission standard. https://webstore.iec.ch/publication/21817&preview=1
2. Huawei to Build China's first coverage of WCDMA High Speed Rail Project, available via web. http://www.hartenisi.com.cn/en/pronews.asp?ids=201
3. Rivas Tocado FJ, Diaz Zayas A, Merino Gomez P (2013) Performance study of Internet traffic on high speed railways. In: IEEE 14th international symposium and workshops on world of wireless, mobile and multimedia networks, Madrid, pp 1–9
4. Alasali M, Beckman C (2013) LTE MIMO performance measurements on trains. In: 7th European conference on antennas and propagation, Gothenburg, pp 2327–2330

5. Hsin-Ta C et al (2012) WiFi multicast streaming using AL-FEC inside the trains of high-speed rails. In: 2012 IEEE international symposium on broadband multimedia systems and broadcasting, Uxbridge, pp 1–6
6. The European Market for Train-Based Broadband Wireless Services 2012. Available via web. http://www.bwcs.com/documents/report_broadband_wireless_services_2012.pdf
7. Free wifi. http://www.viarail.ca/en/travel-info/onboard-train/wireless-internet
8. Hao L (2007) In-flight entertainment system: state of the art and research directions. In: Second international workshop on semantic media adaptation and personalization, London, pp 241–244

Chapter 80
Research on Rail Track Maintenance Management Information Systems

Tingting Jiang, Xuehui An and Changyun Liu

Abstract With the rapid development of information technology (IT), information construction has become one of the main works in different industries. Varieties of IT tend to be integrated gradually. Maintenance management information systems (MMIS) application challenges for subway operation companies still exist; however, it creates a need for improved evaluation methods and intensive research. This study proposes a model for evaluating MMIS at the maintenance management process level. Considering the characteristics of maintenance management processes, discrete event simulation coupled with keystroke-level model is adopted in the model. MMIS exerts different levels of effects on different maintenance management processes. The case study verifies that the effects of MMIS on maintenance performance can be assessed quantitatively and objectively through the application of the model. Decision makers may apply the model in MMIS planning phase or in optimizing the available MMIS.

Keywords Maintenance · Information systems · Evaluation · Discrete event simulation

80.1 Introduction

Information technology (IT) applications have assisted architectural/engineering/construction (A/E/C) organizations in improving their service quality and productivity. Investment in IT has become a primary expenditure of numerous A/E/C organizations [1]. Irani et al. [2] also observed a significant increase in information systems (IS) investment to support organization survival in increasingly competitive business environments.

T. Jiang (✉) · X. An · C. Liu
School of Civil Engineering, Tsinghua University, Beijing 100084, China
e-mail: 545405025@qq.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_80

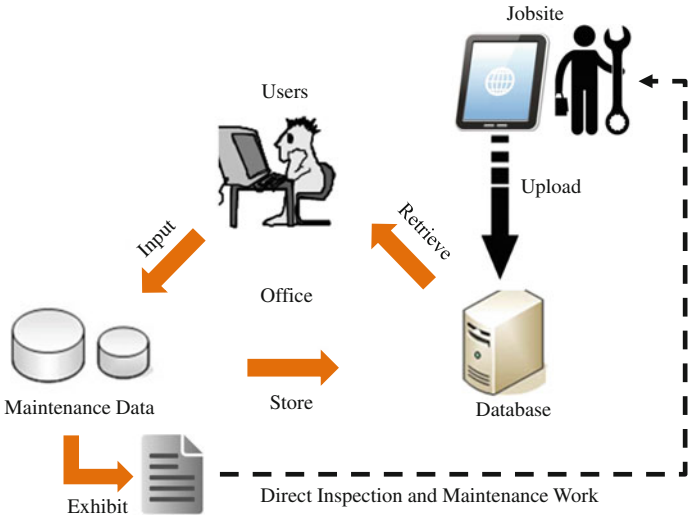


Fig. 80.1 Information management mode of maintenance data

Continuous and long-term maintenance work is necessary to guarantee the safety and usability of railway throughout its entire service life. The fragmentation of data [3], massive scale of railway features with long-term service life, and intensive dispersed maintenance needs intensify the challenges of traditional maintenance management methods, including extensive management, high labor intensity, low efficiency, and information isolation. Railway operation companies are increasingly realizing the benefits and urgency of IT applications as an integral fine management tool. In the past decade, information and communication technology (ICT) [4] has been replacing paper-based communication in various industries with the rapid development of the telecommunication network and the Internet. The integration of mobile devices, personal computers, and enterprise office software enables users to access, retrieve, transmit, and manipulate data and information conveniently and efficiently, as shown in Fig. 80.1. Xuehui [5] proposed the single pocket principle for realizing further communication efficiency improvements. Following this principle, all data generated during the life cycle of a project can be stored in the same database in the Internet cloud and accessed by anyone at any time.

The development and integration of ICT have evolved in recent decades, and its application in the A/E/C industry has significantly progressed. However, the application of IT in companies and organizations continues to confront challenges. Previous researchers have discovered that most maintenance management systems fail to provide the expected benefits [6, 7]. An investigation on three Chinese subway operation companies revealed that IS did not achieve adequate performance expectations. Moreover, work efficiency was not improved, information isolation was not solved effectively, and workloads actually increased.

As such, IT application processes present a dilemma to decision makers within their organizational structure. Continuous IT innovations may eventually contribute to organizational performance and may maintain competitive capabilities. However, increasing IT investment introduces the added task of evaluating its overall business value. Quantifying the effects of IT on organizational performance may resolve the abovementioned issues [8]. Based on the investigation of maintenance management information systems (MMIS) of ten Japanese metro companies and five Chinese metro companies, this paper summarized the characteristics and features of current MMIS. An evaluation model was proposed for objectively and comprehensively quantifying the effects of MMIS on maintenance management processes. Combining the investigation and simulation results, suggestions and solutions were proposed for future information construction of railway operation companies.

80.2 Investigation on Development and Application of MMIS

The first MMIS appeared in the early 1980s, which were mainly administratively oriented. Progressively, more maintenance software containing specific modules (i.e., equipment data, work order planning) became available. In nowadays, maintenance software is commercially available either as an independent package or as a module in integrated systems [9]. With the development of graphic user interface (GUI) and experience economy, user experience (UE) [10] has been an essential aspect in IS design and evaluation in the last ten years for improving the ease of using IS and satisfying user requirements. Garrett [11] suggested that UE must be considered in the entire process of designing and developing IS. The increasing attention given to UE indicates that the design and development of software have been transformed from being production-centered to user-centered.

80.2.1 Application of MMIS in Japan

Based on the investigation of ten Japanese metro companies, the feature of MMIS applied in Japan was summarized as shown in Fig. 80.2. The inspection module and equipment module are two basic modules, which provide equipment data and all kinds of inspection results (i.e., patrol inspection, monthly inspection, and dynamic inspection) for analysis and budget work. Maintenance plan is made afterward, which would further direct execution of maintenance work and personnel schedule. On-site maintenance work and materials consumption are recorded and would cause information update of the equipment module. The equipment information and status data can be depicted as diagram to realize visualized management. The

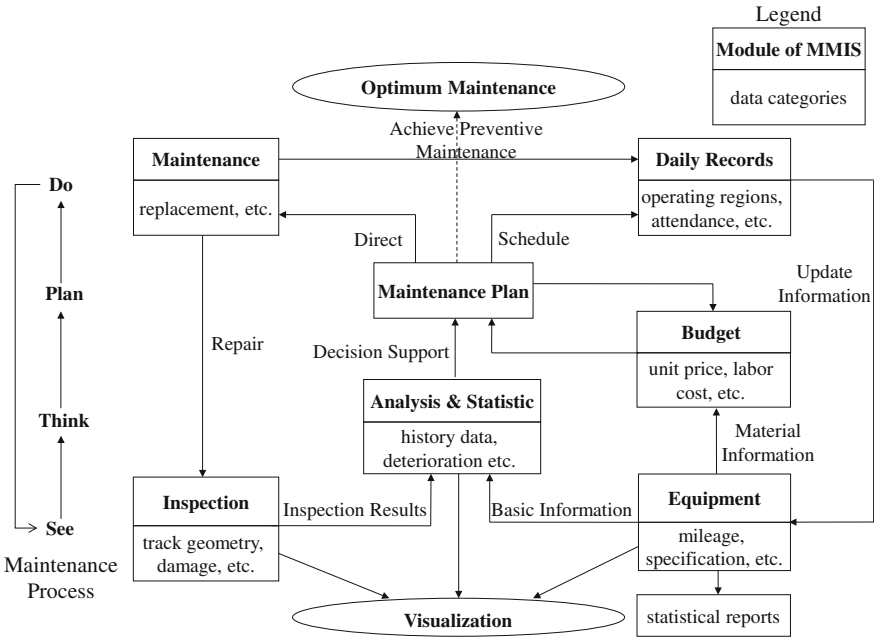


Fig. 80.2 Modules of MMIS applied in Japanese metro companies

ultimate objective is to transform from breakdown maintenance (BM) to preventive maintenance (PM).

