

Songlin Sun
Na Chen
Tao Tian
Editors

Signal and Information Processing, Networking and Computers

Proceedings of the 3rd International
Conference on Signal and Information
Processing, Networking and Computers
(ICSINC)



Lecture Notes in Electrical Engineering

Volume 473

Board of Series editors

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

About this Series

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

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

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

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

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

Songlin Sun · Na Chen · Tao Tian
Editors

Signal and Information Processing, Networking and Computers

Proceedings of the 3rd International
Conference on Signal and Information
Processing, Networking and Computers
(ICSINC)

 Springer

Editors

Songlin Sun
Beijing University of Posts
and Telecommunications
Beijing
China

Tao Tian
Beijing University of Posts
and Telecommunications
Beijing
China

Na Chen
Beijing University of Posts
and Telecommunications
Beijing
China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-7520-9 ISBN 978-981-10-7521-6 (eBook)
<https://doi.org/10.1007/978-981-10-7521-6>

Library of Congress Control Number: 2017961806

© Springer Nature Singapore Pte Ltd. 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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

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

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The 3rd International Conference on Signal and Information Processing, Network and Computers (ICSINC 2017) provided a forum for researchers, engineers, and industry experts to discuss recent development, new ideas, and breakthrough in information and communications, computer technologies.

We were honored to have the keynote speakers invited to present their outstanding achievements and understanding on the following topics: spatiotemporal video resolution enhancement and recent advances in 5G/B5G.

ICSINC 2017 received 177 papers submitted by authors, 63 papers were accepted, and finally, 61 papers were included in the final conference proceedings. The accepted papers were presented and discussed in seven regular technical sessions, one workshop, two special sessions, and posters.

On behalf of the ICSINC 2017 committee, we would like to express our sincere appreciation to the TPC members and reviewers for their tremendous efforts. In particular, we appreciate all the sponsors for their generous support and advice, including Beijing University of Posts and Telecommunications, Chongqing University of Technology, China Unicom, Springer, HuaCeXinTong, and ShunXingChangMao. Finally, we would also like to thank all the authors and participants for their excellent work and cooperation.

Songlin Sun
Yue Wang
ICSINC 2017 General Co-chairs

Organization

International Steering Committee

Songlin Sun	Beijing University of Posts and Telecommunications, China
Michel Kadoch	Universite du Quebec, Canada
Takeo Fujii	The University of Electro-Communications, Japan
Ju Liu	Shandong University, China
Chenwei Wang	DOCOMO Innovations (DoCoMo USA Labs), USA

General Co-chairs

Songlin Sun	Beijing University of Posts and Telecommunications, China
Yue Wang	China Academy of Space Technology, China

Technical Program Committee Chairs

Xinzhou Cheng	China Unicom Network Technology Research Institute, China
Wei Liu	Chongqing University of Technology, China
Guowei Yang	China Academy of Space Technology, China

Organizing Committee Chair

Tao Tian	Beijing University of Posts and Telecommunications, China
----------	---

General Secretary

Na Chen	Beijing University of Posts and Telecommunications, China
---------	---

Sponsors

Beijing University of Posts and Telecommunications



Chongqing University of Technology



China Unicom



Springer



HuaCeXinTong



ShunXingChangMao



Contents

Wireless Communication Systems

Application of NSGA-II Algorithm to Energy and Spectral Efficiency Trade-off in Massive MIMO Systems with Antenna Selection	3
Qiaoqiao Zhang, Xuebin Sun, and Dianjun Chen	
A PHY-Based Secret Sharing Scheme in MIMO Systems	11
Zhuoru Jian, Hai Huang, Xiaojun Jing, and Jia Li	
Long Short-Term Memory Network for Wireless Channel Prediction	19
Xiaoyun Tong and Songlin Sun	
A Practical Implementation of TD-LTE and GSM Signals Identification via Compressed Sensing	27
Jianyi Yang, Liang Yin, Lin Sang, Xin Zhang, Siqing You, and Hongjie Liu	
Monitoring and Avoidance of Atmospheric Duct on Interference Optimization in TD-LTE System	36
Ao Shen, Yang Zhang, Bao Guo, Guozhi Wang, Yan Gao, Jixiang Liu, Dayang Liu, Yi Liu, Xiaochun Hu, and Tao Xie	
Load Balancing and Interference Management in Heterogeneous Networks	46
Rihan Wu, Songlin Sun, and Yasir Ullah	
QoE Evaluation Model for Wireless Video Network Business	53
Tao Tian, Songlin Sun, Chenwei Wang, and Xinzhou Cheng	
A Spectrum Sensing Scheme with Multiple Users	65
Junsheng Mu, Xiaojun Jing, Chenchen Sun, and Jia Li	

A New Method of Spectrum Sensing in Cognitive Radio Based on Statistical Covariance Matrix	72
Zhaocong Sun, Xiaojun Jing, and Jia Li	
An Enhanced Double Threshold Energy Detection in Cognitive Radio	80
Xuan Zhou, Xiaojun Jing, Hai Huang, and Jia Li	
An Energy Detection Based on Coefficient of Variation for Spectrum Sensing in Cognitive Radio	87
Wenwen Zhang, Xiaojun Jing, and Jia Li	
Performance Analysis for User-Centric Cloud Radio Access Network in Millimeter Wave	95
Yangying Zhang, Hai Huang, Xiaojun Jing, and Jia Li	
Millimeter Wave Cloud Radio Access Network Coverage and Capacity	103
Jinxia Hu, Xiaojun Jing, and Jia Li	
Optimal Downtilts for 3D Beamforming Based on Greedy Algorithm in Massive MIMO Networks with Imperfect CSI	112
Kang Zhang, Kwahak Jon, and Songlin Sun	
Dynamic User Scheduling Algorithms for Massive MIMO Multicast System	121
Xinran Zhang and Songlin Sun	
A Bat Algorithm Based on Centroid Strategy	129
Siqing You, Tongjuan Liu, Fei Xue, Hongjie Liu, and Zhaoqun Qi	
Algorithms Optimization and Implementation	
An Optimized Daisy-chain Topology for Multi-load Interconnection in High-speed and High-density Electronic Systems	139
Fengrui Guo, Xingming Li, Shanqing Hu, and Yanyan Qin	
Realization and Optimization of Pulse Compression Algorithm on OpenCL-Based FPGA Heterogeneous Computing Platform	147
Jiacheng Yu, Xingming Li, Shanqing Hu, and Yuwei Wang	
The Research of SAR Processing Performance Based on Multi-core GPU	156
Yuwei Wang, Xingming Li, Shanqing Hu, and Jiacheng Yu	
Ship Detection in Optical Satellite Images Based on Sparse Representation	164
Haotian Zhou, Yin Zhuang, Liang Chen, and Hao Shi	

Application of Back Propagation Neural Network with Simulated Annealing Algorithm in Network Intrusion Detection Systems 172
Chen Chang, Xuebin Sun, Dianjun Chen, and Chenwei Wang

An Improved Binary Bee Colony Algorithm for Satellite Resource Scheduling Method 181
Pan Zhao, Xuebin Sun, and Ping Chen

Automatic Liver Segmentation on CT Images 189
Torecan Celik, Hong Song, Lei Chen, and Jian Yang

An Improved Blind Spectrum Sensing Algorithm Based on QR Decomposition and SVM 197
Yaqin Chen, Xiaojun Jing, Wenting Liu, and Jia Li

Sentiment Analysis Using Modified LDA 205
Jingyi Ye, Xiaojun Jing, and Jia Li

Co-training Based on Multi-type Text Features 213
Wenting Liu, Xiaojun Jing, Yaqin Chen, and Jia Li

GNSS Spoofing Jamming Recognition Based on Machine Learning 221
Pan Gao, Songlin Sun, Zhen Zeng, and Chenwei Wang

TinyPEP: Tiny Pairwise-Key Establishment Protocol for Wireless Sensor Networks 229
Wei Liu and Rong Luo

An Innovative Indoor Location Algorithm Based on Supervised Learning and WIFI Fingerprint Classification 238
Cong Chao and Men Xiaoran

Satellites and Remote Sensing

Deep Learning and Machine Learning for Object Detection in Remote Sensing Images 249
Guowei Yang, Qiang Luo, Yinding Yang, and Yin Zhuang

Design and Implementation of Automatic Interpretation System for Remote Sensing Image 257
Linna Ni and Yu Jiang

An Improved Cloud Detection Method of Optical Remote Sensing Image 265
Yang Gao, Hao-tian Zhou, and Liang Chen

A Novel Method to Analyze Dual Camera Pointing Direction Difference of Remote Sensing Satellite 272
Kan Cheng, Zihao Cui, Tao He, and Mengjie Shi

Application of Wi-Fi Wireless Network on Attitude and Orbit Control System of Satellite 281
 Jinpeng Wang, Yue Wang, Yi Zhan, Mingyu Xie, and Jianzhao Ding

Design and Implementation of the GEO Remote Sensing Satellite for Intelligent Applications 290
 Fengjing Liu, Ning Liu, Xiang Li, and Guo Li

Application of Computer Simulation Technology in Observing Effectiveness Analysis of the Satellite 298
 Yunhe Liu, Fengjing Liu, Kuai Yu, Jian Liu, Yongheng Zou, and Guo Li

Investigation on the Electromagnetic Field Leakage Effect Dominated by the Cable Penetration for a Spacecraft 309
 Ran Li, XiaoYong Yang, Xiang Li, and XuYang Du

Big Data Workshop

Research on Impact of LTE RSSI Based on Network Data Correlation Analysis and Optimization Practice 319
 Mingxin Li, Tianbiao Tang, Juanjuan Tan, Hao Guo, and Hongxi Liao

Customer Churn Analysis for Telecom Operators Based on SVM 327
 Runsha Dong, Fei Su, Shan Yang, Xinzhou Cheng, and Weiwei Chen

Evaluating LTE Service Performance for High-Speed Rail Cells via User Classification Model 334
 Alexis Huet, Mantian (Mandy) Hu, Jibin Wang, and Ye Ouyang

Research on High-Efficient Dynamic Evaluation Method of Operation Stability in Mobile Radio Network 342
 Jian Guan, Wensheng Li, Haina Ye, Jie Gao, Yongfeng Wang, and Xinzhou Cheng

A Novel PCI Optimization Method in LTE System Based on Intelligent Genetic Algorithm 350
 Ao Shen, Bao Guo, Yan Gao, Tao Xie, Xiaochun Hu, Yang Zhang, Jinhu Shen, Yuan Fang, Guozhi Wang, and Yi Liu

Data Mining for Base Station Evaluation in LTE Cellular Systems 356
 Lexi Xu, Xueqing Zhao, Yanli Yu, Yuting Luan, Xinzhou Cheng, Jie Gao, Jian Guan, and Kun Chao

Blind Video Quality Assessment Based on Human Visual Speed Perception and Nature Scene Statistic 365
 Shiyu Zhou, Xiuyan Xia, Meng Ran, Luhan Wang, and Chen Cheng

A Downlink Coverage Self-optimizing Algorithm for LTE Cellular Networks Based on Big Data Analytics 373
 Jie Gao, Xinzhou Cheng, Lexi Xu, Lijuan Cao, and Chen Cheng

Research on Wireless Network Planning of Railway TD-LTE System 381
 Kai Yu, Yanli Yu, and Jie Xiong

Data Mining Based Modeling and Application of Mobile Video Service Awareness 389
 Kun Chao, Pengfei Wang, Haina Ye, Lexi Xu, Xinzhou Cheng, Mingjun Mu, and Chen Cheng

The Research of Virtual Drive Test Based on MR and CDR 397
 Zhiqiang Lv, Saibin Yao, Ling Li, Yongjia Qi, and Jialong Liang

Big Data Research on Driving Behavior Model and Auto Insurance Pricing Factors Based on UBI 404
 Heng Zhang, Lexi Xu, Xinzhou Cheng, Weiwei Chen, and Xueqing Zhao

Compliance Testing for Data Quality Assurance: Definitions, Models and Applications 412
 Xu Mao, Fei Su, Kuitong Xian, and Kaicheng Xu

Telecom Big Data Based Electromagnetic Wave Research Under Haze and Rainstorm 420
 Xinzhou Cheng, Lexi Xu, Tao Zhang, Chen Cheng, Weiwei Chen, Heng Zhang, Yuwei Jia, and Haina Ye

Analysis and Optimization of Video Fluency Based on Big Data 429
 Mingjun Mu, Yuwei Jia, Weiwei Chen, and Yongfeng Wang

Clothing Recommendation System Based on Advanced User-Based Collaborative Filtering Algorithm 436
 Yu Liu, Jingwen Nie, Lexi Xu, Yue Chen, and Bingyu Xu

OSS Data Based LTE Wireless Coverage Efficiency Analysis Method 444
 Xingyu Fan, Jun Lu, and Weiwei Chen

A Radio Network Differentiation Parameter Optimization Algorithm Based on Clustering 452
 Guanggen Guo, Baisong Ren, Gang An, Wendong Wu, Lexi Xu, Hongxing Bai, and Zhongxi Zhao

A Value-Added Service Strategy for 3G Mobile Network Based on Network Resource Utilization 460
 Lijuan Cao, Yuwei Jia, Chuntao Song, Jie Gao, and Xinzhou Cheng

Big Data and Location Based Dynamic Power Control for Small Cell Networks 468
Yi Li, Yucang Yang, Xingyu Fan, Jun Lu, and Weiwei Chen

Big Data Based Recommendation Scheme in APP Marketing Field 475
Yuwei Jia, Kun Chao, Lijuan Cao, Mingjun Mu, and Xinzhou Cheng

Research on Propagation Prediction Model Localization Over the Tropical Maritime Environment 483
Lianbo Song, Xiaofei Chen, Jian Wang, Jialin Huang, Yafang Xu, Chao Dou, and Cheng Yang

A Novel Architecture and Machine Learning Algorithm for Real Estate 491
Chen Cheng, Xinzhou Cheng, Mingqiang Yuan, Kun Chao, Shiyu Zhou, Jie Gao, Lexi Xu, and Tao Zhang

Multi-index Evaluation Analysis of Region Network Development: A Cluster Empirical Study 500
Haina Ye, Wensheng Li, Jian Guan, Xiaodong Cao, Xinzhou Cheng, Mingqiang Yuan, and Kun Chao

Author Index 509

Wireless Communication Systems

Application of NSGA-II Algorithm to Energy and Spectral Efficiency Trade-off in Massive MIMO Systems with Antenna Selection

Qiaoqiao Zhang^(✉), Xuebin Sun, and Dianjun Chen

Key Lab of Universal Wireless Communications,
Beijing University of Posts and Telecommunications, Beijing, China
qq_0913@sina.com, {sunxuebin,djchen}@bupt.edu.cn

Abstract. In the massive multiple input multiple output (MIMO) system with transmit antenna selection, considering large-scale fading and power consumption, the trade-off between energy efficiency (EE) and spectral efficiency (SE) is important for green communication and meeting the traffic request. However, the diversities of the solution spaces (decision and objective domain) which the existing method obtained are not enough. In this paper, we adopt the non-dominated sorting genetic algorithm version II (NSGA-II) algorithm to study the EE-SE trade-off, which is summarized as a multi-objective optimization (MOO) problem. Simulation results show that compared with the weighted-sum particle swarm optimization (WS-PSO) algorithm, NSGA-II has obvious advantages on the diversity of solution set in both decision and objective domains.

Keywords: Massive MIMO · Energy efficiency · Spectrum efficiency
NSGA-II · Antenna selection

1 Introduction

Massive MIMO has become a key technology of 5G communication to meet the demand of explosive growth of wireless data traffic. With the increase in the number of transmit antennas, the base station (BS) will be configured with a large number of expensive radio frequency (RF) links, together with a dramatic increase of the complexity and hardware costs of the wireless communication system. Antenna selection technique [1] can achieve high energy efficiency as well as high spectral efficiency and reduce the costs. When the active antennas amount exceeds the number of users seriously, it has been proved that simple random transmit antenna selection (TAS) and linear precoding algorithms performs closely to the optimal strategy [2, 3]. As the number of antennas equipped at BS side is quite large, we cannot ignore impact of circuit power consumption, related work is studied in the scene of multi-cell [4, 5]. [6] has studied the transmit power and the number of active antenna selections to optimize the trade-off between EE and SE in massive MIMO, and summarized it as a MOO problem of both EE and SE. It proposed an effective WS-PSO algorithm by transforming it into a single objective optimization (SOO) problem. However, the solution space has poor diversity in both decision and objective domains.