80.2.2 Application of MMIS in China

Considering the investigation in several Chinese subway operation companies and literature review, MMIS applied in China can be classified into two categories. One is enterprise asset management (EAM) system, which threads through the generation, submission, approval, and execution of work order. Through the establishment of enterprise database, equipment management, maintenance management, purchasing management, inventory management, and other modules are integrated in a data sharing information system, so as to realize the assets' life cycle management. However, the asset management of EAM in subway companies is wide range but not refined, which fell short of performance expectations.

The other type of MMIS focuses on the management (i.e., input, retrieve, storage, and display) of various maintenance data. This kind of MMIS has already been developed and used in several first-tier cities, such as Beijing [12], Shanghai, and Shenzhen.

80.3 Evaluation Model of MMIS

This study proposes a model for evaluating MMIS at the maintenance management process level. The two categories of MMIS were comparatively assessed using the model in the following case study.

Figure 80.3 presents the proposed evaluation model for objectively quantifying the effects of IS on maintenance management processes. First of all, workflow diagram was developed based on the interview and investigation involving different maintenance workers. Considering the characteristics of maintenance work, the maintenance management processes were simulated using discrete event simulation (DES) [13]. The workflow diagram was composed of two categories of work units: the work units accomplished by applying MMIS (Category 1) and the work units accomplished without using MMIS (Category 2). The work time of Categories 1 and 2 was obtained by incorporating keystroke-level model (KLM) [14] and by investigating the actual maintenance work, respectively. Users were assumed to have the same degree of familiarity with the evaluated MMIS because the time of the operational units in KLM (as shown in Table 80.1) represented the level of general operators. The key performance indicators were selected according to the objective of applying IS.

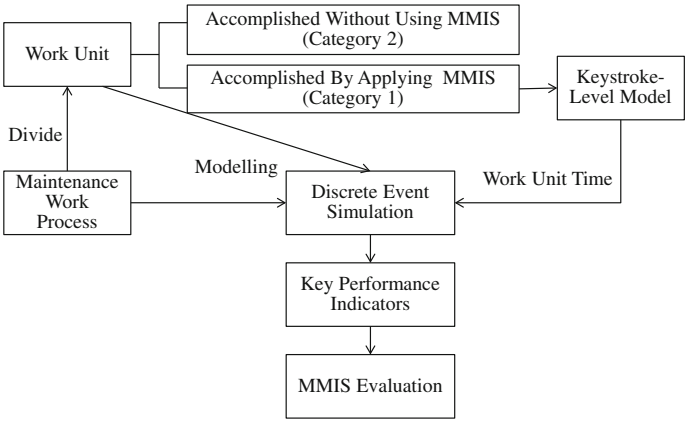


Fig. 80.3 Evaluation model of MMIS

Table 80.1 Average time of each operational unit

Operational units	Time (s)
Keying, *K*/*B*	0.2
Pointing, *P*	1.1
Homing, *H*	0.4
Mentally preparing, *M*	1.35

80.4 Case Study

80.4.1 Introduction of the Subway Company

The focus of this case study is a metropolitan subway company which was founded in 2004 and which implemented initial IT applications in 2009. The maintenance issues are extensively integral to the operations responsible for upholding the safety of the subway systems. The subway lines are closed between 11:00 p.m. and 6:30 a.m. every night to allow the workers to conduct inspections and repair work. Over time, the data and documents on this work accumulate in the maintenance department and are geographically distributed and saved. Meeting the challenges of retaining the data management accuracy is imperative to satisfying the safety requirements; however, massive daily data records have been consistently time-consuming and cumbersome for effective management. The company sought methods of improving data management with the IT application process and introduced a customized IS in 2013. In this study, a quantitative comparison was conducted, the results of which were evaluated to analyze the effect of IS on maintenance performance in the maintenance management process during four phases, as shown in Table 80.2.

80.4.2 System Boundaries

Establishing the model scope is one of the important issues for DES studies in the context of the construction industry [15]. By considering the scope of MMIS application in the company, this study set the maintenance management process boundaries from the daily railway equipment maintenance production of crew workers to the technical management work of professional engineers, directors, and deputy manager during an entire month (Fig. 80.4). The crew workers obtained data from the job site and provided them to engineers for technical analysis. Through technical management work, the data were transformed from being effective information into direct daily production work.

Table 80.2 Introduction of IT applications in the subway company

Cases	Time interval	IT applications
Case 1	2005–2008	Office, corporation email
Case 2	2009–2012	EAM
Case 3	2013–2014	MMIS with several modules (Web)
Case 4	2014–2015	MMIS with several modules (Web and Pad)

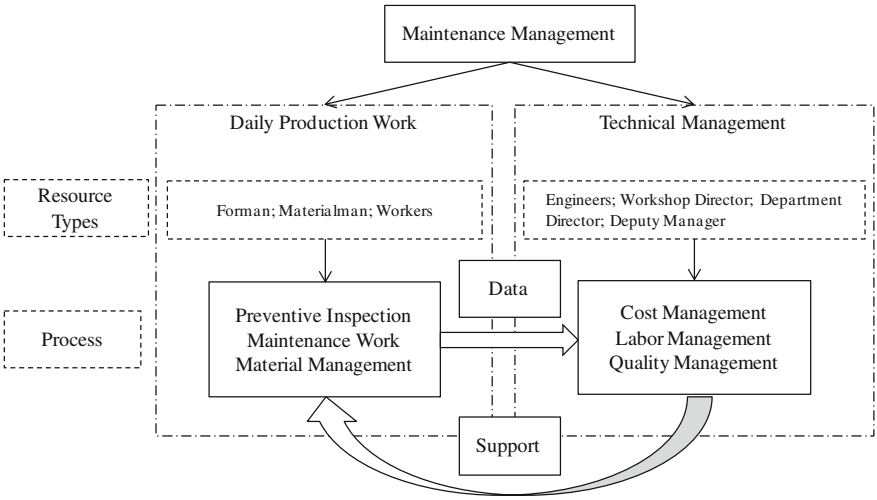


Fig. 80.4 Boundary of the maintenance management processes

80.4.3 Building the Maintenance Process Diagram

As shown in Fig. 80.5, maintenance management can be decomposed into six processes: material management, preventive maintenance inspection, maintenance work, labor management, quality management, and cost management. Material management refers to storeroom arrangement, supplies purchasing plan, and tools for repair work. Preventive maintenance inspection consists of patrol inspection (once every two days), track geometry inspection (once a month), dynamic inspection (once every quarter), and so on. Subway maintenance involves hundreds of repair work categories because of different equipment failures. In this study, maintenance work was classified into eight categories according to the crew size. Technical management was divided into quality, cost, and labor management. Quality management is mainly implemented by technical engineers and management staff, and it refers to the statistics and summary work of equipment failures and maintenance plans for assuring traffic safety. Cost management reduces inventory accumulation and controls the material budget. Labor management applies to monthly staff scheduling and attendance statistic work.

80.4.4 Model Validation

The simulation model displays the daily work of the crew workers over an entire month. The simulated results were compared with the actual work time obtained

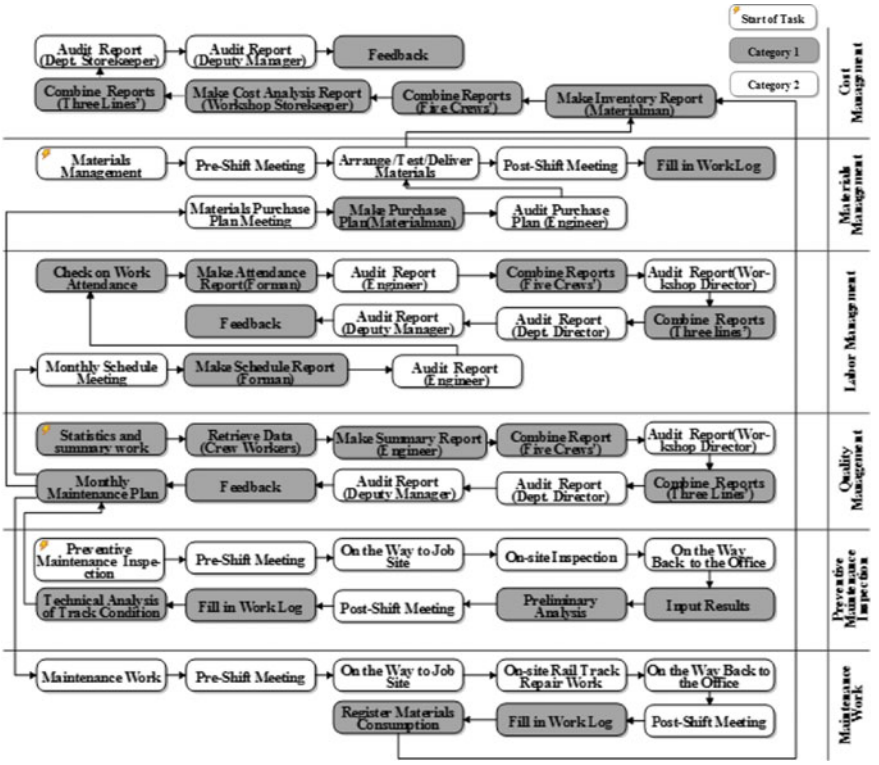


Fig. 80.5 Workflow of maintenance management processes of the subway company

Table 80.3 Work time of an entire month

Resource type (count)	Total work hours		Relative differences (%)
	Actual time	Simulated results	
Foreman (1)	160	156.27	2.33
Materialman (1)	184	182.40	0.87
Workers (10)	1672	1619.88	3.12
Total	2016	1958.55	2.85

through investigation. The findings are displayed in Table 80.3; they specify that the simulated results of different resource types are extremely close to the actual work time. A few crew work items (e.g., training), which were not considered in the model, account for the 2.85 % of the variance between the total actual work time and simulated results.

80.4.5 Results and Discussion

Total work time of Category 1 work units during an entire month was calculated based on the simulation results as displayed in Table 80.4. Comparing to Case 1, the decrease ratios of Cases 2, 3, and 4 were calculated, respectively. EAM applied in Case 2 exerts greatest effect on material management. The MMIS applied in Case 3 caused the improvement of work efficiency of technical management (cost, manpower, and quality control), each of which is over 90 %. The MMIS was customized according to the maintenance workers' requirements. Once the data were inputted to the MMIS, the engineers and workers could obtain the information conveniently rather than retrieving data from an overwhelming volume of documents. Furthermore, various analysis reports can be generated automatically. The introduction of portable terminal system (Pad) in Case 4 can further save the manpower of data input.

The total cycle time of the six processes is displayed in Fig. 80.6. In summary, the application of MMIS in Case 3 and Case 4 resulted in 53 % savings in technical management cycle time. Besides, the introduction of Pad in Case 4 resulted in almost 20 % savings in production work as compared to Case 1.

Table 80.4 Total work time of Category 1 work units

Maintenance processes	Total work time of an entire month (min)						
	Case 1	Case 2		Case 3		Case 4	
	Time	Time	Decrease ratio (%)	Time	Decrease ratio (%)	Time	Decrease ratio (%)
Material management	101	20.46	79.74	6.93	93.14	6.93	93.14
Preventive inspection	6696	6661.81	0.51	5010.51	25.17	4035.59	39.73
Maintenance work	7565	7516.14	0.65	7344.59	2.91	5897.29	22.05
Cost management	1280	1235.74	3.46	8.06	99.37	8.06	99.37
Labor management	158	108.92	31.06	14.11	91.07	14.11	91.07
Quality management	2322	1763.17	24.07	7.02	99.70	7.02	99.70
Total	18,122	17,306.2	4.50	12,391.2	31.62	9969	44.99

Fig. 80.6 Total cycle time of maintenance management processes

80.5 Conclusion

This study proposed an evaluation model for MMIS prior to its formal application in organizations. The case study verified that the effects of MMIS on maintenance performance can be assessed quantitatively and objectively by applying the model. This study also determined that MMIS exerts different levels of effects on the six maintenance management. With its development, ICT has become an integral component of company strategies. Decision makers of subway companies may apply the model in the MMIS planning phase in consideration of redesigning the organizational process to obtain the best fit between MMIS and the organization. Generating the model at the work process level has facilitated the identification of essential production systems and has provided an effective approach to optimizing resource allocation and workflow and to realizing modern enterprise fine management. Object-oriented decision support system for maintenance management is another potential trend in the future.

Acknowledgments This work was supported by the State Key Laboratory of Hydroscience and Engineering (No. 2012-Ky-02). The authors would like to thank the subway company and the staff of their maintenance department for providing the data for this study.

References

1. Peña-Mora F, Tanaka S (2002) Information technology planning framework for Japanese general contractors. J Manage Eng 18(3):138–149
2. Irani Z, Sharif A, Kamal MM, Love PE (2014) Visualising a knowledge mapping of information systems investment evaluation. Exp Syst Appl 41(1):105–125
3. Wong FW, Lam PT (2010) Difficulties and hindrances facing end users of electronic information exchange systems in design and construction. J Manag Eng 27(1):28–39
4. Brewer GJ, Gajendran T, Chen SE (2005) The use of ICT in the construction industry: critical success factors and strategic relationship in temporary project organizations. CIBW78 Inf Commun Technol Construct, pp 19–21

5. Xuehui A (2003) Construction project lifecycle management based on information technology. Intell Build City Inf 4:70–72 (in Chinese)
6. Smith RA (2003) Why CMMS implementations fail. Proc Water Environ Feder (5):717–729
7. Jafarnejad A, Soufi M, Bayati A (2014) Prioritizing critical barriers of computerized maintenance management system (CMMS) by fuzzy multi attribute decision making (F-MADM). J Bus Manag Rev 4(3):11–27
8. Mukhopadhyay T, Javier Lerch F, Mangal V (1997) Assessing the impact of information technology on labor productivity a field study. Decis Support Syst 19(2):109–122
9. Garg A, Deshmukh SG (2006) Maintenance management: literature review and directions. J Quality Maint Eng 12(3):205–238
10. Norman DA (2007) The design of future things. Basic Books, New York
11. Garrett JJ (2010) The elements of user experience, user-centered design for the web and beyond. Pearson Education, NJ
12. Bin L, Yao Y (2006) The design of the integrated information platform of Beijing rail transit system. Urban Rapid Rail Transit 19(4):47–49 (in Chinese)
13. Jahangirian M, Eldabi T, Naseer A, Stergioulas LK, Young T (2010) Simulation in manufacturing and business: a review. Eur J Oper Res 203(1):1–13
14. Card SK, Moran TP, Newell A (1980) The keystroke-level model for user performance time with interactive systems. Commun ACM 23(7):396–410
15. Martinez JC (2009) Methodology for conducting discrete-event simulation studies in construction engineering and management. J Constr Eng Manag 136(1):3

Chapter 81
Placing-In and Taking-Out Method for Working Wagons Under the Random Environment

Bo Qi

Abstract The placing-in and taking-out operation technology for wagons on railway sidings is one of the important works in the technic station. Determining the reasonable timing and sequence of the wagons can not only ensure the train plan is completed but also effectively shorten the time of placing-in and taking-out and save the shunting locomotive hour cost. The travel time of locomotive in private sidings is the problem of input parameters. In the actual process of operation placing-in and taking-out, the cost of each section of the sidings varies with locomotive traction weight and the driver's manipulation levels, which will inevitably affect the placing-in and taking-out time and the order of working wagons. Based on reviewing the traditional model of the existing literature, the paper constructs the optimization model of wagon placing-in and taking-out under the condition of the fluctuant travel time. Then, the hybrid intelligent algorithm is presented based on stochastic simulation, neural network, and genetic theory. The example proves the feasibility of the random model and the effectiveness of hybrid intelligence algorithm.

Keywords Railway private sidings · Placing-in and taking-out operation · Chance-constrained programming · Hybrid intelligence algorithm · Neural network · Genetic algorithm

81.1 Introduction

The placing-in and taking-out of wagons is an important part of railway transportation and production process. Its efficiency is directly related to the completion of the turnover of vehicles and goods delivery speed, the train formation plan and the completion of transportation and production index. The scientific and

B. Qi (✉)
School of Traffic and Transportation, Lanzhou Jiaotong University,
Lanzhou, Gansu, China
e-mail: qib@mail.lzjtu.cn

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*,
Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_81

reasonable arrangement of placing-in and taking-out's timing and sequencing of shunting operation is one of the key links in this goal.

In terms of layout form of private sidings, it can be divided into two categories of radial and dendritic. The difference of radial and dendriform on placing-in and taking-out operation is mainly that the former must return to the station to go to placing-in (taking-out) wagons on another private sidings when the locomotive after placing-in (taking-out) a private sidings and while the latter doesn't have to return to the station among a group of works unless in special circumstances. Both radial and dendriform's heart of the problem are seeking the best timing and sequence of placing-in and taking-out wagons. A large number of railway technological workers do some deep research on the problem of working wagons' placing-in and taking-out. The literatures [1–3] used the theory of locomotive scheduling model to study the problem of placing-in and taking-out on special line for radial and dendritic. The literature [4, 5] turned the problem of placing-in and taking-out wagons on private sidings with dendritic into the problem of looking for Hamilton loop of least shunting locomotive total trip time. It is easy to see that all the above is under the assumption that link travel time on special lines is constant.