NSGA-II [7] algorithm is a popular multi-objective genetic algorithm with low complexity. We try to apply NSGA-II to this MOO problem directly without transforming it into a SOO one to achieve a diverse solution set of the optimized EE-SE trade-off. NSGA-II utilizes non-dominated sorting and elite reserved strategy to find the Pareto optimal set quickly, calculates and compares crowding distances to obtain a diverse solution set. It has been successfully applied to optimize multi-objective problems such as generation expansion planning (GEP) problem [8] and resource allocation in MIMO-OFDMA systems [9].

In this paper, we mainly investigate energy and spectral efficiency trade-off in the scene of massive MU-MIMO (multi-user MIMO) with antenna selection. We apply NSGA-II algorithm to obtain a diverse Pareto solution set, considering the active antennas amount and transmit power at BS side. Numerical simulation shows that the proposed algorithm is suitable and effective in solving this problem and performs more superior in the result diversities compared with WS-PSO.

2 System Model

2.1 Massive MU-MIMO System

In this paper, a massive MU-MIMO system with K single-antenna users distributed evenly over an annulus single-cell with radius (d_{min}, d_{max}) and the BS configures N ($N \gg K$) antennas is considered. The downlink of the massive MU-MIMO system, whose channel bandwidth is B Hz, and perfect channel state information (CSI) is known to the BS will be discussed.

As random TAS and ZF precoding is adopted, N_t antennas is randomly selected from N ones, and the received signals can be combined in a $K \times 1$ vector as

$$\mathbf{y} = \mathbf{G}_s \mathbf{W}_s \mathbf{P}_s \mathbf{s} + \mathbf{n} \quad (1)$$

where \mathbf{s} is transmit symbols vector, $\mathbf{P}_s = \text{diag}\{\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_K}\}$ is the power allocation matrix and $P_t = \text{tr}(\mathbf{W}_s \mathbf{P}_s^2 \mathbf{W}_s^H)$ is equally allocated to all users, \mathbf{n} is additive white Gaussian noise subject to $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_K)$. \mathbf{G}_s and \mathbf{W}_s are the channel gain and precoding matrix respectively. \mathbf{G}_s is given by $g_{k,n} = h_{k,n} \sqrt{\beta_k}$, where $h_{k,n}$ and $\beta_k(x_k) = \frac{\bar{d}}{\|x_k\|^\kappa}$ represent to the small- and large-scale fading, and x_k refers to the physical location of the k -th user, κ is the path-loss exponent [6]. From [10], we have $\mathbb{E}\{\beta_k^{-1}\} = \frac{d_{max}^{\kappa+2} - d_{min}^{\kappa+2}}{\bar{d}(1 + \frac{\kappa}{2})(d_{max}^2 - d_{min}^2)}$. As ZF precoding is employed, \mathbf{W}_s is determined by

$$\mathbf{W}_s = \mathbf{W}_{ZF} = \mathbf{G}_s^H (\mathbf{G}_s \mathbf{G}_s^H)^{-1} \quad (2)$$

2.2 Power Consumption Model

As a large amount of antennas is equipped at the BS side, the circuit power consumption cannot be ignored. Each active antenna needs a separate expensive RF chain,

and we adopt the conventional circuit power consumption model [10, 11]. The total power consumption P_{Tot} can be represented as a function of the amount of active antennas N_t and the transmit power over a symbol period P_t as:

$$P_{Tot} = \frac{B \cdot P_t}{\eta} + N_t \cdot P_{Bc} + P_{etc}, \quad (3)$$

where $B \cdot P_t$ represents the transmit power, η is the power amplifier efficiency, the other power consumption is P_{etc} and each active antenna will increase P_{Tot} by P_{Bc} .

3 Analysis of EE-SE Relations and Trade-off

3.1 EE-SE Relations with ZF Precoding

The spectral efficiency is obtained by the achievable rate divided by the channel bandwidth. With Jensen's inequality, we have the lower bound of SE from [10] as (4) in which $N_t \geq K + 1$.

$$f_{SE}^{ZF}(N_t, P_t) = K \cdot \log_2 \left(1 + \frac{P_t(N_t - K)}{K\bar{\beta}_{-1}\sigma^2} \right) \quad (4)$$

Where $\bar{\beta}_{-1} = \mathbb{E}\{\beta_k^{-1}\}$ is same for all users, and (4) is tight as $N_t \rightarrow \infty$ [10]. According to Subsect. 2.2, we formulize the corresponding lower bound of EE as

$$f_{EE}^{ZF}(N_t, P_t) = \frac{B \cdot K \cdot \log_2 \left(1 + \frac{P_t(N_t - K)}{K\bar{\beta}_{-1}\sigma^2} \right)}{\frac{B \cdot P_t}{\eta} + N_t \cdot P_{Bc} + P_{etc}}. \quad (5)$$

The spectral efficiency increases with active antennas amount N_t and it is same for the total power consumption P_{Tot} , so that the energy efficiency EE respected to N_t is a concave curve [12]. For a given N_t , the curve of EE-SE is a quasi-concave function, and a trade-off between EE and SE can be calculated [6].

3.2 Multi-Objective Optimization for EE-SE Trade-off

The trade-off between EE and SE is mainly related to an integer variable N_t and a continuous variable P_t . This MOO problem is mathematically expressed as follows [6].

$$\begin{aligned} \mathcal{P}(1) : \max F(N_t, P_t) &= [f_{SE}(N_t, P_t), f_{EE}(N_t, P_t)]^T \\ s.t. (N_t, P_t) &\in \Omega \end{aligned} \quad (6)$$

where $\Omega = \{(N_t, P_t) | K + 1 \leq N_t \leq N, N_t \in \mathbb{Z}, 0 < P_t \leq P_{max}\}$ is the feasible region, and P_t is limited by the maximum transmit power of the base station.

Figure 1 shows EE-SE relationship generated by different and the red line illustrates Pareto front of EE-SE tradeoff problem. Although the Pareto front has been obtained in Fig. 1, we still don't know what the responding optimal solutions (N_t^*, P_t^*) are [6].

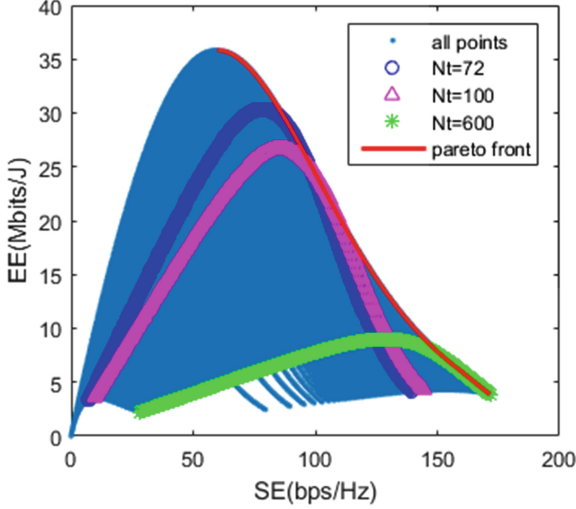


Fig. 1. EE-SE relationship generated by different N_t and the Pareto front

4 NSGA-II for EE-SE Trade-off

As the problem in Subject. 3.2 is a mixed-integer-continuous-variable MOO problem, so that in the operation of genetic operators we need to pay attention to the integer constraint of N_t . The pseudocode of the proposed NSGA-II algorithm for EE-SE trade-off is summarized in Algorithm 1, in which $x_m^i = (N_{t,m}^i, P_{t,m}^i)$ represents the m -th individual in the i -th generation, and non-domination sort is based on the objective $F_m^i = (f_{SE,m}^i, f_{EE,m}^i)$.

Algorithm 1. NSGA-II.

- 1: Initialize: the maximum iterations gen_max , the population size pop_N .
 - 2: Set the iteration counter $i = 0$. For each individual, randomly generate $x_m^i = (N_{t,m}^i, P_{t,m}^i)$, where $P_{t,m}^0 \sim \mathcal{U}(0, P_{t,max})$ and $N_{t,m}^0$ range $[K + 1, N]$.
 - 3: Sort the initialized population by non-dominated sort. Generate the offspring population by binary tournament selection, genetic operators of crossover and mutation.
 - 4: Combine the offspring population with the current generation population as a new population. For the new population, do the following steps.
 - 4.1: Use fast non-dominated sort to divide the population into different fronts as the fitness value, which will be discussed in detail in Algorithm 2.
 - 4.2: Assign the crowding distance to rank the individuals in the same front.
 - 4.3: Update the current generation population by adding better individuals into it until the population size goes to pop_N .
 - 5: Generate offspring population by selection, crossover and mutation.
 - 6: Increment i by one. If $i > gen_max$, exit loop; otherwise, go to Step 4.
-

In fast non-dominated sort, for each individual p , we preserve a set of individuals dominated by p (S_p) and the number of individuals that dominates itself (n_p). Since NSGA-II utilizes the information about S_p and n_p , its computational complexity is much lower than conventional NSGA.

Algorithm 2. Fast Non-Dominated Sort

- 1: For each p in the current population P , initialize $S_p = \phi$ and $n_p = 0$.
 - 2: For each q in P , judge the domination relation between q and p update S_p and n_p .
 - 2.1: if p dominates q , i.e. $f_{SE,p} \geq f_{SE,q}$, $f_{EE,p} \geq f_{EE,q}$ and at least one of the equal sign is not satisfied, then add q to S_p i.e. $S_p = S_p \cup \{q\}$.
 - 2.2: else if q dominates p , increment n_p by one, i.e. $n_p = n_p + 1$.
 - 3: After Step 1 and 2, each p in P has S_p and n_p . If $n_p = 0$, p is included in the first front. We set first front $F_1 = F_1 \cup \{p\}$.
 - 4: Initialize: the front counter $i = 1$.
 - 5: Set the next front $Q = \phi$. For each p in F_i , then go Step 6.
 - 6: For each q in S_p , decrement n_q i.e. $n_q = n_q - 1$.
 - 7: After the loops in Step 5 and 6, if $n_q = 0$, then q belongs to the $(i + 1)$ -th rank, $q_{rank} = i + 1$. Update Q i.e. $Q = Q \cup \{q\}$.
 - 8: $i = i + 1$, update F_i i.e. $F_i = Q$, until $\forall p \in P$ has its own rank. Otherwise, go to Step 5.
-

Then, we assigned crowding distance, which represents the Euclidean distance between an individual and its neighbors in the same front and the distance of boundary values is set to be infinite. In the binary tournament selection, individual i is superior to j only if $i_{rank} < j_{rank}$ or $i_{rank} = j_{rank}$ and $i_d > j_d$ at the same time. After NSGA-II algorithm, we obtain the optimal Pareto solution set of EE-SE trade-off with good diversity

5 Simulation Results

In this section, the performance of the proposed algorithm is compared with WS-PSO algorithm [6] in solving the EE-SE trade-off MOO problem.

We set parameters as follows, users number $K = 10$, bandwidth $B = 20\text{MHz}$, transmit power $P_{tB} \in [10\text{dBm}, 54\text{dBm}]$, active antennas $N_t \in [K + 1, N]$, $pop_N = 400$ and $gen_max = 20$.

The solution set of WS-PSO and NSGA-II with total antennas amount N of 100 and 600 are respectively illustrated in Fig. 2. Compared with the actual Pareto front in Fig. 1, we can recognize that the two algorithms are both effective to obtain the optimal Pareto solution set on the Pareto front. It's obvious that the optimal Pareto solution set obtained by WS-PSO is concentrated at the convex regions. By contrast, the solution of NSGA-II is distributed along the Pareto front more evenly, so that NSGA-II algorithm has superior diversity to WS-PSO in the objective space.

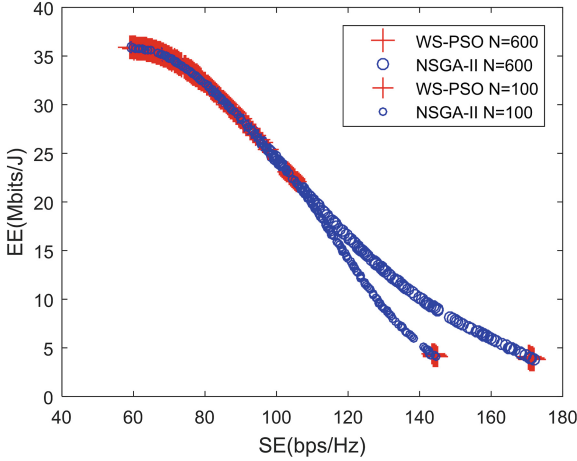


Fig. 2. Comparison of WS-PSO and NSGA-II algorithms in solving EE-SE trade-off

With the given parameters, the probability density in decision-domain (N_t, P_t) for WS-PSO and NSGA-II with total available antennas $N = 600$ is shown in Fig. 3. In the solution set obtained by WS-PSO, P_t is concentrated in the vicinity of $-38dBm/Hz$ and $-19dBm/Hz$, and N_t is concentrated in the vicinity of $K + 1$ and N , where the particles can achieve the largest EE or SE. By contrast, the solution set obtained by NSGA-II is relatively evenly distributed across the feasible region, satisfying the diversity requirement of decision variables in this paper.

From Fig. 4, it can be known that the mean of SE is decreasing with the increase of the generation, and the mean of EE is increasing until the 11th generation. In other words, in massive MU-MIMO systems, it is feasible to sacrifice a certain amount of SE in exchange for an increase in EE with antenna selection to achieve green communication.

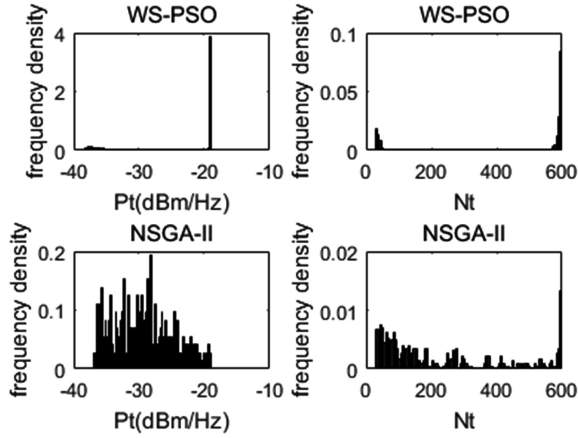


Fig. 3. Probability density in decision-domain of WS-PSO and NSGA-II

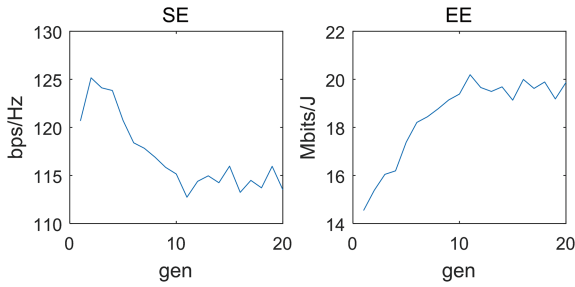


Fig. 4. Mean of SE and EE varies with the number of generation

6 Conclusion

We studied the trade-off between energy and spectral efficiency in massive MU-MIMO systems in this paper. The trade-off issue can be concluded as a MOO problem of active antennas amount (integer variable) and transmit power (continuous variable). Since the Pareto fronts of EE-SE contains both convex and concave sections, we applied NSGA-II algorithm to solve this MOO problem. Numerical simulation shows that NSGA-II algorithm can obtain the Pareto optimal EE-SE trade-off with improved diversities of the solution set in both decision-domain and objective-domain.