However, in the actual process of the working wagons to placing-in and taking-out, the running time of each road section in private sidings of locomotive is fluctuating because of the affection of various subjective and objective factors, such as the change of quantity and weight of the vehicles attached on the shunting locomotive, driver control level of high and low, and train set of shunting locomotive in various loading and unloading station, for goods such as technical operation time. So the optimal solution for working wagons' timing and sequence which are obtained by the above literature theory is often not feasible scheme in actual transport production and ultimately affects the speed of vehicle flow and cargo delivery and train marshaling plan and the completion of the transportation and production quota. Based on reviewing the traditional model of the existing literature, the paper constructs the optimization model of wagon placing-in and taking-out under the condition of the travel time fluctuant and sets up the theory of hybrid intelligent algorithm which is based on stochastic simulation, neural network, and genetic theory.

81.2 Establishment of the Mathematical Model

Assumes that the station is only 1 shunting locomotive as a send operation, shunting locomotive once sends vehicle to load and unload station on private sidings and through each operating point only once, and finally the shunting locomotive back to the station. The question now is how to arrange the shunting lines and enable the shunting locomotive total travel time at least. As for the determination of the sequence of taking-out locomotive, the method is similar [5].

In order to establish a suitable model, simplified constraint, according to the practical requirements of working wagon to placing-in and taking-out on railway private lines, the paper puts forward the constraints and definitions as follows.

81.2.1 Constraint

1. The best theoretical operation of the job that taking-out wagon and placing-in wagon to separate proposed according to the literature [2, 3].
2. The loading and unloading points at each loading and unloading point are not less than 2.
3. The number of wagons waiting for each loading and unloading point is known.
4. The travel time (or distribution) of the locomotive in the private sidings is known.

81.2.2 Definition

The station number is 0, and the number of operating points that required to send wagons is 1, 2, …, n. Then, the problem of the placing-in and taking-out for working wagons in railway private sidings is expressed as the network map, which is indicated $G = [V, A, C]$.

$V = \{0, 1, \ldots, n\}$, the operating point that shunting locomotive served and the 0 is the source.

$A = \{(i,j)|i,j = 0, 1, \ldots, n, i \neq j\}$, the set of private lines that the shunting locomotive may go through.

$C = \{c_{ij}(i,j) \in A\}$, the spending time of shunting locomotive goes through the corresponding arc (i, j), including the travel time of shunting locomotive at (i, j) in pure and the operating time of the slotting technology for picking and hanging wagons.

At this point, the problem of wagon operation on the railway private sidings is turned into a traveling salesman problem [6]. The problem is changed to find the Hamilton loop of least shunting locomotive total trip time in the empowerment undirected figure G.

Set the following variables:

$$x_{ij} = \begin{cases} 1 & \text{Shunting route through the arc } (i,j) \\ 0 & \text{Shunting route does not through the arc } (i,j) \end{cases}$$

W—The total number of vehicles sends to a job for a shunting locomotive;

w_i—The number of vehicles sends to the operating point i, $i \in V$, $W = \sum w_i$.

81.2.3 The Take-Delivery Model of Traditional Operation Wagons

The optimal placing-in and taking-out sequence model of railway wagon on private sidings under the condition of regardless of the random parameters can be described as a model as follows.

$$\min Z = \sum_{(i,j)\in A} c_{ij}x_{ij} \tag{81.1}$$

s.t.

$$\sum_{i\in V} x_{ij} = 1, \quad \forall j \in V \tag{81.2}$$

$$\sum_{j\in V} x_{ij} = 1, \quad \forall i \in V \tag{81.3}$$

$$\sum_{i\in S}\sum_{j\notin S} x_{ij} \geq 1, \quad S \subset V, \ S \neq \emptyset \tag{81.4}$$

$$\sum_{j\in V} x_{0j} = 1 \tag{81.5}$$

$$\sum_{i\in V} x_{ih} - \sum_{j\in V} x_{hj} = 0, \quad \forall h \in V \tag{81.6}$$

$$\sum_{i\in V} x_{i,n+1} = 1 \tag{81.7}$$

$$W < G \tag{81.8}$$

$$x_{ij} = 0 \ \text{ or } \ 1, \quad \forall (i,j) \in A \tag{81.9}$$

where, formula (81.1) is the objective function which minimizes shunting locomotive total travel time, formula (81.2) and formula (81.3) give a limitation that each loading and unloading point on private sidings are only visited once in take (send) operation at a time, formula (81.4) shows each loading and unloading point can be connected, formula (81.5), (81.6) and (81.7) shows shunting locomotive from the station served each loading and unloading point on private sidings after back to the station, formula (81.8) is an integer type constraints, formula (81.9) is an constraints of shunting locomotive traction capacity.

81.2.4 The Stochastic Optimization Model of Placing-In and Taking-Out for Working Wagons

In the actual process of placing-in and taking-out for working wagons on railway private sidings, the time of the shunting locomotive going through the section of private line is fluctuating, and these parameters are accurate values after the implementation of scheme. The question now is that there is need to make a decision for the behavior of the system before observing parameter fluctuations. So the travel time of shunting locomotive walking on the road can be viewed as random variables, and the railway private siding is a typical stochastic system. At the same time, the paper takes into account the station dispatchers and maximum hope possibly forecasting out the worst (pessimistic) situation so as to take timely measures such as early start to the work of take or deliver and then ensure the operation of placing-in and taking-out for wagons on private sidings impregnability in order to achieve that the wagons can be taken out and set out in the time range of diagram defining. According to the chance-constrained programming form proposed by Liu [7], the placing-in and taking-out model of traditional working wagons can improve stochastic optimization model as follows for maximization objective function pessimistic value.

$$\max_{x} \min_{\overline{Z}} \overline{Z} \tag{81.10}$$

s.t.

$$\Pr\left\{ \sum_{(i,j)\in A} c_{ij} x_{ij} \le \bar{Z} \right\} \ge \alpha \tag{81.11}$$

$$\sum_{i\in V} x_{ij} = 1, \quad \forall j \in V \tag{81.12}$$

$$\sum_{j\in V} x_{ij} = 1, \quad \forall i \in V \tag{81.13}$$

$$\sum_{i\in S} \sum_{j\notin S} x_{ij} \ge 1, \quad S \subset V, \ S \neq \emptyset \tag{81.14}$$

$$\sum_{j\in V} x_{0j} = 1 \tag{81.15}$$

$$\sum_{i\in V} x_{ih} - \sum_{j\in V} x_{hj} = 0, \quad \forall h \in V \tag{81.16}$$

$$\sum_{i \in V} x_{i,n+1} = 1 \tag{81.17}$$

$$W < G \tag{81.18}$$

$$x_{ij} = 0 \text{ or } 1, \quad \forall (i,j) \in A \tag{81.19}$$

where $\min \bar{Z}$ is a pessimistic value for objective function and α is a given confidence level for decision makers in advance.

81.3 The Solution Based on the Hybrid Intelligent Algorithm

When the mathematical programming problem includes random variable, the objective function and constraint conditions become complicated and must be given a set of corresponding algorithm. The literature [8] proposed that the chance-constrained programming is transformed into a deterministic equivalent form of the traditional method, but this method has certain limitations. For this reason, the hybrid intelligent algorithm is designed to solve the above chance-constrained programming model, which is based on random simulation, neuron network, and genetic algorithm.

The hybrid intelligent algorithm flow is as follows.

Step 1: Using stochastic simulation techniques to generate input and output data for the following uncertain functions:

$$U: x \to \max \left\{ \bar{Z} \middle| \Pr \left\{ \sum_{(i,j) \in A} c_{ij} x_{ij} \le \bar{Z} \right\} \ge \alpha \right\} \tag{81.20}$$

Step 2: Training a neural network approximation to an uncertain function based on the generated input and output data.

Step 3: Initialize pop_size chromosomes and test the feasibility of chromosomes using the trained neural network.

Step 4: The chromosome was updated by cross and mutation operation [9], and the feasibility of the offspring chromosome was tested by trained neuron network.

Step 5: Calculate the target values of all chromosomes by using the trained neuron network.

Step 6: Calculate the fitness of each chromosome according to the target values.

Step 7: Chromosome selection by rotating roulette wheel.

Step 8: Repeat Step 4 to Step 7 until the completion of the number of cycle times given.
Step 9: The best chromosome as the best solution.

81.4 Analysis of Numerical Example

To compare the difference of decision scheme of placing-in and taking-out for wagons, basing on the example in the literature [5], the design examples of this paper are as follows: the travel time that the shunting locomotive after the private sidings is a random variable which obeys the normal distribution, μ is the mean value of the example in the literature [5] and the corresponding variance δ^2 see Table 81.1 and other parameters are exactly the same. Requirements are identified in a given confidence level, to determine the optimal shunting route so that the total locomotive' traveling time is least in a working process of delivering a wagon. The sketch map of joining the sidings in one railway marshaling station is shown in Fig. 81.1.

Examples are solved by C++ programming language. The parameters are as follows: Population size pop_size = 30, crossover probability P_c = 0.3, mutation probability P_m = 0.2, and genetic algebra MaxGen = 1500. After 6000 times simulation, the optimal adjustment path at a given confidence level is obtained by the hybrid intelligent algorithm. The satisfactory solution is $0 \rightarrow 4 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 6 \rightarrow 5 \rightarrow 7 \rightarrow 8 \rightarrow 0$, this time shunting locomotive spend 241 min on the whole shunting route.

And in the literature [5], the optimum dispatch path did not consider the fluctuation of spend time is $0 \rightarrow 4 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 6 \rightarrow 5 \rightarrow 7 \rightarrow 8 \rightarrow 0$, shunting locomotive spends 219 min. Comparison of the scheme of take-delivery under this two conditions, we can know, on the one hand, corresponding optimal order of the two schemes for wagons is consistent and the consistency shows that despite the fluctuation of traveling time that the locomotive on the private sidings has impact on the placing-in and taking-out schemes, as a result of the time of fluctuation is not

Table 81.1 The distribution of locomotive traveling time c_{ij} between the sidings

c_{ij}	0	1	2	3	4	5	6	7	8
0	–	(45,4)	(40,3)	(35,2)	(30,3)	(25,4)	(20,3)	(22,2)	(25,3)
1	(45,4)	–	(16,4)	(27,4)	(31,3)	(52,3)	(48,2)	(60,2)	(68,4)
2	(40,3)	(16,4)	–	(19,3)	(24,4)	(45,2)	(37,3)	(56,4)	(62,2)
3	(35,2)	(27,4)	(19,3)	–	(23,2)	(38,3)	(29,4)	(50,2)	(55,3)
4	(30,3)	(31,3)	(24,4)	(23,2)	–	(36,4)	(26,2)	(44,3)	(51,4)
5	(25,4)	(52,3)	(45,2)	(38,3)	(36,4)	–	(18,3)	(30,4)	(39,4)
6	(20,3)	(48,2)	(37,3)	(29,4)	(26,2)	(18,3)	–	(27,4)	(33,3)
7	(22,2)	(60,2)	(56,4)	(50,2)	(44,3)	(30,4)	(27,4)	–	(21,4)
8	(25,3)	(68,4)	(62,2)	(55,3)	(51,4)	(39,4)	(33,3)	(21,4)	–

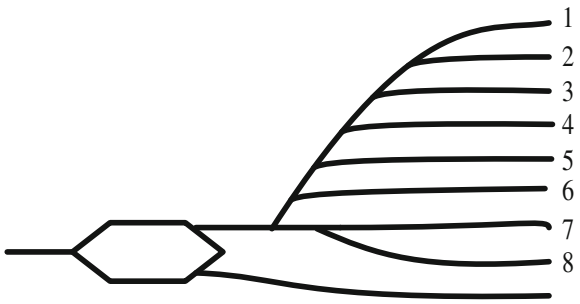


Fig. 81.1 The sketch map of joining the sidings in one railway marshaling station

purely random but the organization of a controlled stochastic and the time can be protected by asked the driver to strictly grasp the driving speed, strictly controlling the shunting locomotive in the loading and unloading point for picking and hanging, corresponding the location for cargo. On the other hand, for the same optimal placing-in and taking-out order scheme, the time of the shunting route that considering the random factors is 22 min longer than doesnot take it into account. It means that a station dispatchers should take measures such as early start the work of placing-in or taking-out, otherwise it will likely affect the realization of the train formation plan and lead to the train can not take and set out in the time range of train diagram defining.

81.5 Conclusion

1. Based on a large number of the domestic and foreign literatures, the paper presents the optimal placing-in and taking-out sequence model of railway wagons on private sidings under the condition of regardless of the random parameters.
2. The random factors in the process of taking and sending on railway private sidings are analyzed, and the stochastic optimization model of maximizing the pessimistic value of objective function is presented.
3. The hybrid intelligent algorithm is proposed, and the algorithm process is designed in detail.
4. The example with normal distribution parameters is designed, and the optimization problem is resolved by C++ language programming. The simulation results show the model, and hybrid intelligence algorithm is effective.

In this paper, the fluctuation influence of parameters is not considered, and the problem should be studied deeply in the future.

Acknowledgements This work is supported by the Program of Humanities and Social Science of Education Ministry of China (No. 12YJC630152), the Natural Science Foundation of Gansu Province (No. 1506RJZA083), and the Higher School Science Foundation of Gansu Province (No. 2014A-044).

References

1. Jing Y (1998) Research on the algorithm of placing-in and taking-out for wagons on private sidings of the branch type. J Chengdu Univ 17(4):36–42 (in Chinese)
2. Li W, Du W (1997) Mathematical model and algorithm for placing-in and taking-out wagons on private sidings of dendritic type. J Henan Univ 27(2):1–8 (in Chinese)
3. Du W, Li W (1995) The algorithm for the direct train on sidings of radial. J Southwest Jiao Tong Univ 30(5):503–508 (in Chinese)
4. Shi H, Peng Q, Guo H (2005) Hamilton graphic method for taking and delivering wagons on private sidings of branch type. China Railway Sci 26(2):132–135 (in Chinese)
5. Yu S, Li X (2002) Transportation model and algorithm for operation wagons. Railway Transp Econ 24(12):46–48 (in Chinese)
6. Papadimitriou CH, Steiglitz K (1982) Combinatorial optimization: algorithms and complexity. Prentice-Hall Inc, USA
7. Liu B (1999) Uncertain programming. Wiley, New York
8. Baoding L, Ruiqing Z, Gang W (2003) Uncertainty programming and application. Tsinghua University Press, Beijing (in Chinese)
9. Mou H, Liu L (2006) Fuzzy stochastic shortest path problem model and algorithm. J LanZhou JiaoTong Univ 25(3):118–122 (in Chinese)

Chapter 82
Risk Early Warning and Control of Rail Transportation Based on Fuzzy Comprehensive Evaluation Method in Wuhan City

Wenman Chen, Houyuan Ye and Jun Yun

Abstract As one of the most important parts of city public transportation, rail transportation has many operational risks which are increasingly being paid attention to by the government and research development institutions. How to evaluate efficiently the risk of rail transportation in Wuhan is the main content of this paper. In this paper, the fuzzy comprehensive evaluation method is used to evaluate the risk early warning level of rail transportation in Wuhan. According to the results of the assessment, the relevant and effective suggestions for the development of rail transportation in Wuhan are given.

Keywords Rail transportation · Risk early warning · Fuzzy comprehensive evaluation · Risk control

82.1 Introduction

Urban rail transit refers to having the characteristics of large capacity, high speed, security, punctuality, environmental protection, energy saving, and other modes of transportation, including the subway, light rail, rail, and tram. The government generally believes that it is necessary to give priority to the development of rail transportation as the backbone of urban public transport system in order to solve the traffic problem [1]. In 1984, the delegation of the Belgian Rail Transportation company visited Wuhan which became the beginning of the construction of Wuhan rail transit. The State Planning Commission officially approved the planning of the first phase of Wuhan Rail Transit Line 1 in 1999. The construction cycle up to 10 years in Wuhan Urban Rail Transit Line 1 project was finally opened to traffic in

W. Chen (✉) · H. Ye · J. Yun
Management School, Wuhan University of Technology, Building Four, Wuhan 100029, People's Republic of China
e-mail: 619472911@qq.com

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_82

July 2010. Between 2010 and 2014, Wuhan had successively opened four rail transit lines and the traffic capacity of Wuhan has greatly improved.

As the center of economic development in middle area, Wuhan has put tremendous work to develop rail transit to be the backbone of urban public transport system. And as a result, the rail transportation can promote economy to a certain extent and solve the traffic flow problems in Wuhan. In recent years, the government of Wuhan increases investment to develop rail transportation and is committed to building rail transportation into the mainstream of Wuhan public transport system traffic patterns.

82.2 Rail Transportation Risk Early Warning

As a form of transportation, rail transportation may have many uncontrollable factors in the operation process, which brings risks to the safety of the rail transportation. Rail transportation is a complete, continuous, and dynamic traffic system, and we can divide the risk of the system into two aspects: system risk and non-system risk [2]. From the two angles, this paper finds out indicators related to the two aspects and uses comprehensive evaluation method to evaluate the status of the risk early warning of Wuhan rail transit.