References

1. Sanayei, S., Nosratinia, A.: Antenna selection in mimo systems. *IEEE Commun. Mag.* **42**(10), 68–73 (2004)
2. Lee, B.M., Choi, J.H., Bang, J.H., Kang, B.C.: An energy efficient antenna selection for large scale green MIMO systems. In: *IEEE International Symposium on Circuits and Systems*, pp. 950–953. IEEE (2013)
3. Marzetta, T.L.: Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wirel. Commun.* **9**(11), 3590–3600 (2010)
4. Liu, W., Han, S., Yang, C., Sun, C.: Massive MIMO or small cell network: who is more energy efficient?. In: *Wireless Communications and Networking Conference Workshops*, vol. 12, pp. 24–29. IEEE (2013)
5. Li, C., Zhang, J., Letaief, K.B.: Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations. *IEEE Trans. Wirel. Commun.* **13**(5), 2505–2517 (2014)
6. Liu, Z., Du, W., Sun, D.: Energy and spectral efficiency tradeoff for massive MIMO systems with transmit antenna selection. *IEEE Trans. Veh. Technol.* **66**(5), 4453–4457 (2017)
7. Deb, K.: A fast elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**(2), 182–197 (2000)
8. Kannan, S., Baskar, S., Mccalley, J.D., Murugan, P.: Application of NSGA-II algorithm to generation expansion planning. *IEEE Trans. Power Syst.* **24**(1), 454–461 (2009)
9. Zhou, L.Y., Jiang, L., Jiang, D.F.: On the use of improved NSGA-II for multi-objective resource allocation in MIMO-OFDMA systems. In: *Wireless Communication Technology* (2013)
10. Björnson, E., Sanguinetti, L., Hoydis, J., Debbah, M.: Optimal design of energy-efficient multi-user MIMO systems: is massive mimo the answer? *IEEE Trans. Wirel. Commun.* **14**(6), 3059–3075 (2015)
11. Cui, S., Goldsmith, A.J., Bahai, A.: Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks. *IEEE J. Sel. Areas Commun.* **22**(6), 1089–1098 (2004)
12. Ha, D., Lee, K., Kang, J.: Energy efficiency analysis with circuit power consumption in massive MIMO systems. In: *IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, vol. 27, pp. 938–942. IEEE (2013)

A PHY-Based Secret Sharing Scheme in MIMO Systems

Zhuoru Jian^{1,2(✉)}, Hai Huang¹, Xiaojun Jing^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
jzr@bupt.edu.cn

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ School of Engineering and Computer Science, Oakland University,
Rochester, USA

Abstract. A secure secret sharing scheme distributes a secret among a group of participants and only when a sufficient number of shares are combined together can the secret be reconstructed. In this paper, we present a secure PHY-based secret sharing (PSS) scheme exploiting radio channel fading coefficients in multiple-input multiple-output (MIMO) systems. (n, n) threshold secret sharing scheme is considered and there is no need for the third party to distribute keys in PSS scheme. A two-step randomness sharing is designed, which reduces the time overhead significantly compared to one-by-one randomness sharing when the number of participants increases. Furthermore, we derive a power allocation scheme under power constraints based on particle swarm optimization (PSO) algorithm. Numerical results demonstrate the efficiency of our power allocation strategies on secret key rates.

Keywords: PHY-based secret sharing · Two-step randomness sharing
Power allocation

1 Introduction

Secret sharing is proposed to distribute a secret among a group of participants, each of whom is allocated a share of the secret. Individual shares are of no use on their own and only when a specified number of shares are combined together can the secret be read out. The purpose of secret sharing is to prevent secrets from being excessively concentrated as well as to prevent a copy of an encryption key falling into the wrong hands. In one type of secret sharing schemes which is called (t, n) threshold scheme [1], a secret is divided into n shares and every t shareholders out of n can reconstruct the secret while shareholders less than t should not be able to gain any information about the secret.

Security of classic cryptographic-based key generation in secret sharing schemes usually depends on the computational hardness of some mathematical problems. And most key distribution strategies assume that there may exist a safe wireless transmitting

channel, which, however, is highly challenging in wireless systems. Besides, with the development of hardware technology, these schemes may be compromised. To enhance the security of wireless communications, PHY-based key generation has attracted much attention in recent years.

There have been extensive researches on key generation based on the wireless physical layer. Instead of exploiting the computationally secure nature of the public key cryptography and symmetric encryption schemes, wireless physical layer based key generation is information-theoretically secure. The concept of the secret key agreement exploiting wireless channel characteristics was first proposed in [2]. Key generation schemes such as received signal strength (RSS) based schemes and channel state information (CSI) based schemes have been studied in [3–6]. Secret key agreement in two-way wireless relaying systems was studied in [7, 8]. Group secret key generation was discussed in [9], where one-by-one randomness sharing was exploited, i.e., each participant took turns to broadcast training sequences to estimate channel gains. In general, the secret key generation procedures are usually divided into four stages, i.e., channel probing, quantization, information reconciliation and privacy amplification [10]. Channel reciprocity, which provides similar wireless channel characteristics for the two communication parties within coherence time, is exploited in most systems.

In this paper, we propose a PHY-based secret sharing (PSS) scheme based on fading channel characteristics in multiple-input multiple-output (MIMO) systems. The main contributions are summarized as follows.

- A two-step randomness sharing is designed. Different from letting participants take turns to broadcast training sequences, we divide the coherence time into two time slots in which all the participants transmit signals to reduce the time overhead.
- The status of each participant in the group is equal and anyone within the group can encrypt a file using the obtained key while the others need to decrypt the file jointly, thus can prevent secrets falling into the wrong hands.
- A power allocation scheme based on particle swarm optimization (PSO) algorithm is presented to improve the secret key rates under power constraints.

The rest of this paper is organized as follows. Section 2 introduces the system model. Section 3 presents the proposed PSS scheme in details and provides our power allocation algorithm. The numerical results are shown in Sect. 4 and conclusions are given in Sect. 5.

2 System Model

In this paper, we are mainly concerned with (n, n) threshold secret sharing schemes, when all shares are necessary to recover the secret. Without loss of generality, we consider a secret sharing scheme for a group including three participants named Alice, Bob and Charlie. The system model is depicted in Fig. 1, where the three participants wish to generate a set of secret keys through fading wireless channels in the presence of a potentially dishonest one who may act as a passive eavesdropper. In this model, we assume that the fading channel is a slow fading channel, which means the fading channel remains constant within the channel coherence time, thus the reciprocity of the

channel holds. Let $h_{i,j}$ denotes channel between antenna i and j , and it is assumed to be Gaussian random variables, i.e., $h_{i,j} \sim \mathcal{N}(0, \sigma_{i,j}^2)$. We also assume the noise at each antenna is *i.i.d* complex Gaussian random variable with zero mean and variance σ^2 . Moreover, we assume all the participants can communicate with each other and work in a half-duplex system. Each participant is equipped with two antennas and in each time slot in the proposed scheme, one antenna is used to transmit signals while the other is used to receive signals. Considering channel estimation between antenna i and j , the received signals after the training process at antenna i and j can be written as $\mathbf{y}_{i,j} = h_{i,j}\mathbf{s}_i + \mathbf{n}_j$ and $\mathbf{y}_{j,i} = h_{j,i}\mathbf{s}_j + \mathbf{n}_i$, where \mathbf{s}_i and \mathbf{s}_j are probe signals transmitted from antenna i and j , respectively. Then node j and i can obtain the following estimated channel gains

$$\tilde{h}_{i,j} = \frac{\mathbf{s}_i^T}{\|\mathbf{s}_i\|^2} \mathbf{y}_{i,j} = h_{i,j} + \frac{\mathbf{s}_i^T}{TP_i} \mathbf{n}_j, \quad (1)$$

$$\tilde{h}_{j,i} = \frac{\mathbf{s}_j^T}{\|\mathbf{s}_j\|^2} \mathbf{y}_{j,i} = h_{j,i} + \frac{\mathbf{s}_j^T}{TP_j} \mathbf{n}_i, \quad (2)$$

where $\|\mathbf{s}_i\|^2 = TP_i$ and $\|\mathbf{s}_j\|^2 = TP_j$. P_i, P_j are transmission power of antenna i and j , respectively. T is the duration of the signal transmitted by each antenna.

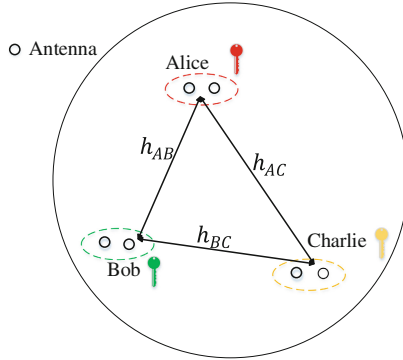


Fig. 1. System model.

After the channel estimation, each antenna pair (i,j) can agree on a nearly uniformly distributed pairwise secret key with arbitrarily small error probability [8], the secret key rate between antenna i and j is $R_{i,j} = I_{i,j}/2T$ where

$$I_{i,j} = I(\tilde{h}_{i,j}, \tilde{h}_{j,i}) = -\frac{1}{2} \log \left(1 - \frac{\sigma_{i,j}^2 TP_i}{\sigma_{i,j}^2 TP_i + \sigma^2} \cdot \frac{\sigma_{j,i}^2 TP_j}{\sigma_{j,i}^2 TP_j + \sigma^2} \right). \quad (3)$$

3 PHY-Based Secret Sharing and Power Allocation

3.1 PSS Scheme

In our scheme, the duration of each channel estimation requires two time slots T_1 and T_2 . It is reasonable to assume $T_1 = T_2 = T/2$, where T represents the channel coherence time. Our scheme includes two major components: two-step point-to-point randomness sharing and secret key generation. Let \mathcal{B} represents the group, i.e., $\mathcal{B} = \{A, B, C\}$, the details can be described as follows.

- **Two-step point-to-point randomness sharing.** As is illustrated in Fig. 2, this phase consists of two time slots. In the first time slot T_1 , each participant $n, n \in \mathcal{B}$ transmits probe signals \mathbf{s}_n with length L using their antenna 1. At the same time, their antenna 2 are used to receive all the signals. Assuming all transmitted sequences are orthogonal to each other, the received signal at each participant can be written as

$$\mathbf{y}_{k_2} = \sum_{n \in \mathcal{B}, n \neq k} h_{k_2, n_1} \mathbf{s}_n + \mathbf{n}_{k_2}, \quad (4)$$

where k denotes the participant k and the subscript 1, 2 denote antenna 1 and 2. \mathbf{n}_{k_2} is additive white Gauss noise at antenna k_2 and $\mathbf{s}_n = [s_{n1}, s_{n2}, \dots, s_{nL}]$ is the probe signal. We assume the three parties are at least half of wavelength away, thus all the channel gains are independent of each other, which indicates that Charlie can get no information about h_{B_2, A_1} , so it is with Alice and Bob.

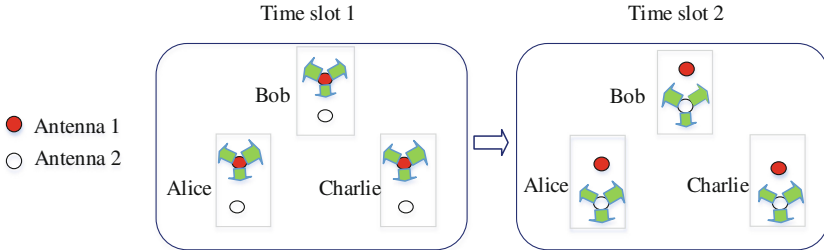


Fig. 2. Two-step randomness sharing in PSS where each participant is equipped with two antennas.

In the second time slot T_2 , each participant n transmits probe signals \mathbf{s}_n using their antenna 2, respectively. At the same time, their antenna 1 are used to receive all the signals. Similar to the first time slot, the received signal at each participant can be written as

$$\mathbf{y}_{k_1} = \sum_{n \in \mathcal{B}, n \neq k} h_{k_1, n_2} \mathbf{s}_n + \mathbf{n}_{k_1}. \quad (5)$$

Due to the channel reciprocity, we have $h_{i,j} = h_{j,i}$. Consequently the shared randomness among Alice, Bob and Charlie can be given as $\tilde{\mathbf{h}}_{BA} = (\tilde{h}_{B_2,A_1}, \tilde{h}_{B_1,A_2})$, $\tilde{\mathbf{h}}_{CA} = (\tilde{h}_{C_2,A_1}, \tilde{h}_{C_1,A_2})$, $\tilde{\mathbf{h}}_{BC} = (\tilde{h}_{B_2,C_1}, \tilde{h}_{B_1,C_2})$.

- **Secret key generation.** Based on the two-step randomness sharing, each participant within the group can obtain the corresponding estimated channel gains, i.e., Alice gets $\tilde{\mathbf{h}}_{BA}$ and $\tilde{\mathbf{h}}_{CA}$, Bob gets $\tilde{\mathbf{h}}_{BC}$ and $\tilde{\mathbf{h}}_{BA}$ and Charlie gets $\tilde{\mathbf{h}}_{CA}$ and $\tilde{\mathbf{h}}_{BC}$. Then according to the pairwise key agreement schemes in [6, 9], the cumulative distribution function (CDF) based quantizer maps $\tilde{\mathbf{h}}_{BA}$, $\tilde{\mathbf{h}}_{CA}$, $\tilde{\mathbf{h}}_{BC}$ to the corresponding Gray code. After information reconciliation and privacy amplification, three pairwise keys $\tilde{\mathbf{h}}_{BA}^\Delta$, $\tilde{\mathbf{h}}_{CA}^\Delta$ and $\tilde{\mathbf{h}}_{BC}^\Delta$ are generated. Next, all the three participants calculate the XOR of the pairwise keys they possess. Thus the final secret keys (K_A, K_B, K_C) can be easily obtained by $K_A = \tilde{\mathbf{h}}_{BA}^\Delta \oplus \tilde{\mathbf{h}}_{CA}^\Delta$, $K_B = \tilde{\mathbf{h}}_{BA}^\Delta \oplus \tilde{\mathbf{h}}_{BC}^\Delta$, $K_C = \tilde{\mathbf{h}}_{CA}^\Delta \oplus \tilde{\mathbf{h}}_{BC}^\Delta$ with the key rate R_{key} .

As the channel between any two participants consists of two sub-channels, concatenating the pairwise secret keys from each sub-channel, the key rates for each channel \mathbf{h}_{BA} , \mathbf{h}_{CA} , \mathbf{h}_{BC} can be written as $R_{BA} = R_{B_2,A_1} + R_{B_1,A_2}$, $R_{CA} = R_{C_2,A_1} + R_{C_1,A_2}$, $R_{BC} = R_{B_2,C_1} + R_{B_1,C_2}$. Since the shorter key is protected by the longer key in XOR operation, the group key rate is limited by the smallest pairwise key rate. Therefore the final keys (K_A, K_B, K_C) in this scheme are obtained with the key rate

$$R_{key} = \min\{R_{BA}, R_{CA}, R_{BC}\}. \quad (6)$$

3.2 Power Allocation

Here we design an algorithm to search for the optimal power allocation for the PSS scheme under total power constraint. The problem we solve can be given as

$$\begin{aligned} & \text{maximize} && R_{key} \\ & \text{s.t.} && P = \sum_{d \in Q} P_d, \\ & && P_d > 0 \end{aligned} \quad (7)$$

where P is the total power of the group and Q is the set of all the antennas within the group. P_d is the transmission power of antenna d . Exploiting the penalty function method, we can rewrite the problem as

$$\begin{aligned} & \text{maximize} && R_{key} + \lambda \left[P, \sum_{d \in Q} P_d \right]^+ \\ & \text{s.t.} && P_d > 0 \end{aligned} \quad (8)$$

where λ is a inertial weight coefficient for the penalty function. $[a, b]^+$ means that $[a, b]^+ = 0$ if $a \geq b$, otherwise $[a, b]^+ = a - b$.

Algorithm 1. PSO algorithm for power allocation

1. Set $n = 0$, and generate the initial position $\{x_{i,d}^n, d \in Q\}$ and velocity $\{v_{i,d}^n, d \in Q\}$ for each particle. $x_{i,d}^n \in [x_{min}, x_{max}]$, $v_{i,d}^n \in [v_{min}, v_{max}]$.
 2. Calculate the fitness F_i^n for each particle. Then we get $\mathbf{P}_i = [x_{i,A_1}^0, x_{i,A_2}^0, \dots, x_{i,C_2}^0]$ and $\mathbf{P}_b = [x_{b,A_1}^0, x_{b,A_2}^0, \dots, x_{b,C_2}^0]$.
 3. Let $n = n + 1$, and update the value of v_i^n and x_i^n according to Eq.(9) and Eq.(10). If $v_{i,d}^n < v_{min}$, set $v_{i,d}^n = v_{min}$, else if $v_{i,d}^n > v_{max}$, set $v_{i,d}^n = v_{max}$. If $x_{i,d}^n < x_{min}$, set $x_{i,d}^n = x_{min}$, else if $x_{i,d}^n > x_{max}$, set $x_{i,d}^n = x_{max}$.
 4. Calculate the new fitness F_i^n . If $F_i^n > F_i^{n-1}$, set $\mathbf{P}_i = [x_{i,A_1}^n, x_{i,A_2}^n, \dots, x_{i,C_2}^n]$. Meanwhile, if $F_i^n > F_b^{n-1}$, set $b = i$, and $\mathbf{P}_b = [x_{b,A_1}^n, x_{b,A_2}^n, \dots, x_{b,C_2}^n]$.
 5. If the number of iterations reaches to the predetermined value, stop and get the optimal power allocation \mathbf{P}_b , otherwise return to step 3.
-

We design the power allocation based on the proposed PSO algorithm to maximize the key rate. PSO is inspired by the birds flocking behaviors and is a population-based optimization method. In Algorithm 1, $\mathbf{x}_i^n = \{x_{i,d}^n, d \in Q\}$ is the position of bird i at time n , and $x_{i,d}^n$ means the transmission power of antenna d . $\mathbf{v}_i^n = \{v_{i,d}^n, d \in Q\}$ is the velocity of bird i . $\mathbf{P}_i = [x_{i,A_1}^n, x_{i,A_2}^n, \dots, x_{i,C_2}^n]$ is the personal best position of bird i and $\mathbf{P}_b = [x_{b,A_1}^n, x_{b,A_2}^n, \dots, x_{b,C_2}^n]$ represents the global best position of the group, where b indicates bird b . Moreover, according to Eq. (8), the fitness F can be given as $F = R_{key} + \lambda [P, \sum_{d \in Q} P_d]^+$. The purpose is to find the best position where the fitness F is maximum through iteration. The position and speed update model is

$$\mathbf{v}_i^{(n+1)} = \mathbf{v}_i^n + c_1 \boldsymbol{\beta}_1 \diamond (\mathbf{P}_i - \mathbf{x}_i^n) + c_2 \boldsymbol{\beta}_2 \diamond (\mathbf{P}_b - \mathbf{x}_i^n), \quad (9)$$

$$\mathbf{x}_i^{(n+1)} = \mathbf{x}_i^n + \omega \mathbf{v}_i^{(n+1)}, \quad (10)$$

where \diamond represents the element-by-element (Hadamard) product operation. ω is the inertial weight coefficient and c_1 and c_2 are the learning factors. We set $\omega = 0.5$, $c_1 = c_2 = 2$. $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors which obey uniform distributions over $[0,1]$. Using Algorithm 1, we update the power allocation and when the number of iterations reaches to a predetermined value, we can get the optimal power allocation.