82.3 The Fuzzy Comprehensive Evaluation

The fuzzy comprehensive evaluation that is based on some concepts of fuzzy mathematics can evaluate the actual comprehensive evaluation problems. The method is firstly to determine the evaluation factor, the evaluation index $U = \{u_1, u_2, \ldots, u_m\}$, and the evaluation grade $V = \{v_1, v_2, \ldots, v_n\}$. The membership degree is r_{ij} of the factor $U = \{u_1, u_2, \ldots, u_m\}$ focus on the degree of $v_j (j = 1, 2, 3, \ldots, n)$, and we can obtain the single factor evaluation set $r_i = (r_{i1}, r_{i2}, \ldots, r_{im})$ from factor u_i. Each evaluation object determines the fuzzy relational judgment matrix from Y to X, and the fuzzy relational judgment matrix is named R:

$$R = (r_{ij})_{m \times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

Besides, it is necessary to empower the evaluation factor and introduce a fuzzy subset $A = (a_1, a_2, \ldots, a_m)$ which said the weight or weight set of U. The decision

matrix and the fuzzy subset are synthesized by the fuzzy operator method and obtain the decision results finally [3].

82.3.1 Determine Evaluation Indexes

Through analyzing the present situation of rail transportation in Wuhan, the corresponding indexes of fuzzy comprehensive evaluation are determined. Then, we can determine the correlation weight of each index by expert scoring method [4]. The following Table 82.1 is the indicators table.

82.3.2 Evaluation

By analyzing the operational risk of rail transportation, the evaluation grade is divided into four levels: severe early warning (SEW), moderate early warning (M1EW), mild early warning (M2EW), and no early warning (NEW). Besides, the evaluation matrix of the evaluation set is determined by the expert analysis method and the results are as follows [5].

Table 82.1 Indicators of rail transportation risk

Target layer	Standard layer	Primary indicator layer	Secondary indicator layer
Rail transportation risk	System risk	Economic risk	Government investment
			Social benefit
		Technical risk	Technical level of contractor
			Technical upgrading
		Natural and social risks	Flood
			Fire
			Social stability
	Non-system risk	Infrastructure risk	Data exchange analysis capability
			Server performance
			Server reliability
		Standardization risk	Scientific standards
			Standards upgrading and maintenance
			Standard application degree
		Management risk	Timely troubleshooting capability
			The rationality of management system and organization

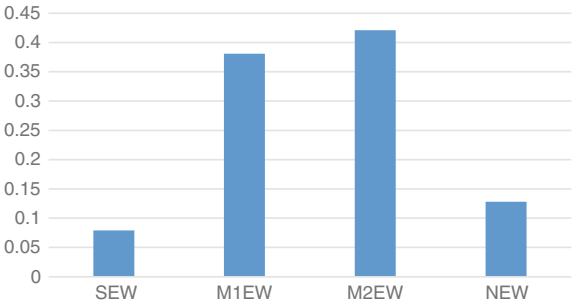


Fig. 82.1 The risk level of rail transportation in Wuhan

$$R = \begin{bmatrix} 0 & 0.667 & 0.333 & 0 \\ 0 & 0 & 0.333 & 0.667 \\ 0.167 & 0.667 & 0.166 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0.333 & 0.334 & 0.333 \\ 0.125 & 0.3 & 0.3 & 0.275 \\ 0.167 & 0.166 & 0.667 & 0 \\ 0.167 & 0.667 & 0.166 & 0 \\ 0 & 0.333 & 0.333 & 0.334 \\ 0.167 & 0.333 & 0.5 & 0 \\ 0.167 & 0.333 & 0.5 & 0 \\ 0 & 0.333 & 0.334 & 0.333 \\ 0.167 & 0.667 & 0.66 & 0 \\ 0 & 0.333 & 0.667 & 0 \\ 0.167 & 0.333 & 0.5 & 0 \\ 0 & 0.167 & 0.667 & 0.166 \\ 0.167 & 0.333 & 0.5 & 0 \end{bmatrix}$$

We can input the relevant data and utilize the fuzzy comprehensive evaluation method to evaluate the risk level of rail transportation. As a result, we obtain the following result of comprehensive evaluation:

$$B = (0.079, 0.381, 0.412, 0.128)$$

After using fuzzy comprehensive evaluation method, the figure of risk level of rail transportation in Wuhan is shown as follows (Fig. 82.1).

82.4 Risk Control Measures

From the results of fuzzy evaluation, we can see the risk level of rail transportation in Wuhan is mainly in mild and moderate warning areas. If the risk prevention is inappropriate, it is easy to turn the risk from the mild warning mode into the

moderate warning mode and lead to the significant increasing cost of rail transportation risk control [6]. According to the result of the risk state of rail transportation in Wuhan, the government should formulate targeted emergency mechanism and policy.

82.4.1 Non-system Risk Control Mechanism

The construction level of rail transportation is directly related to the reliability, security, and stability of the rail transportation operation. So it is extremely essential to strengthen the basic construction in order to reduce the risk of rail transportation and control risk of infrastructure with the help of the maintenance personnel and the platform director of track construction management. The professional staff should develop reasonable and reliable standards to reduce the operation risk of rail transportation. The monitoring system of rail transit operation platform should be controlled by the top management personnel and formulate the corresponding monitoring policies to find the potential risk in the operation process of rail transit and to deal with risks in time [7].

82.4.2 System Risk Control Mechanism

Rail transportation as an important part of urban traffic, the government should increase the fund investment and promote the operation and development level of urban traffic in Wuhan. The rail transportation in China has developed rapidly in recent years, and as the center of economic development in the central city, Wuhan also vigorously develops rail transportation. The construction technology of railway transportation is essential to establish the rail transportation system. Therefore, the government should pay more attention to the core role of the rail transportation technology in the process of the establishment of rail transportation system to reduce the technical risk [8]. In fact, natural disaster is the factor that must be considered in the risk of rail transportation and administrative departments can control risk through various effective measures.

82.5 Conclusion

In this paper, we can see apparently that the risk level of rail transportation in Wuhan is mild by using the fuzzy comprehensive evaluation method. According to the result of the risk level, the relevant departments of rail transportation can improve the rail transportation development mechanism and establish a sound rail transportation risk warning and risk management system.

References

1. Stock JR (1990) Managing computer, communication and information technology strategically: opportunities and challenges for warehousing. Logistics Transp Rev 25(2):133–148
2. Lijie W, Fulin S (2010) Study on risk early warning and risk prevention and control of supply chain. Bus Adm 10:40–42 (in Chinese)
3. Armbrust M, Fox A, Griffith R, Joseph AD, Katz RH, Kaminski A, Lee G (2009) Engineering and computer sciences University of California at Berkeley. Tech Rep 28:168–194
4. Wen-Hong DUE (2007) The application norm research to the digital signature technology of the Hospital information system. ICTE 3294–3298
5. Hongbo W, Guoliang S (2002) Risk early warning mechanism. Economic Management Press, Beijing, pp 150–151 (in Chinese)
6. Millet I, Wedley WC (2002) Modeling risk and uncertainty with the analytic hierarchy process. J Multi-Criteria Decis Anal 11(2):97–107
7. Chengsong W (2011) Study on the financial risk early warning index system of China. Res Technol Econ Manag 1:19–24 (in Chinese)
8. Cuicui L (2012) Risk early warning and control of logistics public information platform. Wuhan Univ Technol 31(4):79–81 (in Chinese)

Chapter 83
Optimization of Fare Gates Layout of Rail Transit Station Based on BP Neural Network

Qingyu Zhang, Haiying Li and Xinyue Xu

Abstract In order to improve the efficiency of fare gates, this paper presents an approach to optimize fare gates layout. Beside shortest distance, the most regular influential factor, another 6 factors are concluded to analyze passengers fare gates choice, including the number of fare gates, queue length and pedestrian-flow-cross index and so on. The model of pedestrians fare gates choices calibrates the weight value of influential factors based on BP neural network, which solves the problem of quantitative presentation between pedestrians fare gates choices and influential factors. Then, optimize fare gates layout to improve the efficiency of fare gates for the most important factor. In the example of a station on Beijing metro, high prediction accuracy (87.206 %) is achieved to choose fare gates and the efficiency of fare gates is improved by 23.7 %. It is proved that the optimization of fare gates choice presented in this paper has better performance.

Keywords Rail transit station · Fare gate · Choice · BP neural network

83.1 Introduction and Literature Review

Along with the increasing demand of urban transportation and the skyrocketing of pedestrian flow in rail transit station, the conflict in equipment and facilities between supply and demand is more obvious and the congestion is becoming more and more serious. Fare gates layout should take long-term pedestrian flow, instal-

Q. Zhang
School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China
e-mail: 13120929@bjtu.edu.cn

H. Li (✉) · X. Xu
State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
e-mail: hyli@bjtu.edu.cn

Q. Zhang · H. Li · X. Xu
Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, 100044 Beijing, China

Y. Qin et al. (eds.), *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Lecture Notes in Electrical Engineering 378, DOI 10.1007/978-3-662-49370-0_83

ling conditions and reserved location into consideration according to Code for Design of Metro. Fare gates layout is becoming modelized and simplistic, resulting to pedestrian flow lines interweaving and the imbalance using of fare gates.

The study of facilities layout in rail transit stations based on the behavior of passengers has some achievements. The rationality of station facilities is evaluated through the delay simulation of all kinds of facilities [1]. A pedestrian walking behavior model based on the minimum cost is presented to evaluate facilities of Lisbon station [2]. Passenger flow assignment and facilities simulation inside the station are presented [3]. The law of travel preferences and path choice behavior of passengers in the rail transit stations are studied based on RP survey [4]. A static calculation method is carried out, which is about fare gates layout of railway passenger station based on ticketing velocity and passenger dispatch volumes [5]. Meticulous researches on pedestrian behavior and facilities are organized in literatures mentioned above. Since passengers are affected by the environment and equipment in stations, their behaviors are very complex. However, the studies about fare gates layout based on the influential factors of fare gates choice are not enough, the achievements are kind of scattered.

Two contributions to the prediction of fare gates choices of rail transit station are offered: (1) Pedestrian-flow-crossing index, wide fare gate and guider are introduced to influential factors of fare gates choice, apart from the number of fare gates, shortest distance, queue length, which are common influential factors. (2) Measure the interreaction mechanism among all of the influential factors and the way how they influence the choosing result quantitatively.

The rest of the paper is organized as follows. In Sect. 83.2, we analyze some assumptions for the study of fare gates layout optimization. Then, we present a prediction model of fare gates choices in Sect. 83.3. Furthermore, Sect. 83.4 provides a case study to verify the model proposed above. The last section concludes the paper and suggests directions for future research.

83.2 Model Assumptions

The most congested period in rail stations is during peak hours, a variety of facilities usage is closely related to passenger route choice. Therefore, it is essential to establish model of pedestrian fare gates choices in rush hours, for assessing the reasonableness of the station facilities layout. Two assumptions of the pedestrian fare gates choices are presented based on pedestrians characteristic as follows: Assume that pedestrians are rational when making fare gates choices and all of the passengers are inbound.

83.3 Optimization of Fare Gates Layout Based on BP Neural Network

83.3.1 Development of the Model

Kadali et al. [6] concludes that ANN has a better prediction with possibility to consider the effect of more number of variables on the pedestrian gap acceptance behavior as compared to the MLR model. Therefore, BP neural network is adopted to establish a prediction model of fare gates choice.

It has been mathematically proven that BP neural network with a single hidden layer is a universal function [7], so a three-layered ANN is adopted to establish the fare gates choice model in this paper, consisting of the input layer, the hidden layer and the output layer, as is shown in Fig. 83.1.

83.3.1.1 Input Layer

The number of neurons in the input layer corresponds to the number of input parameters in the model. In this paper, the following influential factors are selected to predict the probability of passengers choosing fare gates as input parameters. As shown in Fig. 83.2, locations leaving from security check are marked as ST1-2 and fare gates1-2 are marked as FG1-2. The Influential factors and their symbols are shown in Table 83.1.

83.3.1.2 Hidden Layer

The number of neurons in the hidden layer is related to the number of training samples and neurons in the input/output layer, shown in Eq. (83.1).

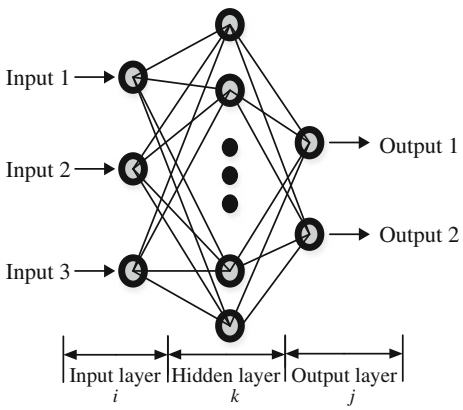


Fig. 83.1 Topological structure of the BP neural network

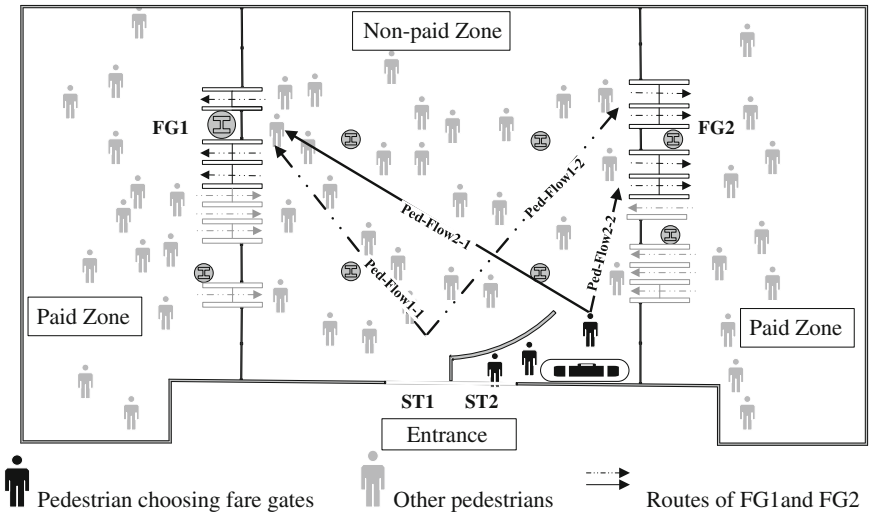


Fig. 83.2 Fare gates choice in non-paid zone

Table 83.1 Influential factors of fare gates choices

Input parameters	Symbol	Remark
Shortest distance	D_{ij}	The shortest distance from the STi to FGj
Number of fare gates	N_j	The number of fare gates in FGj
Queue length	L_j	The queue length before FGj
Pedestrian-flow	M_{ij}	The pedestrian flow crossing frequency from STi to FGj
Wide fare gate	W_j	The number of wide gates in FGj
Guider	Z_j	The number of guiders in FGj

$$N_{\mathrm{h}} = \frac{N_{\mathrm{in}} + N_{\mathrm{out}}}{2} + \sqrt{M_{\mathrm{t}}} \tag{83.1}$$

where $N_{\mathrm{h}}, N_{\mathrm{in}}, N_{\mathrm{out}}$ are the number of neurons in the hidden layer, input layer, and output layer respectively, and M_{t} is the number of training samples.

83.3.1.3 Output Layer

The value of the fare gates unit that pedestrian chooses are considered to be the output Y of the network, "0" and "1" indicating FG1 and FG2 separately.

83.3.2 Model Training

Input and normalize the training samples. Shoot videos of fare gates layout in non-paid zone to obtain the parameters mentioned above. Initialize the weights and threshold of every layer of BP neural network as the random values. In order to minimize the error, 5000 trials are carried out of model training and testing for each number of neurons. At the same time, initialize the learning precision and other parameters.

(1) Obtain the output of hidden layer $\{G_t\} = \{g(1), g(2), \ldots, g(t)\}$ through $g(t) : g(t) = \frac{2}{1+e^{2t}} - 1$. (2) Get the result of output layer$\{Y\}$ via purelin. (3) Analyze whether circulate all of the samples or not. If not, return to (1). (4) Calculate the error ε and analyze whether ε reaches the precisions. Stop learning if $\varepsilon < e$. Anti-normalize $\{Y\}$ and record weight $\{w_{ik}\}$, $\{w_{kj}\}$. Otherwise, modify $\{w_{ik}\}$. and $\{w_{kj}\}$, and turn to (1).

83.3.3 Optimization of Fare Gates Layout

Optimize the layout of fare gates through changing parameters of the most important influential factor. Evaluate fare gates of different layout by Eq. (83.2).

$$Q = 1 - \sqrt{E[(K_j - E(K))^2]}, \quad K_j = \frac{p_j}{c_j}, j = 1 - 2. \tag{83.2}$$

where Q is the effectiveness of different fare gates units. p_j is the number of pedestrians who choose FGj in an hour. c_j refers to the design capacity of FGj.

83.4 Case Study

Usually, fare gates inbound are divided into several units in Beijing metro. A station on Beijing Metro Line is selected to be a case to analyze. The station, a typical commuter station, has heavy inbound flow and light outbound flow in morning peak hours. Fare gates for inbound and outbound are shown in Fig. 83.1.