4 Numerical Results and Analysis

In order to estimate the performance of our proposed scheme, some numerical examples are provided in this section. For simplicity, all additive noise variance are assumed to be 1 (i.e., $\sigma^2 = 1$).

Figure 3(a) compares the secret key rates of our proposed power allocation with the equal power allocation. We assume the variances of each channel gain are set as $\sigma_{A_1,B_2} = 2.7$, $\sigma_{A_2,B_1} = 1.3$, $\sigma_{B_2,C_1} = 1.1$, $\sigma_{B_1,C_2} = 0.09$, $\sigma_{A_1,C_2} = 5.7$, $\sigma_{A_2,C_1} = 4$ and $\sigma_{i,j} = \sigma_{j,i}$. The inertial weight coefficient is set as $\lambda = 4$. It is obvious that our power allocation scheme using PSO algorithm outperforms equal power allocation. We can also see that the secret key rate difference between the two allocations increases with the increased SNR at the low SNR region. However, when the SNR is higher, the difference tends to be a stable value. This phenomenon occurs because the proposed power allocation algorithm tends to assign more power to channels with worse channel quality. When SNR is low, the secret key rate is mainly limited by noise, so with SNR increasing the improvement of the power allocation algorithm can be more significant. However, when SNR is high, the secret key rate is mainly limited by the variances of channel coefficients, thus the improvement of the power allocation algorithm can become stable.

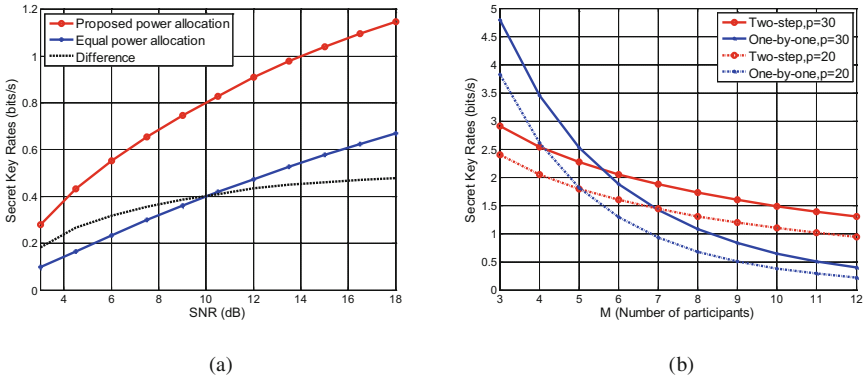


Fig. 3. (a). Comparison of key rates with proposed power allocation and key rates with equal power distribution under total power constraint. (b). Secret key rates using two-step randomness sharing and one-by-one randomness sharing versus the number of participants, the variances of all channel gains are 2.7.

Moreover, Fig. 3(b) shows the secret key rates using two-step randomness sharing and one-by-one randomness sharing as a function of the number of participants (i.e., M) during the coherence time. The variances of all channel gains are 2.7, and the power $P = 20$ or 30. The results indicate that our two-step randomness sharing can reduce the time overhead when the number of participants gets large.

5 Conclusions

In this paper, we have investigated the (n, n) threshold secret sharing scheme based on physical layer in MIMO systems. In our PSS scheme, three legitimate participants within a group employ the two-step randomness sharing scheme to establish keys with

each other exploiting the fading channel coefficients. Once the group secret keys generated, any participant in the group can encrypt confidential message and let the rest of the group work together to restore the message. Moreover, a PSO algorithm based power allocation scheme was derived for total power constraint. Numerical results show that our proposed power allocation is efficient on the secret key rates.

Acknowledgment. Project 61471066 supported by NSFC.

References

1. Lin, C.C., Tsai, W.H.: Secret image sharing with steganography and authentication. *J. Syst. Softw.* **73**(3), 405–414 (2004)
2. Hershey, J.E., Hassan, A.A., Yarlagadda, R.: Unconventional cryptographic keying variable management. *IEEE Trans. Commun.* **43**(1), 3–6 (1995)
3. Liu, H., Yang, J., Wang, Y., Chen, Y.J., Koksai, C.E.: Group secret key generation via received signal strength: Protocols, achievable rates, and implementation. *IEEE Trans. Mob. Comput.* **13**(12), 2820–2835 (2014)
4. Liu, H., Wang, Y., Yang, J., Chen, Y.: Fast and practical secret key extraction by exploiting channel response. In: 2013 Proceedings of the IEEE INFOCOM, pp. 3048–3056 (2013)
5. Wang, Q., Su, H., Ren, K., Kim, K.: Fast and scalable secret key generation exploiting channel phase randomness in wireless networks. In: 2011 Proceedings of the IEEE INFOCOM, pp. 1422–1430 (2011)
6. Chen, K., Natarajan, B.: MIMO-based secret key generation strategies: rate analysis. *Int. J. Mob. Comput. Multimedia Commun.* **6**(3), 22–55 (2014)
7. Shimizu, T., Iwai, H., Sasaoka, H.: Physical-layer secret key agreement in two-way wireless relaying systems. *IEEE Trans. Inf. Forensics Secur.* **6**(3), 650–660 (2011)
8. Zhou, H., Huie, L.M., Lai, L.: Secret key generation in the two-way relay channel with active attackers. *IEEE Trans. Inf. Forensics Secur.* **9**(3), 476–488 (2014)
9. Xu, P., Cumanan, K., Ding, Z., Dai, X., Leung, K.K.: Group secret key generation in wireless networks: algorithms and rate optimization. *IEEE Trans. Inf. Forensics Secur.* **11**(8), 1831–1846 (2016)
10. Zhang, J., Duong, T.Q., Marshall, A., Woods, R.: Key generation from wireless channels: A review. *IEEE Access* **4**, 614–626 (2016)

Long Short-Term Memory Network for Wireless Channel Prediction

Xiaoyun Tong^{1,2,3}✉ and Songlin Sun^{1,2,3}

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
xiaoyun_t@126.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China

³ National Engineering Laboratory for Mobile Network Security,
Beijing University of Posts and Telecommunications, Beijing, China

Abstract. In modern wireless systems, channel prediction is an effective way to overcome the feedback delay of channel state information (CSI). When the receiver performs adaptive transmission based on the feedback CSI, the channel prediction algorithm can reduce the system overhead by predicting the future CSI. In this paper, we provide a long short-term memory (LSTM) network for wireless channel prediction. This method can get a smaller prediction error than other intelligence methods. Experiments show that the LSTM model has a lower normalized mean square error (NMSE) and less running time than support vector machine, artificial neural network, and recurrent neural network prediction approaches.

Keywords: Wireless channel prediction · Channel state information · Long short-term memory network

1 Introduction

With increasing demands of business communications, high communication quality and spectral efficiency become the main objectives and challenges for the future of wireless communication systems. With the increase in the number of base station antennas, feedback overhead increases. For a Time Division Duplexing (TDD) system, which can use channel reciprocity to obtain Channel State Information (CSI), but for a Frequency Division Duplexing (FDD) system, it is difficult to obtain accurate CSI at the base station side, and a large feedback overhead. Channel prediction is an effective way to overcome the feedback delay of CSI. It can predict comparatively accurate CSI to provide reliable data for applying the adaptive transmission technique. Channel prediction methods can be used to predict the coming CSI in advance, so it can reduce the impact of channel feedback delay error.

There are many existing channel prediction methods. The existing methods for channel prediction can be divided into traditional methods and intelligence methods.

The traditional methods include Kalman filtering method, LMS algorithm, particle filters and so on. These methods have some disadvantages as follows.

- Kalman filtering method: it would result in a waste of frequency resources and large amounts of data were needed.
- LMS algorithm: the convergence was slow and the method is not applicable in the fast time varying system.
- Particle filters: there were no pilot symbols, but due to the random variable iteration was used to complete the approximation of the unknown distribution, the computational complexity was further increased.

The intelligence methods include support vector machine (SVM) [1–3], support vector regression (SVR) [4], neural network [5–8], autoregressive (AR) model [9, 10] and so on. Figure 1 shows a block diagram of the intelligence methods for channel prediction. Compared with these intelligence optimization algorithms, the neural network has the advantages of nonlinearity, input-output mapping, adaptability, fault tolerance, high-speed parallelism, and self-learning, which will provide important technical support for wireless communication development.

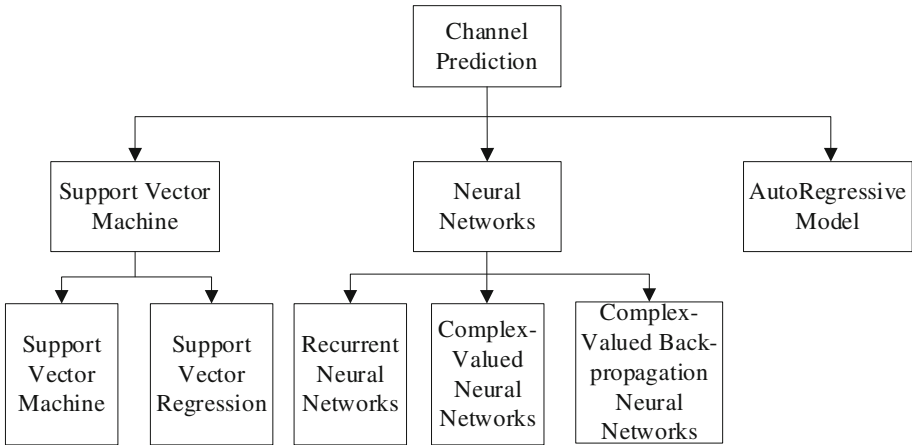


Fig. 1. Channel prediction family tree

LSTM is an excellent variance model of Recurrent Neural Networks (RNN), inheriting the characteristics of most RNN models, and solves the vanishing gradient problem caused by the gradual reduction of gradient back propagation. An LSTM network is a prediction model. It is well suited for dealing with issues that are highly relevant to time series, such as machine translation, dialog generation, encoding, decoding, and so on.

In this paper, we propose an LSTM model to solve wireless channel prediction for real scenarios. We first build a simple input simple output LSTM model which can make full use of pasted CSIs. In addition, we proposed a channel prediction scheme. Finally, the experiment determines the number of training samples and hidden layers.

The normalized mean square error (NMSE) and running time are evaluated under different prediction approaches.

2 Channel Model

In this paper, we study realistic scenes, based on outdoor microcellular channel measurements at 39 GHz for 5th-Generation (5G) channel. The outdoor microcellular measurements were conducted at the research institute of radio wave propagation in Qingdao, China. In this scenario, a line-of-sight (LoS) is measured as shown in Fig. 2. The Base Station (BS) and Mobile Station (MS) are located in a crane and a trolley with heights of 9 m and 1.8 m, respectively. As can be seen from Fig. 2, the BS and the MS are line-of-sight transmission, and they are equipped with one antenna. D_{\min} represents the minimum distance between the BS and the MS. The moving distance of MS is D_t . Also, there are many obstacles in the actual measurement. The antenna parameters are displayed in Table 1.

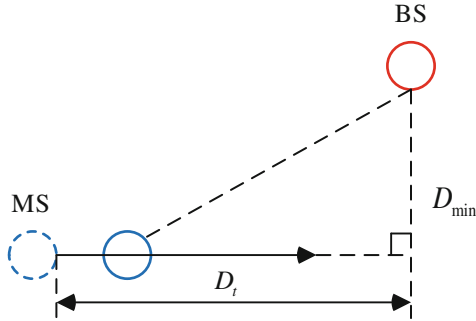


Fig. 2. Outdoor line of sight measurement

Table 1. Antenna parameters.

Frequency	26.5 GHz–40 GHz
Transmit power	1 W
Bandwidth	1 GHz
Antenna gain	24.29 dB–26.97 dB

3 Long Short-Term Memory

LSTM is a special recurrent neural network with the ability to learn long-term dependencies. It is a useful model for predicting. CSI obtained from the feedback channel for FDD wireless communication systems is outdated as a result of processing and feedback delay, and cannot indicate the actual channel state at the time of sending data. In order

to alleviate the problem of outdated CSI, the transmitter predicts CSI at the time of sending data by adopting an LSTM.

LSTM has a chain structure similar to RNN, but the repeating modules are different. There are four instead of a single neural network layer, and they interact in a special way, as shown in Fig. 3.

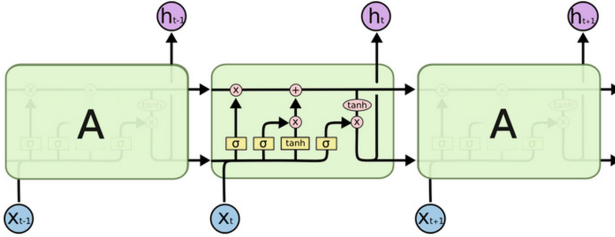


Fig. 3. LSTM chain structure

LSTM contains three layers: an input layer, a cell layer, and an output layer. A Cell consists of three gates (input, forget, output) and a cell unit. Gate uses a *sigmoid* activation function, while input and cell states are usually converted using *tanh*. Figure 4 shows the LSTM cell diagram.

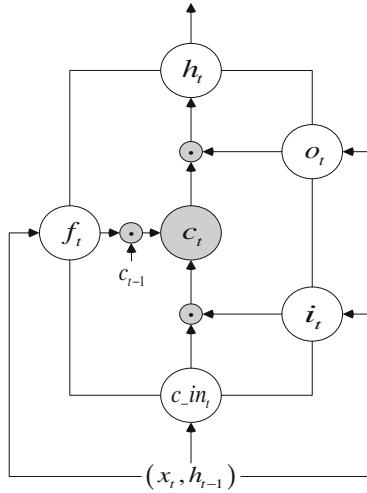


Fig. 4. LSTM cell diagram

The input vector is $(x_t, h_{t-1}) \in C^{M \times 1}$, which is an M-dimensional variable, the output state is $h_t \in C$, a one-dimensional variable. The update of the LSTM cell can be divided into the following steps.

Step 1: Calculating the value of the candidate memory cell at the current time c_in_t , W_{xc} , W_{hc} corresponding to the input data and the last moment LSTM cell output weight, respectively.

$$c_in_t = \tanh (W_{xc}x_t + W_{hc}h_{t-1} + b_c). \quad (1)$$

Step 2: Calculating the value of the input gate i_t . The input gate is used to control the effect of the current data entry on the memory cell status value.

$$i_t = \sigma (W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \quad (2)$$

where σ is usually taken as a logistic *sigmoid* function, the value range (0,1).

Step 3: Calculating the value of the forgotten gate f_t . The forgotten gate is used to control the effect of historical information on the current memory cell status values.

$$f_t = \sigma (W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f). \quad (3)$$

Step 4: Calculating the current time memory unit status value c_t .

$$c_t = f_t \odot c_{t-1} + i_t \odot c_in_t, \quad (4)$$

where \odot represents a point-by-point product.

Step 5: Calculating the output gate o_t , used to control the output of the memory unit status value.

$$o_t = \sigma (W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o). \quad (5)$$

Step 6: Finally, the output of the LSTM cell is given by Eq. (6).

$$h_t = o_t \odot \tanh (c_t). \quad (6)$$

The design of three doors and a separate memory cell allows the LSTM cell to maintain, read, reset and update long distance historical information.

4 Experiment Results and Analysis

At time n , there are M input features for LSTM, and $y(n)$ is the output variable. Thus, input samples are as follows:

$$\begin{cases} x(n) = [h(n), h(n+1), \dots, h(n+M-1)]^T, \\ y(n) = h(n+M) \end{cases}, \quad (7)$$

where $n \in \{1, 2, 3, \dots, N_s\}$, N_s is the total number of samples.

The total samples consist of training data and test data. Training samples are selected from the following proportions: {20%, 40%, 50%, 60%, 80%}. The rest are test data.

In the experiment, the total number of samples and forecast length are set as $N_s = 22528$, $M = 20$, respectively.

The prediction error metrics NMSE is defined as,

$$NMSE = \frac{\sum |h(n + M) - \hat{h}(n + M)|^2}{\sum |h(n + M)|^2}, \tag{8}$$

where $\hat{h}(n + M)$ is the predicted CSI, and $h(n + M)$ is the actual CSI.

The number of hidden layers and training samples in the LSTM are determined by experiment. Training samples are selected as described above. Training set repeated training model 1000 times. Figure 5 shows that under different cell layers, NMSE gradually decreases as the sample increases. As can be seen from Fig. 5, when the number of cell layers is 1, samples ratio is 80%, the LSTM can obtain a smallest NMSE performance. when the samples ratio is 50%, the NMSE get the minimum at $cell = 3$. Therefore, in this experiment, the numbers of hidden layers and training samples are set as 1 and $0.8 \times N_s$, respectively. So we can use the least hidden layer to obtain a lower mean square error, the network is simpler.

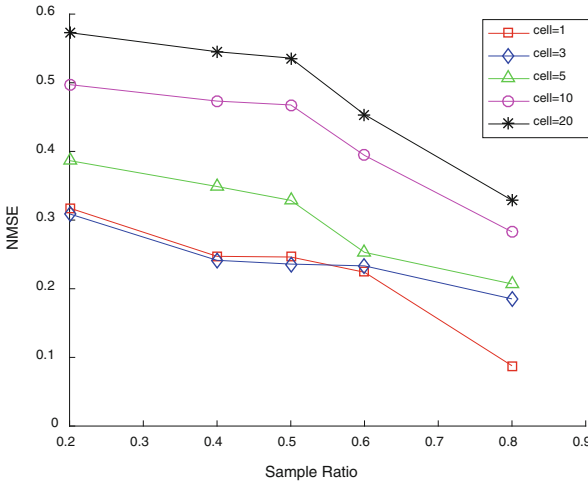


Fig. 5. NMSE versus number of training samples under different hidden layers.

Figure 6 shows the predicted results for LSTM when the samples ratio is 80%. The blue line represents the original CSI, and the red line indicates the predict CSI. From Fig. 6 we can see that the CSI trend is consistent, and the difference is small between original CSI and predicts CSI.

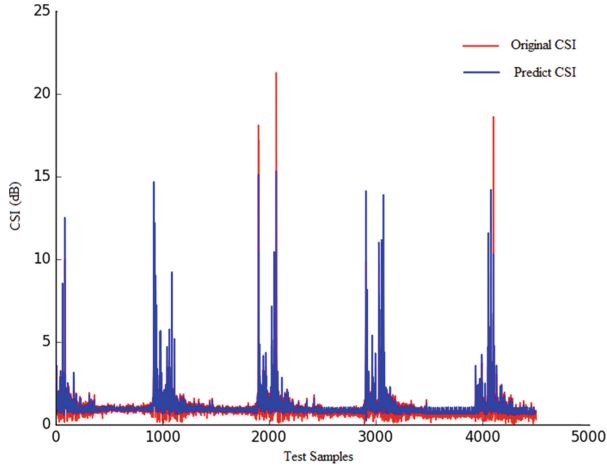


Fig. 6. The predicted results for LSTM when the samples ratio is 80%.

In Table 2, the existing SVM, Artificial Neural Network (ANN), RNN prediction methods are compared in terms of NMSE and running time. Running time is the time for the model to train once. In this experiment, SVM uses the radial basis function kernel function, the hidden neurons of ANN and RNN were set to 30, the hidden layer activation function is set to *sigmoid* and the output layer is *tanh*. The samples ratio is 80%. And the operating system is inter core Central Processing Unit of 3.40 GHz. The program runs 10 times to solve the average of the NMSE. It can be observed that LSTM outperforms SVM, ANN, and RNN and running time is acceptable. The reason is that LSTM has the ability to learn long-term dependencies. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn.

Table 2. Compare SVM, ANN, RNN, with LSTM in NMSE and running time.

Methods	NMSE	Time
SVM	0.9786	175.08 s
ANN	0.4842	12.94 s
RNN	0.4847	0.04 s
LSTM	0.0878	0.82 s

5 Conclusions

Channel prediction provides a more accurate prediction of channel state information, which can improve the system capacity and system performance of the channel. In this paper, we present an LSTM model to apply the channel prediction to the actual scene. The proposed LSTM model has lower NMSE and running time than previous intelligence methods.

Based on the algorithm of the model, compared with traditional methods, the biggest benefit is that it does not need Doppler frequency shift, the physical properties of scatter and number of channel information, it only predicted CSI values on the basis of past CSI, the algorithm is relatively simple, and has a smaller amount of calculation.

References

1. Sankhe, K., Pradhan, C., Kumar, S., et al.: Machine learning based cooperative relay selection in virtual MIMO. *OALib J.* (2015)
2. Chen, X., Meriaux, F., Valentin, S.: Predicting a user's next cell with supervised learning based on channel states. In: *Signal Processing Advances in Wireless Communications*, pp. 36–40. IEEE, (2013)
3. Zhao, X., Hou, C., Wang, Q.: A new SVM-based modeling method of cabin path loss prediction. *Int. J. Antennas Propag.* **2**, 718–720 (2013)
4. Djouama, A., Zochmann, E., Pratschner, S., et al.: Predicting CSI for link adaptation employing support vector regression for channel extrapolation. In: *International ITG Workshop on Smart Antennas* (2016)
5. Ding, T., Hirose, A.: Fading Channel Prediction Based on Combination of Complex-Valued Neural Networks and Chirp Z-Transform. *IEEE Trans. Neural Netw. Learn. Syst.* **25**, 1686–1695 (2014)
6. Li, S., Jiang, M.: Time-varying channel tracking and prediction based on complex-valued back-propagation neural network. *J. Inf. Comput. Sci.* **11**(18), 6579–6589 (2014)
7. Potter, C., Kosbar, K., Panagos, A.: MIMO channel prediction using recurrent neural networks. In: *International Foundation for Telemetry* (2008)
8. Zhao, Y., Gao, H., Beaulieu, N., et al.: Echo state network for fast channel prediction in ricean fading scenarios. *IEEE Commun. Lett.* **21**(3), 672–675 (2016)
9. Dong, C., Dong, Y., Wang, L.: Autoregressive channel prediction model for cognitive radio. In: *International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4 (2009)
10. Eyceoz, T., Duel, A., Hallen, H.: Prediction of fast fading parameters by resolving the interference pattern. In: *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 167–171 (1998)

A Practical Implementation of TD-LTE and GSM Signals Identification via Compressed Sensing

Jianyi Yang^{1,2}, Liang Yin^{1,2(✉)}, Lin Sang², Xin Zhang², Siqing You^{1,3},
and Hongjie Liu¹

¹ Beijing Advance Innovation Center for Future Network, Beijing 100124, China
Siqingyou@139.com, liuhongjie2008@139.com

² School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing 100876, China
{yangjianyi,yinl,sanglin,zhangxin}@bupt.edu.cn

³ School of Information, Beijing Wuzi University, Beijing 101149, China

Abstract. Signal identification is a crucial subject in cognitive radio (CR) systems. In GSM spectrum refarming or spectrum monitoring scenarios, CR is required to identify on-the-air signals like long term evaluation (LTE), global system mobile (GSM). Second-order cyclostationary detection is an identification method robust to noise uncertainty and used widely in spectrum sensing. However, it requires high sampling rate and long processing time. In this paper, we first propose a compressed sensing (CS) based sampling structure to reduce the sampling rate using the second-order cyclostationary features of Time Division-LTE (TD-LTE) and GSM signals. Furthermore, an identification method for TD-LTE and GSM signals based on CS is employed to reduce sensing time. The performance of the method is evaluated by the practical on-the-air-signals measured with a spectrum analyzer. Numerical results show that our method can achieve a high detection probability with a low sampling complexity.

Keywords: Signal identification · Compressed sensing
Second-order cyclostationarity · TD-LTE · GSM

1 Introduction

Cognitive radio (CR) is an effective technology to tackle the spectrum scarcity problem. It allows secondary users to access the spectrum opportunistically without interfering with the primary users [1, 2]. In some applications, in order to adapt to the complex radio environment, CR is required not only to identify the presence but also the type of on-the-air signals [3]. And in spectrum sharing scenario, such as GSM spectrum refarming for LTE co-channel deployments,

it is more important to identify who is occupying the spectrum. And then corresponding decisions can be made to avoid interference with licensed users and improve spectrum sharing efficiency. Thus it is important for CR to identify LTE and GSM signals [4].

Second-order cyclostationary detection is one of the signal identification methods. It identifies different signals by detecting the second-order cyclostationary features related to frame structures, modulation, etc. [5]. Since reference signals (RS) used for channel estimation and acquisition repeat every LTE frame, significant peaks are formed at certain fixed cyclic frequencies. This has been verified in [4, 6] by acquiring the cyclic autocorrelation function (CAF). Second-order cyclostationary detection is robust to noise uncertainty and needs less prior information of the target signal. However, cyclostationarity based signal identification requires large signal samples and high computation complexity.

To reduce the complexity of second-order cyclostationary detection, compressed sensing techniques have been investigated in many studies [7–9]. Specifically, in [7], Tian proposed a compressive sampling scheme to recover a two dimensional cyclic spectrum in wideband sensing. However, this method relies on sparsity in both cyclic domain and frequency domain. In [8, 9], the authors exploited the sparse feature of the CAF, upon which a CS based spectrum sensing method was proposed. This method shortens the sensing time and has a relatively low complexity. However, this method is only verified in the identification of simulated man-made narrow-band signals. So, real world experiments are needed to verify compressed sensing based signal identification using cyclostationary features. Different from the prior work on CS based cyclostationary identification, our paper evaluates the performance of the method by the measured on-the-air data. So our conclusions are more practical and can be used to guide piratical spectrum sharing deployments.

In our paper, we implement CS algorithm using real world TD-LTE and GSM signals acquired with a spectrum analyzer Tektronix 306B. The CAF of TD-LTE and GSM signals are obtained using the measured data and then the second-order cyclostationarity of the measured signals are verified for our proposed CS structure. Our contributions are as follows:

1. Using cyclostationary features, we are the first to propose a low-rate compressed sampling structure for TD-LTE and GSM signals identification.
2. Real-world experiments are implemented to prove our proposed signal identification method. Verification results show that a high right identification probability can be obtained with short observation time and low false alarm probability.

2 Preliminaries of LTE and GSM Signal Identification

2.1 Second-Order Cyclostationarity

Signal identification is a binary hypothesis-testing problem [8]:

$$\begin{aligned} H_0 : \quad & \mathbf{y}(t) = \mathbf{n}(t) \\ H_1 : \quad & \mathbf{y}(t) = \mathbf{x}(t) + \mathbf{n}(t). \end{aligned} \tag{1}$$

Under the hypothesis H_0 , the received signal only contains Gaussian noise $\mathbf{n}(t) \sim \mathcal{N}(0, \sigma_n^2)$, where $\mathcal{N}(0, \sigma_n^2)$ is the Gaussian distribution with mean 0 and variance σ_n^2 . Under the hypothesis H_1 , the received signal is LTE or GSM signal $\mathbf{x}(t)$ plus the additive Gaussian noise $\mathbf{n}(t)$.

Given the second-order cyclostationarity of the random process $x(t)$, the autocorrelation function $R(t, \tau)$ can be expressed by the Fourier series as

$$R(t, \tau) = \sum_{\alpha} c(\alpha, \tau) e^{j2\pi\alpha t}, \tag{2}$$

where $\{\alpha\}$ consists of the integer multiples of the cyclic frequency α_0 . If T_0 is a cycle of $R(t, \tau)$, then $\alpha_0 = \frac{1}{T_0}$. The CAF can be expressed as

$$c(\alpha, \tau) = \frac{1}{T_0} \int_{t=0}^{T_0} R(t, \tau) e^{-j2\pi\alpha t} dt. \tag{3}$$

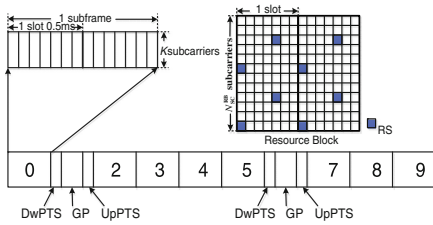


Fig. 1. The frame structure of TD-LTE.

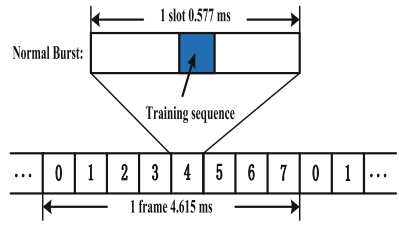


Fig. 2. The frame structure of GSM.

The sampling sequence of the signal $x(t)$ is denoted as $x(n)$. If the observation time is T_{total} and the sampling period is T_s , the length of $x(n)$ is $N = \frac{T_{total}}{T_s}$. $F_s = \frac{1}{T_s}$ is the sampling frequency. The unbiased estimation of $c(\alpha, \tau)$ is

$$\hat{c}(\alpha, \tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) x^* \left(n + \frac{\tau}{T_s} \right) e^{-j2\pi\alpha n T_s}, \tag{4}$$

where τ is the integer multiply of T_s . The second-order samples are $r^\tau(n) = x(n) x^* \left(n + \frac{\tau}{T_s} \right)$. Therefore, the vector consisted of samples of $\hat{c}(\alpha, \tau)$ in the cyclic frequency domain, which is defined as $\mathbf{c}^\tau = \left[\hat{c}(0, \tau), \hat{c}\left(\frac{F_s}{N}, \tau\right), \dots, \hat{c}\left(\frac{(N-1)F_s}{N}, \tau\right) \right]$ is

$$\mathbf{c}^\tau = \frac{1}{N} \mathbf{D}_N \mathbf{r}^\tau, \tag{5}$$

where \mathbf{D}_N is an N dimensional DFT matrix and $\mathbf{r}^\tau = [r^\tau(0), r^\tau(1), \dots, r^\tau(N-1)]^T$ is the vector form of $r^\tau(n)$.

2.2 The Frame Structures of TD-LTE and GSM

The second-order cyclostationarity based signal identification is closely depends on the frame structure of the signal. The frame structure of TD-LTE is shown in Fig. 1 [10]. A TD-LTE frame contains 20 time slots, each with 6 or 7 OFDM symbols. In each time slot, $N_{\text{SC}}^{\text{RB}}$ subcarriers constitute a resource block (RB). The reference signals (RSs) used for cell identification are embedded in the RBs as shown in Fig. 1. Since reference signals are repeated in each time slot, LTE signal exhibits second-order cyclostationarity.

The frame structure of GSM signal is depicted in Fig. 2 [11]. Each GSM frame contains 7 time slots. A time slot might be the normal bursts for data transmission and the control bursts for synchronization, access, etc. In each normal burst, the training sequence used for channel estimation occupy 26 bits. The second-order cyclostationarity of the GSM signal comes from the periodicity of the training sequences.

For LTE signals, the period of reference signals is $T_{0,\text{LTE}} = 0.5$ ms. So the cyclic frequency is $\alpha_{0,\text{LTE}} = \frac{1}{T_{0,\text{LTE}}} = 2$ kHz. And for GSM signals, the period of training sequences is $T_{0,\text{GSM}} = 0.577$ ms. So the cyclic frequency is $\alpha_{0,\text{GSM}} = \frac{1}{T_{0,\text{GSM}}} = 1.733$ kHz. These features are used to identify signals.

3 The Compressed Sensing Based Identification Method for LTE and GSM Signals

In this section, we present an enhanced method to identify LTE and GSM signals as depicted in Fig. 3. First, a novel structure is introduced to achieve sub-Nyquist sampling rate. Then, a low-complexity algorithm based on compressed sensing is used to estimate the cyclic frequencies. Finally, we elaborate the decision rule. The detection diagram is shown in Fig. 3.