83.4.1 Data Collection and Process

Conduct a survey during the morning peak (07:00–8:00) on 5 weekdays, when the hourly volume reached a maximum. 385 effective samples are obtained eventually. 67, 16 and 17 % of all of the samples are selected for training, validation and test, separately. The input and output variables are shown in Table 83.2.

Table 83.2 Total sample table

			Input												Output
	i	j	D_{i1}	D_{i2}	N_1	N_2	L_1	L_2	M_{i1}	M_{i2}	W_1	W_2	Z_1	Z_2	Y
1	2	2	1764	1442	3	4	5	3	1	0	1	0	1	0	1
2	2	1	1764	1442	3	3	4	6	1	0	1	0	1	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
385	1	1	1554	1871	2	4	4	5	0	1	0	0	1	0	0

83.4.2 Model Training and Validation

The final topological structure of network is 12-26-1 according to Eq. (83.1). The expected error, learning rate, adding momentum factor and training times are 0.001, 0.05, 0.5, 7000, respectively The training, validation and test samples are packet randomly in every trail, acquiring 7000 performance indexes eventually. Evaluate performance of the model with the output matching degree of prediction and target (prediction accuracy), namely Eq. (83.3) and the distribution of prediction accuracy shown in Fig. 83.3.

$$\bar{e} = \frac{\sum_{k=1}^{u} e_k}{u}, \quad e_k = \begin{cases} 1, Y_k = C_k \\ 0, Y_k \neq C_k \end{cases} \tag{83.3}$$

where u refers to the number of all of the test samples and $Y_k = C_k$ represents that target and prediction are equal.

The prediction accuracy is relatively high, for the average of prediction accuracy being 87.206 % and minimum of prediction accuracy of confidence level of 95 % being 82.310 %, according to Fig. 83.3.

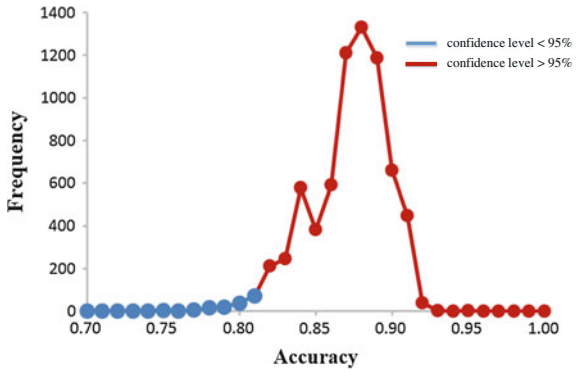


Fig. 83.3 Distribution of prediction accuracy

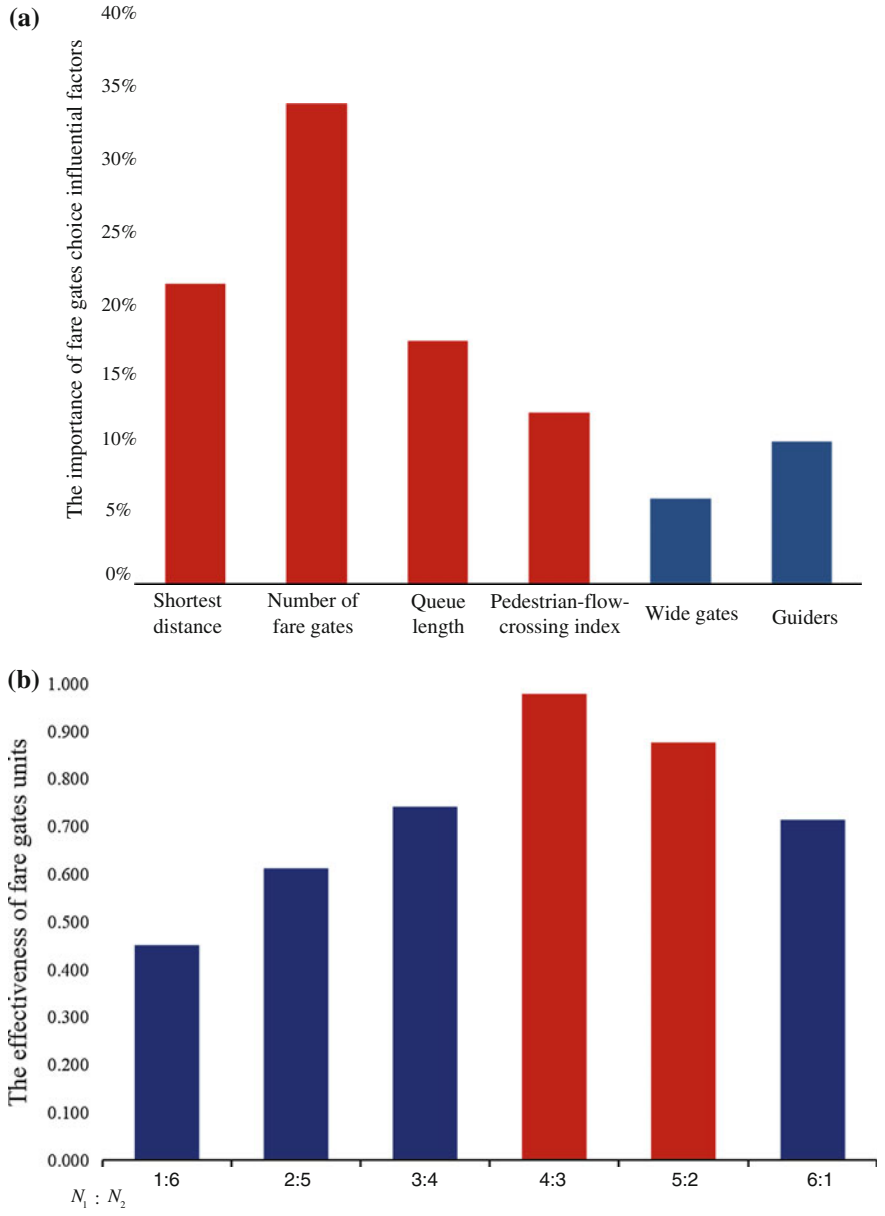


Fig. 83.4 **a** Distribution of prediction accuracy. **b** Effectiveness of different fare gates layout

83.4.3 Optimization of Fare Gates Layout

The importance of input parameters to output variables are calculated through the product of connection weights, according to Garson [8]. The weights of BP neural network (influence coefficients) are shown in Fig. 83.4a.

According to Fig. 83.4a, we can see that the fare gates choice influential factors are ranked by number of fare gates, shortest distance, queue length, pedestrian-flow-crossing index, guiders and wide gates. The number of fare gates is the most important influential factor according to Fig. 83.4a. Hence, this model designed to optimize fare gates layout for the number of fare gates. The effectiveness of fare gates units is shown in Fig. 83.4b, with different numbers of fare gates. We can see that, the effectiveness of fare gates units is significantly increased with the layout of N_1:N_2 = 4:3, which is much higher than the current layout of N_1:N_2 = 3:5. And the optimization is practicable.

83.5 Conclusion

From the perspective of pedestrian-flow organization and fare gates layout, this paper sorts influential factors of pedestrian's fare gates choices in non-paid zone by the importance. And the optimization of fare gates layout based on BP neural network achieves ideal result through validation. The optimization model solves the problem of relationship between fare gates choice with influential factors, realizes the goal of improving the utilization of fare gates. However, it is necessary to investigate more types of stations in the future to improve the universality of the model instead of a few stations included.

Acknowledgments This work is supported by the research fund of State Key Lab of Rail Traffic Control and Safety (RCS2015ZZ002).

References

1. Fox H (1994) User guide to PEDROUTE 3.15. Halcrow Fox, London
2. Hoogendoorn S (2004) Pedestrian flow modeling by adaptive control. Transp Res Record J Transp Res Board 1878:95–103
3. Lee JYS, Lam WHK, Wong SC (2001) Pedestrian simulation model for Hong Kong underground stations. In: Proceedings of intelligent transportation systems. IEEE, New York, pp 554–558
4. Gwynne S, Galea ER, Lawrence PJ et al (2001) Modeling occupant interaction with fire conditions using the building EXODUS evacuation model. Fire Saf J 36(4):327–357
5. Ehtamo H, Heliövaara S, Hostikka S, et al (2010) Modeling evacuees' exit selection with best response dynamics. In: Pedestrian and evacuation dynamics 2008. Springer, Berlin, pp 309–319

6. Kadali B R, Vedagiri P, Rathi N (2015) Models for pedestrian gap acceptance behavior analysis at unprotected mid-block crosswalks under mixed traffic conditions. Transp Res Part F Traffic Psychol Beh, pp 114–126
7. Hornik K, Stinchcombe M, White H (1989) Multilayer feed forward networks are universal approximators. Neural Netw 2(5):359–366
8. Garson DG (1991) Interpreting neural network connection weights