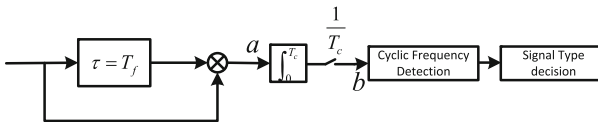


Fig. 3. The proposed detection diagram.

3.1 Signal Sampling

The band of the target signal is usually broad, but the cyclic frequencies may not be very high. As our aim is to identify the signal, the sampling rate can be much lower than the Nyquist rate.

In our proposed structure, the samples are obtained from the second-order signals, which is obtained by multiplying the original signal with the delayed

Algorithm 1. Cyclic frequencies detection via OMP

Input: Second-order signal samples \mathbf{r}_2 ; dictionary matrix \mathbf{Q}

Initialization: $\mathbf{res} = \mathbf{r}_2$; $\mathbf{I} = []$

1: **for** $k = 1 : K$ **do**

2: $i = \operatorname{argmax}(\mathbf{res}^H \mathbf{Q})$ // Find the most relevant item form the dictionary.

3: $\mathbf{I} = [\mathbf{I}, i]$, $\Psi = \mathbf{Q}(:, \mathbf{I})$

4: $\hat{\alpha}_k = (i - 1) \Delta\alpha$ // Calculate the cyclic frequency corresponding to the selected index.

5: $\mathbf{res} = \hat{\mathbf{c}} - \Psi(\Psi^{-1}\Psi)\Psi^{-1}\hat{\mathbf{c}}$ // Remove the influence of the estimated cyclic frequencies.

6: **end for**

Output: Estimated cyclic frequencies $\hat{\alpha}_k (k = 1, 2, \dots, K)$.

signal. The delay is $\tau = T_f$, where T_f is the frame length of the target signal. In this way, the second-order cyclostationarity of the target signal is more significant. The signal at a in Fig. 3 can be represented as

$$r(t) = y(t)y(t + T_f). \quad (6)$$

In order to sample the second-order signal with a low rate, the mean of the second-order signal is calculated every T_c period before sampling. According to the Nyquist sampling theory, the maximum cyclic frequency should be smaller than $\frac{1}{2T_c}$. The signal at b is denoted as

$$r_1(p) = \frac{1}{T_c} \int_{t=(p-1)T_c}^{pT_c} r(t) dt. \quad (7)$$

The vector form of $r_1(p)$ is $\mathbf{r}_1 = [r_1(1), r_1(2), \dots, r_1(N_1)]$, where $N_1 = \frac{T_1}{T_c}$. T_1 is the observation time. The CAF vector can be denoted as $\mathbf{c} = \mathbf{D}_{N_1} \mathbf{r}_1$, whose resolution is $\Delta\alpha = \frac{1}{T_1} = \frac{1}{T_c N_1}$. The observation time should be long enough to obtain a high CAF resolution.

3.2 The Identification Method

As mentioned above, signal identification requires a long observation time, which is generally many frame periods. In fact, a short observation time is enough because CAF is sparse in cyclic domain. This means that the recovery of CAF is a compressed sensing problem. We define a short-time observation \mathbf{r}_2 that is the first N_2 elements of \mathbf{r}_1 . Namely, \mathbf{r}_2 contains the samples within the observation time $T_2 (T_2 \ll T_1)$. If we use the knowledge of \mathbf{r}_2 to recover CAF \mathbf{c} , the compression ratio is $\beta = \frac{N_2}{N_1} = \frac{T_2}{T_1}$. Thus, the recovery problem is

$$\min_{\mathbf{c}} \|\mathbf{c}\|_0, \text{ subject to } \mathbf{r}_2 = \mathbf{Q}\mathbf{c} \quad (8)$$

where $\|\cdot\|_0$ is the l_0 norm to count the number of non-zero entries, and $\mathbf{Q} \in \mathbb{C}^{(N_2 \times N_1)}$ is the first N_2 columns of N_1 order IDFT matrix.

Orthogonal matching pursuit (OMP) is used so solve the recovery problem, which is given in Algorithm 1. K estimated cyclic frequencies are worked out in K loops. K should be larger than the number of the cyclic frequencies. But larger K will lead to larger false alarm probability. In each loop, the index of the column in \mathbf{Q} which has the largest correlation with the residue is selected. The estimated cyclic frequency is calculated according to the index and the expected cyclic frequency resolution $\Delta\alpha = \frac{1}{T_1}$. Then a new residue is obtained by eliminating the contributions of the estimated cyclic frequencies.

The identification needs to decide whether the target signal has the cyclic frequencies of LTE or GSM $\{\alpha_z | \alpha_z = z\alpha_0, (z = 1, 2, \dots, Z)\}$, where α_0 is $\alpha_{0,\text{LTE}}$ or $\alpha_{0,\text{GSM}}$, and Z is the number of cyclic frequencies to be detected. The decision rule is that if an estimated cyclic frequency $\hat{\alpha}_k$ ($k = 1, 2, \dots, K$) satisfies $|\hat{\alpha}_k - \alpha_z| < \eta$, where η is a predefined threshold, the target signal has the cyclic frequency α_z .

The signal type is decided using a voting rule. If the target signal has an cyclic frequency α_z , we have $\Delta_z = 1$. If $\sum_{z=1}^Z \Delta_z \geq V$, the occasion is decided as H_1 . Else, the occasion is decided as H_0 .

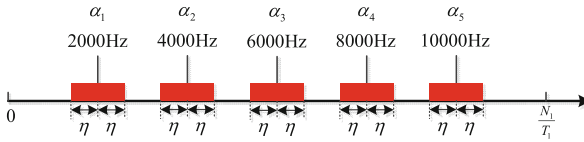


Fig. 4. An example of neighborhoods of cyclic frequencies. $\alpha_0 = 2000$ Hz, $Z = 5$.

The false alarm probability is determined by the two threshold η and V . For the noise case H_0 , the possible positions of the peaks are uniformly distributed in the cyclic domain with the employment of OMP. Considering the decision rule, the false alarm probability is the probability that no less than V estimated cyclic frequencies locate in the neighborhoods $U(\alpha_z, \eta)$, ($z = 1, 2, \dots, Z$), which are the red ranges in Fig. 4. From this figure, we can easily find that the probability that one cyclic frequency locates in $U(\alpha_z, \eta)$ is $P_e = \frac{2Z\eta T_1}{N_1}$. So the false alarm probability is

$$P_f = P(H_1|H_0) = 1 - \sum_{k=0}^{V-1} C_K^k P_e^k (1 - P_e)^{(K-k)}. \quad (9)$$

4 Performance Evaluation by Measured Data

The experiment conditions are given first. The on-the-air signals were measured by the spectrum analyzer RSA306B from tektronix company. Real-world cellular signals are down converted from RF to baseband by RSA306B. And IQ data can then be processed with MATLAB software. We measured TD-LTE signal of

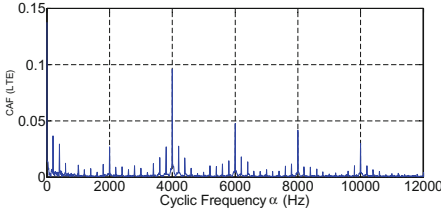


Fig. 5. CAF of TD-LTE with the observation time $T_1 = 100$ ms, the central frequency is 1895 MHz and the sampling rate $F_s = 56$ MHz.

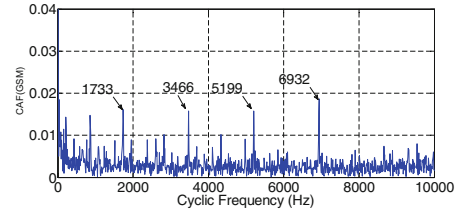


Fig. 6. CAF of GSM with the observation time $T_1 = 100$ ms, the central frequency is 945 MHz and the sampling rate $F_s = 28$ MHz.

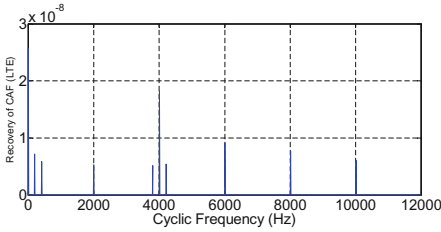


Fig. 7. The recovery of CAF of TD-LTE with the compression ratio $\beta = 0.1$.

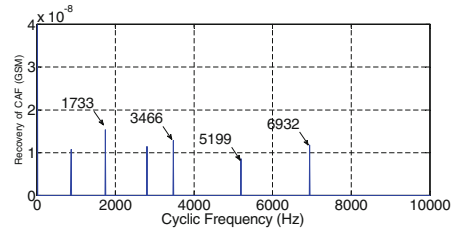


Fig. 8. The recovery of CAF of GSM with the compression ratio $\beta = 0.1$.

Band 33 with the frequency range 1885–1905 MHz and the downlink GSM signal with the frequency range 935–954 MHz. The CAF of the measured signal with length $T_1 = 100$ ms is calculated and shown in Figs. 5 and 6. Significant peaks exist at some integer multiplies of the two cyclic frequencies $\alpha_{0,LTE} = 2$ kHz and $\alpha_{0,GSM} = 1.733$ kHz, respectively.

Then, the measured data is used to simulate the proposed sampling structure and the identification method. Since the cyclic frequencies ($\alpha_z \leq 10$ kHz) is much smaller than the signal bandwidth (20 MHz), the sampling rate can be lower than the Nyquist rate with the proposed sampling structure if the only purpose is identification. In our simulations, the sampling rate is set as $F_s = \frac{1}{T_c} = 56$ kHz, and the compression ratio is set as $\beta = 0.1$, which means the observation time is $T_2 = 10$ ms. The number of iterations of OMP is set as $K = 20$ in order to recover all peaks. The recovery of CAF is shown in Figs. 7 and 8. The positions of the peaks are precisely estimated by OMP algorithm.

In the decision step, the voting threshold is $V = 2$. The false alarm probability is decided by the threshold η . In our simulations, η is set as 10 Hz, 20 Hz and 30 Hz, corresponding to different false alarm probabilities. The number of cyclic frequencies to be detected is set as $Z = 5$ for LTE identification. And for GSM identification, $Z = 4$. The probability of right identification $P_d = P(H_1|H_1)$ versus signal to noise ratio (SNR) is illustrated in Fig. 9. We adjusted the SNR in our simulations by adding emulational noise to the measured data. The compression

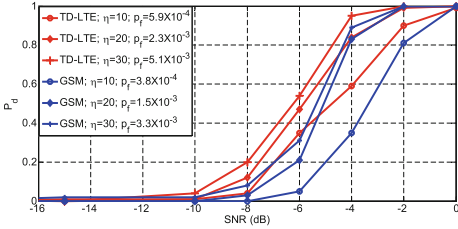


Fig. 9. The probability of correct detection versus SNR with different false alarm probabilities. The compression ratio is 0.1.

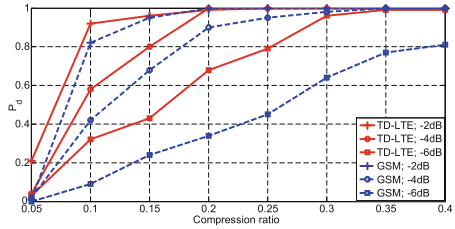


Fig. 10. The probability of correct detection versus compression ratio with different SNR. False alarm probability is 5.9×10^{-4} for TD-LTE and 3.8×10^{-4} for GSM.

ratio is 0.1 in this simulation. As shown in this figure, a high performance is achieved with small false alarm probability.

The influence of the compression ratio on the probability of correct identification is illustrated in Fig. 10. The SNR is set as -2 dB, -4 dB and -6 dB. The threshold η is set as 10 Hz. Namely the false alarm probability is 5.9×10^{-4} for TD-LTE and 3.8×10^{-4} for GSM. From this figure, we get the conclusion that a high performance can be achieved with low compression ratio (short observation time) when the SNR is not very small.

5 Conclusion

In this paper, we focused on the low-complexity identification of the practical LTE and GSM signals. We measured TD-LTE and downlink GSM signals by a spectrum analyzer. Then a low-rate sampling structure and an OMP-based method were employed to identify the signals. The performance of the compressed sensing based method was evaluated by measured data. Numerical results show that a high right identification probability is obtained with short observation time and low false alarm probability. For future work, it would be interesting to identify signals in the condition that diverse signals possibly exist.

References

1. Haykin, S.: Cognitive radio: brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **23**, 201–220 (2005)
2. Zeng, Y., Liang, Y., Hoang, A.T., Zhang, R.: A review on spectrum sensing for cognitive radio: challenges and solutions. *EURASIP J. Adv. Signal Process.* (2010)
3. Dobre, O.A., Inkol, R.: Blind signal identification achievements, trends, and challenges. In: *International Conference on Communications (COMM)*, pp. 349–352 (2012)
4. Al-Habashna, A., Dobre, O.A., Venkatesan, R., Popescu, D.C.: Cyclostationarity-based detection of LTE OFDM signals for cognitive radio systems. In: *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–6 (2011)

5. Like, E., Chakravarthy, V., Ratazzi, P., Wu, Z.: Signal classification in fading channels using cyclic spectral analysis. *EURASIP J. Wireless Commun. Networking*. (2009)
6. Karami, E., Dobre, O.A., Adnani, N.: Identification of GSM and LTE signals using their second-order cyclostationarity. In: *Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1108–1112 (2015)
7. Tian, Z.: Cyclic feature based wideband spectrum sensing using compressive sampling. In: *IEEE International Conference on Communications (ICC)*, pp. 1–5 (2011)
8. Khalaf, Z., Nafkha, A., Palicot, J.: Blind spectrum detector for cognitive radio using compressed sensing. In: *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–5 (2011)
9. Khalaf, Z., Nafkha, A., Palicot, J.: Blind spectrum detector for cognitive radio using compressed sensing and symmetry property of the second order cyclic autocorrelation. In: *International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, pp. 291–296 (2012)
10. TS36.211: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation, 3GPP Std., Rev. 10.0.0, December 2010
11. European Telecommunications Standard Institute (ETSI), Rec. ETSI/GSM 05.02, Multiplexing and Multiple Access on the Radio Path. version 3.5.1, March 1992

Monitoring and Avoidance of Atmospheric Duct on Interference Optimization in TD-LTE System

Ao Shen¹, Yang Zhang², Bao Guo³(✉), Guozhi Wang⁴, Yan Gao⁵, Jixiang Liu⁶, Dayang Liu⁷, Yi Liu⁸, Xiaochun Hu³, and Tao Xie⁹

¹ China Mobile Group Design Institute Corporation, Harbin, China
shenao@cmdi.chinamobile.com

² China Mobile Communication Corporation, Beijing, China

³ China Mobile Communication Corporation, Shanxi Branch, Taiyuan, China
13934150003@139.com

⁴ China Mobile Communication Corporation, Zhejiang Branch,
Hangzhou, China

⁵ China Mobile Communication Corporation, Henan Branch, Zhengzhou, China

⁶ China Mobile Communication Corporation, Liaoning Branch,
Shenyang, China

⁷ China Mobile Communication Corporation, Guangdong Branch,
Guangzhou, China

⁸ China Mobile Communication Corporation, Shandong Branch, Jinan, China

⁹ China Mobile Communication Corporation, Chongqing Branch,
Chongqing, China

Abstract. The interference of mobile communication network optimization system is mainly for the interference of intra-system, interference of inter-system and short distance external interference. The interference localization of the mobile communication network, especially the Time Division Duplexing (TDD) system, becomes very complicated. This paper mainly expounds the basis of the generation of atmospheric duct, and how to monitor the interference of atmospheric duct of the Time Division Long Term Evolution (TD-LTE) system. On the basis of this, the paper proposes a centralized scheme to avoid the interference of atmospheric duct, focusing on the optimization measures of avoiding and reducing the interference of atmospheric duct through parameter optimization.

Keywords: Atmospheric duct · Special sub-frame
Time division duplexing · Power control

1 Introduction

Under certain meteorological conditions, the propagation of electromagnetic waves in the atmosphere, especially in the near surface layer, is bent due to the influence of atmospheric refraction. When the curvature over the surface of the earth curvature, electromagnetic wave will be part of the ground settlement by the thin atmosphere with

a certain thickness, like electromagnetic wave in metal duct propagation, the atmosphere is called atmospheric duct. Propagation of electromagnetic waves in atmospheric duct is called atmospheric duct propagation. It can make the electromagnetic wave in the smaller attenuation along atmospheric duct to a very far place.

Atmospheric duct is a layer different from the standard atmosphere. Atmospheric duct has a negative high index of refraction gradient. Atmospheric duct layer can make the part of the regional electromagnetic wave propagate to become the original electromagnetic blind, also can make the part of the electromagnetic wave with a low attenuation spread far away. And the electromagnetic wave signal maintains strong electric field strength, the problem of interference in mobile communication system is very complicated.

Because of the change of air temperature, pressure and humidity in the wireless environment, the characteristic parameters of the duct can be changed, and the interference analysis will be more difficult.

2 Interference Characteristics and Monitoring of Atmospheric Duct

2.1 Interference Characteristics of Atmospheric Duct

Atmospheric duct mainly occurs in summer and autumn, with the temperature around 20° . It usually starts at night and disappears around eight o'clock the next morning. Atmospheric duct is more easier to spread in the open environment, therefore, the plain area is greatly affected, and the interference of rural area is stronger than the interference of urban area [1, 2]. The atmospheric duct propagation diagram is shown in Fig. 1.

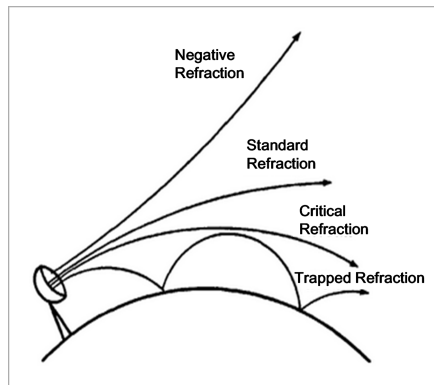


Fig. 1. The atmospheric duct propagation

Because downlink frequency of the Global System for Mobile Communication (GSM) network is isolated, and is using the same frequency, so atmospheric duct has no effect on it; Time Division-Synchronous Code Division Multiple Access (TD-SCDMA)

and TD-LTE are both using TDD standard, but the degree of interference of TD-SCDMA network, which is using the same frequency is less than that of TD-LTE network; Early rural TD-LTE base stations are few, atmospheric duct has little effect on users. With the expansion of network scale and the growth of the number of users, the degree of interference of rural area is increasing, and users are complaining more and more.

TD-LTE system downlink frequency consistent in atmospheric duct environment propagation loss, propagation distance, the uplink and downlink Guard Period (GP) protection network mainstream distance is about 64.2 km, the same frequency reuse interference can be completely isolated wireless base station, but according to the temporal characteristics of atmospheric duct environment analysis, spread even more than 200 km GP, before the apparent lack of distance protection [3].

As the user’s sensitivity to speech is much greater than that to data service, the user’s complaint may suddenly increase.

2.2 Atmospheric Duct Interference Monitoring

Atmosphere duck interference, disturbed cell interference level rise obviously, in part, interference mean even higher than -100 dbm, while the wireless connection rate and dropping rate significantly worse, seriously affect the user perception. TD-LTE network performance indicators are mainly manifested in the following aspects [4, 5].

1. wireless access rate

Daily mean remained at more than 99.5%, a large area of the worst interference time deteriorated to 88.22%. The success rate of Evolved Radio Access Bearer (ERAB) and Radio Resource Control (RRC) has serious deterioration, the main reason is that the wireless layer caused the failure of RRC and ERAB.

2. wireless dropping rate

The daily mean wireless dropping rate of less than 0.5%, a large area of the worst interference period deteriorated to 1.93%, as shown in Fig. 2, the red column diagram of the wireless drop rate (using the coordinate axis). Among them, the wireless layer is mainly due to the uplink error rate is too high (more than 3%), the Automatic Repeat-reQuest (ARQ) retransmission rate also has serious deterioration.

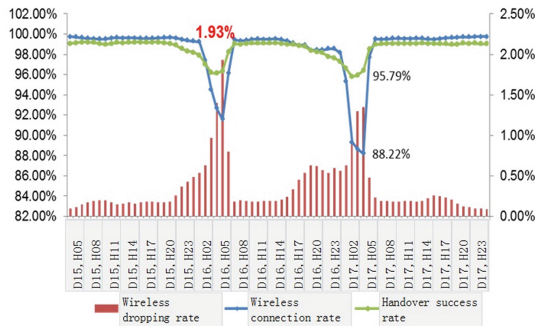


Fig. 2. Influence of atmospheric duct on TD-LTE network index

3. the success rate of switching

Handover success rate of the daily average of about 99%, a large area of the worst interference time deteriorated to 96%.

The influence of atmospheric duct interference on TD-LTE is collected on the network index, as shown in Fig. 2.

It can be seen from the above analysis, the atmospheric duct wireless access, TD-LTE network with switching have caused a serious impact, from the random access process, causes of the decline in performance index of atmospheric duct interference network is mainly MSG3 not successfully sent or received by. It can be considered to mitigate the impact of interference on the performance of the network by improving the MSG3 reception reliability.

2.3 Accurate Identification of Atmospheric Duct

Based on the causes of interference and temporal frequency domain characteristics, the interference of waveguides is precisely and quickly identified. Because the interference of atmospheric duct can not be eliminated fundamentally, the main objective of this thesis is to minimize the impact on users [6].

1. Time domain characteristics

The time domain shows obvious power ramp characteristics, and the GP begins to be disturbed until the last symbol interference of the uplink subframe decreases gradually.

2. Frequency domain characteristics

In the frequency domain, the 6 PRB perturbations are more serious, the 100 PRB are disturbed, and the 6 PRB perturbations in the middle are more serious, mainly the primary and secondary synchronization, the MIB, etc., as shown in Fig. 3.

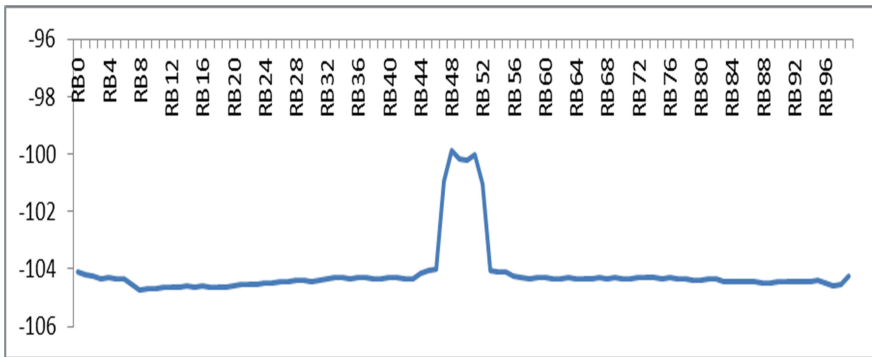


Fig. 3. Schematic diagram of frequency domain characteristics of atmospheric waveguide interference.

3. Spatial characteristics

With the increase of the antenna height, both the perturbation and the perturbation increase, and the smaller the dip angle of the base station antenna is, the larger the interference level is.

3 Atmospheric Duck Interference Mitigation Measures

3.1 Time Domain Dimension Mitigation Measures

In time domain, the mitigation measures of atmospheric duck are mainly realized by changing the sub frame ratio of special sub frame, and the special sub frame of TD-LTE is composed of 3 special slots [7, 8].

Downlink pilot time slot (DwPTS) is a special sub frame in the downlink time slot. The minimum duration of DwPTS is 3 Orthogonal Frequency Division Multiplexing (OFDM) symbols, and the main synchronous channel Physical Shared Channel (PSCH) is placed on third symbols of DwPTS. Because DwPTS in TD-SCDMA for downlink synchronization, PSCH is also used for downlink synchronization, so put the DwPTS in the PSCH; at the same time, on the third symbol is conducive to the configuration of GP. In addition to the synchronization symbol resource, the other DwPTS resources can transmit data, pilot and downlink control signals.

Uplink pilot time slot (UpPTS) is a special sub frame in the uplink time slot, used to access as a random access channel (RACH) occupy 6 resource block (RB) bandwidth, the bandwidth of the system in the case of sufficient UpPTS can be assigned to more than one Physical Random Access Channel (PRACH) channel. In addition, the allocation of RACH after the remaining resources can be transmitted pilot (Sounding and demodulation pilot). Physical Uplink Control Channel (PUCCH) is not transferred in this time slot, and the control overhead can be minimized.

Period Guard time slot (GP) is a special protection time of TDD system. The main reason is that there is a downlink uplink bidirectional transmission delay between the base station and the User Equipment (UE) when the uplink conversion. When the GP is determined, the bidirectional transmission delay of the cell coverage radius is determined. In order to overcome the special symbol interference Inter Symbol Interference (ISI) in OFDM system, LTE introduces the cyclic prefix CP (Prefix Cyclic). CP length configuration: the length of CP is related to the coverage radius, the general situation of the Normal CP can meet the requirements of the wider coverage and other small radius of the scene can be configured to expand the CP (CP Extended). The greater the length of the CP configuration means the greater the system overhead. Third Generation Partnership Project (3GPP) agreement provides for the 9 GP configurations, as shown in Table 1, the unit is the symbol. GP maximum time is 10 OFDM symbols (CP Normal), the value of which is mainly affected by cell coverage radius.

Currently TD-LTE network F band use special sub frame mode is configured 6, that is, 9:3:2, protection slot for the 3 symbols, a symbol for the 71.4us, corresponding to 21.4 km, so the protection of the 3 symbols of the range of 64.3 km. If it is a special sub frame is configured 5, that is, 3:9:2, the protection of time slot for the 9 symbols, the corresponding protection distance of 192.6 km.

Table 1. TD-LTE special sub frame configuration

Special sub frame	Normal CP			Extended CP		
	DwPTS	GP	UpPTS	DwPTS	GP	UpPTS
0	3	10	1	3	8	1
1	9	4	1	8	3	1
2	10	3	1	9	2	1
3	11	2	1	10	1	1
4	12	1	1	3	7	2
5	3	9	2	8	2	2
6	9	3	2	9	1	2
7	10	2	2	–	–	–
8	11	1	2	–	–	–

In addition, can also use the downlink sub frame part shutdown occurred during the atmospheric interference duck in each frame 0, closed 5, 1, 6, and RB downlink sub frame non symbol must be sent, only broadcast channel, the primary synchronization channel and secondary synchronization channel. But the downlink rate will drop significantly, the laboratory test to close down the downlink part of the sub frame, the terminal will have a serious deterioration of the downstream level, do not consider the deployment.

3.2 Dimensional Mitigation Measures in Frequency Domain

1. F+D/F+E double network

In the atmospheric duck interference serious regional scale deployment F + D/F + E double network, when the buck interference, through parameter optimization, the user will migrate from the F area to a clean D/E cell, improve the user's perception. Ministry approved the E band can only be used indoors, there is a risk of policy. Follow up with the increase of network capacity, large-scale use of D band, there will be the same frequency, the anti-interference effect is weakened [9].

2. F band 10 M same frequency network

The F band of 30 M bandwidth is divided into 3 10 M band, a different use of different frequency points. There is more interference from outside the duck region, the F frequency bandwidth of 30 M were used, the F band same frequency network limited effect: from the current network data, during the 10 M F band duck interference area (frequency 38544) average level of interference with 20 M cell (frequency 38400), indicating that F band 30 M bandwidth are effect. 10 M different frequency network will lead to cell capacity, half rate in half, 4G users continued rapid growth, unfavorable to the user perception, competing for the underdog.

3.3 Spatial Dimension Mitigation Measures

1. control of high power station

Township rural areas F band high station (40 m above), the unified specification of its power is not more than 80 W. During the period of interference, the province's unity to the F band high station temporary further down power, interference regression power recovery. Implementation results show that the intensity of interference has a decrease of 1–2 dB. According to the assessment, the high station power is not more than 80 W, the surrounding traffic and indicators have no effect.

2. control antenna elevation angle

Antenna height of 45 m above the angle of not less than 10° , 40 to 45 m is not less than 8° , 30–40 m according to whether the interference area to adjust to 8° , 30 m below the temporarily not adjust. Implementation results show that the interference intensity decreased by 5–10 dB. Measured single point covering distance is reduced by about 10–15%, and if it is replaced by remote electrically tunable antenna, the inclination angle of the lower pressure can be disturbed during the interference, and the inclination angle of the interference is dissipated, and the coverage and interference response can be considered [10].

3. control antenna height

Antenna height is not more than 40 m, the current tower company built a height of 40 m, to meet the requirements; for the stock of the station, considering the coverage of competing factors, only antenna height above 55 m to consider the platform. Implementation results show that the intensity of interference can be decreased by 3–4 dB. Coverage radius is reduced by about 5%, which can be solved by adjusting the inclination angle.

4 Optimization Measures to Avoid Atmospheric Duct

By increasing the GP protection time slot to increase the protection distance, reduce the atmospheric duct interference scheme and the scheme of F band 10 M same frequency network will affect the overall throughput, affect the user perception. If we can achieve the monitoring of atmospheric duct in time, we can deploy the optimization measures such as uplink access, power control, to avoid the impact of atmospheric duct interference.

4.1 Access Channel Parameter Optimization

PRACH power control purpose is to ensure the success rate of eNodeB in the premise of random access, UE in order to minimize the power of the pilot, to reduce the interference of the adjacent area and make UE power. The PRACH transmit power calculation formula for each UE is as follows:

$$P_{PRACH} = \min\{P_{CMAX}, P_{O_pre} + PL + \Delta_{preamble} + (N_{pre} - 1) \cdot \Delta_{step}\} \quad (1)$$

Among them, P_{CMAX} is the maximum transmission power of UE;

$P(O_{pre})$ indicates that the Evolved Node B (eNodeB) is expected to be the target power level when the PRACH leading scheme is 0. Setting initial value through parameter PreambInitRcvTargetPwr.

PL estimates for the downlink path loss value of UE, through the Reference Signal Receiving Power (RSRP) measurement and the cell reference signal transmit power;

The $\Delta_{preamble}$ preamble said that the current configuration of the preamble format based on preamble format power offset value between 0;

N_{pre} indicates the number of times that the UE sends a leader, involving a parameter of PreambleTransMax, which cannot exceed the maximum number of leading times;

The Δ_{step} said the leading power rising step by setting PwrRampingStep parameters.

Through the network performance index of TD-LTE monitor atmospheric disturbance, receiving target power through the parameters of the PreambInitRcvTargetPwr uplift leading initial value, recommended by the current network around -104 dBm uplift to -94 dBm; to increase a maximum number of transmission by setting parameters PreambleTransMax, the net 8 times increased to 10; by setting the parameter PwrRampingStep to increase the power of PRACH climb step increased from 2 dB to 4 dB network.

4.2 Uplink Control and Optimization of Shared Channel Parameters

The purpose of Physical Uplink Shared Channel (PUSCH) power control is to reduce the interference and improve the cell throughput, and to ensure the rate of cell edge users. The formula for calculating the transmission power of UE in the PUSCH channel is:

$$P_{PUSCH}(i) = \min\{P_{MAX}, 10 \log_{10}(M_{PUSCH}(i)) + P_{o_PUSCH}(j) + \alpha(j) \cdot PL + \Delta_{TF}(i) + f(i)\} \quad (2)$$

Among them, (i) represents the first (i) uplink sub frame; P_{MAX} for the system through the RRC signaling to the UE configuration of the maximum transmission power;

$M_{PUSCH}(i)$ represents the number of RB resources allocated by the UE in the PUSCH channel on the I sub frame;

$P_{o_PUSCH}(j)$ is the power benchmark 1.5 static setting value, the value of J is 0, 1 or 2, PUSCH transmission $j = 0$ corresponding to the semi static scheduling authorization or retransmission; PUSCH transmission $j = 1$ corresponding to the dynamic scheduling authorization or retransmission; $J = 2$ corresponds to the random access response PUSCH transfer authorization.

α is indicated as the compensation amount of the path loss, and the downstream path loss is estimated by PASS LOSS (PL) for the UE.

After monitoring the interference of atmospheric duct, the initial PUSCH power is set up to ensure the successful access of cell edge users, and to maintain the business. By setting the nominal P0 value of PUSCH and the path loss factor to resist the interference of atmospheric duct. PUSCH nominal P0 value from the current network of -87 dBm adjusted to -93 dBm (recommended value), the path loss compensation factor from the current network of 0.8 adjusted to 1 (recommended value).

Accordingly, for the uplink control channel PUCCH, the parameter values need to be adjusted for the nominal P0 value of PUCCH, adjusted to -95 dBm \sim -98 dBm (recommended value) by the -100 dBm \sim -105 dBm of the current network.

4.3 Optimize the Implementation Effect

The optimization measures of time domain, frequency domain and airspace are carried out in a region seriously disturbed by atmospheric waveguide.

In the time domain, breaking through the TDD specification, a special subframe ratio is proposed, which is automatic continuous station adjustment and downlink subframe partial turn off.

In the frequency domain, the power control parameters and double layer network parameters are adjusted to improve the user access capability during the interference period.

In the space domain, the use of remote electrically controlled antennas to cope with interference from the waveguide, taking into account interference and coverage.

Through the implementation of the measures, the province atmospheric interference waveguide frequency from 19 times to 7 times and no mass complaints, waveguide interference level period significantly decreased, cell interference accounted for before the implementation of the 31% to 10.6%, and the effect is very obvious.

5 Conclusion


At present, in the strict sense of atmospheric duct interference in TD-LTE network, it is a kind of long distance system interference, which is the same frequency interference caused by uplink and downlink of TDD system. Adjust the special sub frame slot ratio and F band 10 M same frequency network can effectively alleviate the interference of atmospheric duct. In addition, through the network monitoring of atmospheric interference disease through the uplink access channel, control channel and shared channel power control optimization can also be carried out to avoid the interference of atmospheric duct.

References

1. IEEE Standard for Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks Specific Requirements Part 15.4: Wireless Medium Access Control (Mac) and Physical Layer (Phy) Specifications for Low-Rate Wireless Personal Area Networks (Lr-Wpans), IEEE Std 802.15.4-2003, p. I-670 (2003)

2. Supplement to IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks- Specific Requirements- Part 11: Wireless Lan Medium Access Control (Mac) and Physical Layer (Phy) Specifications: Higherspeed Physical Layer Extension in The 2.4 Ghz Band, IEEE Std 802.11b-1999, p. 1-90 (2000)
3. Pollin, S., Tan, I., Hodge, B., et al.: Harmful coexistence between 802.15.4 and 802.11: a measurement-based study. In: IEEE 3rd International Conference On Cognitive Radio Oriented Wireless Networks and Communications (Crowncom), pp. 1–6, May 2008
4. Liang, C.J.M., Priyantha, N.B., et al.: Surviving wi-fi interference in low power zigbee networks. In: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys), pp. 309–322 (2010)
5. Kim, Y., Shin, H., Cha, H.: Y-Mac: an energy efficient multi-channel mac protocol for dense wireless sensor networks. In: Proceedings of the 7th International Conference on Information Processing in Sensor Networks, pp. 53–63, April 2008
6. Dutta, P., Dawson-Haggerty, S., Chen, Y., et al.: A-Mac: a versatile and efficient receiver initiated link layer for low-power wireless. *ACM Trans. Sens. Netw.* **8**(4), 30 (2012)
7. Luo, T., Motani, M., Srinivasan, V.: Cooperative asynchronous multichannel mac: design, analysis, and implementation. *IEEE Trans. Mob. Comput.* **8**(3), 338–352 (2009)
8. Xu, W., Trappe, W., Zhang, Y., et al.: The feasibility of launching and detecting jamming attacks in wireless networks. In: Proceedings of The 6th ACM International Symposium On Mobile Ad Hoc Networking And Computing (Mobihoc), pp. 46–57 (2005)
9. Borms, J., Steenhaut, K., Lemmens, B.: Low-overhead dynamic multi-channel mac for wireless sensor networks. In: *Wireless Sensor Networks. LNCS*, vol. 5970, pp. 81–96 (2010)
10. Angrisani, L., Bertocco, M., Fortin, D., et al.: Experimental study of coexistence issues between IEEE 802.11b and IEEE 802.15.4 wireless networks. *IEEE Trans. Instrum. Meas.* **57**(8), 1514–1523 (2008)

Load Balancing and Interference Management in Heterogeneous Networks

Rihan Wu^{1,2,3}, Songlin Sun^{1,2,3}, and Yasir Ullah¹

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
1052287742@qq.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China

³ National Engineering Laboratory for Mobile Network Security,
Beijing University of Posts and Telecommunications, Beijing, China

Abstract. Load balancing and interference management influence the performance of heterogeneous network (HetNet). Cell range expansion (CRE) scheme has been introduced to achieve better load balance by adding a positive bias to the small-cell BSs. Enhanced inter cell interference coordination (eICIC) is recognized as a solution that reduces cross-tier interference by setting the almost blanking subframes (ABS). In this paper, we divide pico area into pico inner part (the normal coverage area of pico) and pico CRE part (bias area assigned to pico). We propose that the macro only transmits data during non-ABS subframes. The pico CRE part transmits data to its users in ABS subframes, and pico inner part transmits over all subframes. The users within different base stations have different interference patterns, and therefore we compute it independently. Then we derive the optimal ABS value by optimizing the system throughput equation. After performing the extensive simulation, we get the proper CRE value. And the proposed ABS scheme has shown remarkable results.

Keywords: Inter cell interference · Load balancing · Heterogeneous network

1 Introduction

To solve the boosting increase of the data demand, many solutions such as massive multiple-input multiple-output (MIMO) [1], millimeter wave, heterogeneous network (HetNet) have been proposed. Low power base station, such as micro, pico or femto cells are deployed into conventional macro cell systems in [2, 3]. In [4, 5], small cells can expand the macro coverage by deploying them into cell edge areas, and it can be deployed into high traffic demand areas to offload users. For a given HetNet topology, a cross tier interference is also an important issue. If macros and picos share the same frequency, the users within pico cells, especially within pico cell range expansion (CRE) area, suffer severe cross-tier interference from macro cells. To guarantee an acceptable quality of service (QoS) for users within picos, enhanced inter cell interference

coordination (eICIC) scheme was introduced. In [6], eICIC can significantly mitigate the inter-cell interference (macro to pico cell). Despite it may decrease the system throughput, but it can mitigate interference to improve the QoS for users. In [7], the authors propose an algorithm to derive the optimal value of eICIC parameters based on data of actual propagation data and traffic load. The results have shown that the method can obtain system gain by a real RF plan using joint optimization design of ABS and UE-association based on actual network data. The author has explored the system performance with different eICIC parameters in different simulation scenarios. The authors in [8] formulate the problem of choosing a proper user association method and ABS value jointly for sum weighted logarithmic utility maximization while maintaining proportional fairness of users. In [9], the authors use stochastic geometry tool to provide the probabilistic characterization of the downlink coverage and single user throughput of the mobile user that in the randomly located situation, and the system can offload users to small cells.

The main contributions of this paper can summarize as follows. We put forward a method to find the optimal ABS value in HetNet. The goal is to maximize the sum throughput of the system and mitigate inter-cell interference simultaneously. Then, simulation has been performed for the proposed ratio of ABS subframe in different environments (different CRE values) to get better CRE value while achieving max throughput.

The rest of this paper is organized as follows. Section 2 presents the system model and formulates the system interference. Section 3 derives the optimal ABS proportion. Finally, Sect. 4 gets a simulation results and concluding remarks are given in Sect. 5.

2 System Model

We consider a two-tier HetNet system that consists of a number of macro BSs and corresponding picos. As shown in Fig. 1, the picos are deployed into the cellular to offload users from macro base stations. In such a deployment environment, users in the pico coverage area may receive interference from other picos and strong cross-tier interference from the macro BSs, especially the users within the CRE. If the interference

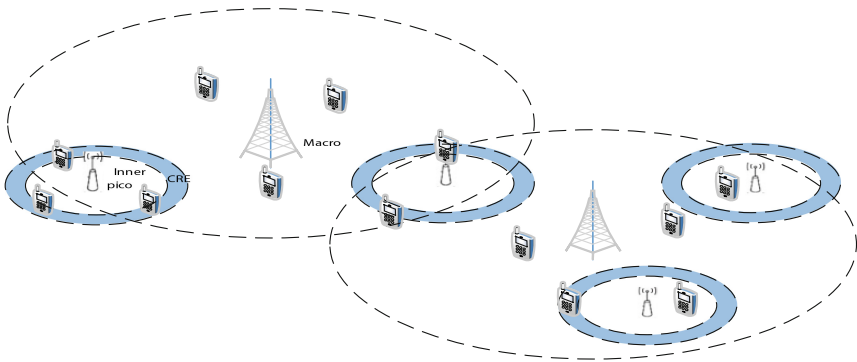


Fig. 1. Two-tier heterogeneous network with cell range expansion.

is not properly controlled, the network performance may be limited severely. However, the proportion of ABS is also very important. If the proportion is too large, it will affect the transmission of the macro base station. If too small, it did not play its usefulness. Moreover, a reasonable proportion of distribution can improve system performance, so as the value of CRE.

The users can be divided into MBSU (users within macro base station), PINU (users within the pico cell inner part) and PCREU (users within pico CRE). Depending on the different base station they connected to.

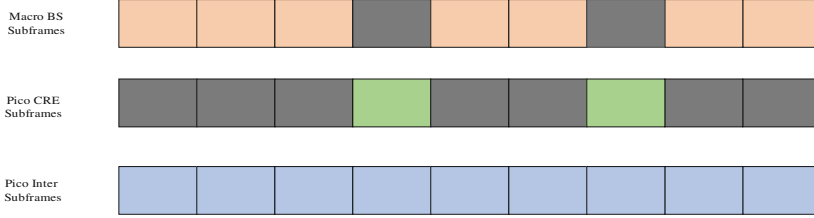


Fig. 2. Illustration of proposed frame structure with ABS subframes.

As seen from Fig. 2, the ABS presents by black subframe, and it means the BS doesn't transmit data during it. In addition to ABS, we can call the normal subframes NABS. Macro BS only transmit data during NABS and PINU transmits data during all subframes. PCREU transmit only during ABS subframes. Based on this observation, we consider the following policy. MBSU suffer interference from the rest of macros and all picos. PINU suffer interference from the rest of pico stations and all Macro stations. PCREU only transmits data during ABS subframes, and all macros keep silent. The PCREU only interfered by pico base stations.

According to the above description, we can list the interference functions of different category of users. In Eq. (1), the I_{ub} illustrates the interference of user u connected to station b . The sets of macro and pico stations are M and P . Then, $m(k)$ presents one of the stations belongs to $M(P)$. $P_m(P_k)$ presents the transmit power of macro station (pico station), and $G_{um}(G_{uk})$ means the channel pathloss of user u connected to macro base station m (pico station k). The I_{ub} can be written as

$$I_{ub} = \begin{cases} \sum_{m \in M, m \neq b} P_m G_{um} + \sum_{k \in P} P_k G_{uk}, & \text{macro BS User} \\ \sum_{m \in M} P_m G_{um} + \sum_{k \in P, k \neq b} P_k G_{uk}, & \text{pico inner part User} \\ \sum_{k \in P, k \neq b} P_k G_{uk}, & \text{pico CRE User} \end{cases} \quad (1)$$

In this system, the combination of the eICIC and CRE divide users into three parts and compute interference independently. It can mitigate the inter-tier interference and increase the spectral reuse simultaneously. So selecting the proper ABS proportion is very important and will be shown in the next section.

3 Problem Solution

The signal to interference plus noise ratio $SINR_{ub}$ of user u within BS b can be written as equation

$$SINR_{ub} = \frac{P_b h_{ub}}{I_{ub} + N_0}, \quad (2)$$

where P_b denotes the transmit power of base station (Macro, pico inner, pico CRE). N_0 denotes the noise spectral power density. h_{ub} denotes the channel gain from base station b to the user u , according to Eq. (2), the rate of user u within BS b can be written as

$$r_{ub} = W \log_2(1 + SINR_{ub}), \quad (3)$$

then proposed value of $A_n(\alpha)$ is formulated in Eq. (4), the proportion of ABS is α and NABS is $1 - \alpha$, $A_n(\alpha)$ is as follow

$$A_n(\alpha) = \begin{cases} 1 - \alpha, & \text{NABS} \\ \alpha, & \text{ABS} \end{cases}. \quad (4)$$

From (3) and (4), we can get the throughput equation of the whole system easily. It is shown as

$$R = \sum_{b \in B} \sum_{u \in U} A_n(\alpha) r_{ub}, \quad (5)$$

and assume that the system contains B base stations. B consist of n macro BSs and $B-n$ pico BSs. And the amount of users in the whole system is U . The max value of system throughput can be written as

$$\max_{\alpha} \sum_{b=1}^B \sum_{u=1}^U A_n(\alpha) r_{ub}, \quad (6)$$

and it is the main expression that we want to optimize. The value range of ABS is $0 \leq \alpha \leq 1$, then add it as a constraint of the (6). The problem becomes a joint optimization problem that can show as

$$\begin{aligned} & \max_{\alpha} \sum_{b=1}^B \sum_{u=1}^U A_n(\alpha) r_{ub}, \\ & s.t. 0 \leq \alpha \leq 1 \end{aligned} \quad (7)$$

and we make a little change that written (7) as the product of dependent variable of α and constant part without α . It is shown as

$$Q(\alpha, r_{ub}) = \sum_{b=1}^B \sum_{u=1}^U A_n(\alpha) f(r_{ub}). \quad (8)$$

We derived the two-order-partial derivative of function Q . Also, the process of calculating is shown as

$$\frac{\partial^2 Q(\alpha, r_{ub})}{\partial \alpha^2} = \sum_{b=1}^B \sum_{u=1}^U r_{ub} \frac{\frac{\partial^2 A_n(\alpha)}{\partial \alpha^2} A_n(\alpha) + r_{ub} \left(\frac{\delta A_n(\alpha)}{\delta \alpha}\right)^2}{A_n^2(\alpha)} = \sum_{b=1}^B \sum_{u=1}^U \frac{-r_{ub} \left(\frac{\delta A_n(\alpha)}{\delta \alpha}\right)^2}{A_n^2(\alpha)}. \quad (9)$$

After calculating the two-order-partial derivative, we can find the result is less than zero. So that the above optimization problem is a concave function, moreover the constraint is a linear function. According to above points, we can see that formula (7) is a convex optimization problem. Then we calculate the partial derivative of Q to α and let the result equal to zero. It can be express in

$$\sum_{u=1}^U \frac{u_1}{1 - \alpha^\zeta} + \sum_{b=1}^{b_m} \sum_{u=1}^U \frac{r_{ub}}{1 - \alpha^\zeta} = \sum_{b=b_m+1}^B \sum_{u=1}^U \frac{r_{ub}}{\alpha^\zeta}, \quad (10)$$

finally, we can get the unconstrained optimal expression of ABS proportion α^ζ as

$$\alpha^\zeta = \frac{u_{CREU}}{U}. \quad (11)$$

The approximate value of α^ζ is the ratio of the number of PCREU to all cellular users.

4 Performance Evaluation

We consider a two-tier HetNet with same frequency deployment of 3 macros and 18 p. There are three-sector in a macro BS, and each sector consists of 2 p. The ratio of user distributed in pico inner area and pico CRE area is 2:1. In each cellular of macro, there are 40 users connected with macro BSs and 18 users with pico. The users uniformly distributed in each eNB. The pathloss model and other parameters are displayed in Table 1.

Table 1. Summary of simulation parameters.

Cell deployment	3 macro cells (18 p cells)
Transmit power	Macro BS: 46 dBm; pico BS:30 dBm
Carrier frequency	2 GHz
Bandwidth	20 MHz
Pathloss model	128.15 + 37.6log ₁₀ (d)dB (macro) 136.74 + 39.2log ₁₀ (d)dB (pico)
Thermal noise	-174 dBm/Hz
Antenna gain	Macro(12 dBi), pico(8 dBi), user(0 dBi)

System throughputs with different CRE values are displayed in Fig. 3. With CRE, more users can be offloaded to small cells like picos that have weaker signal strength but more available resources. By assigning different bias (CRE value) to picos, we can get different system throughputs. Through observation, we can find that when CRE = 3 dB, the CDF (cumulative distribution function) of system throughput reaches the maximum. Furthermore, higher CRE is not always better.

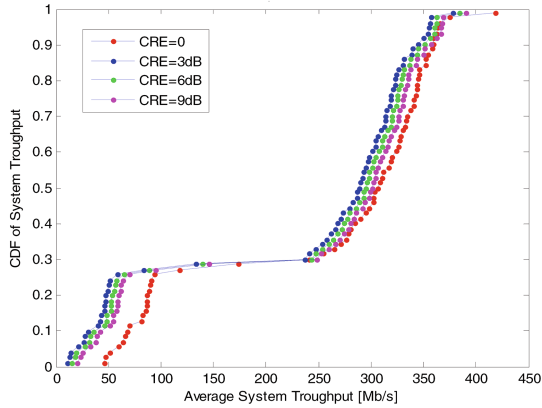


Fig. 3. CDF of system throughput with different CRE values.

An appropriate value can achieve the best system performance. In Fig. 4, we analyze the system throughput influenced by different ABS proportion. The users connected to pico CRE areas are 18, and the whole system has 174 users, approximately, this ratio

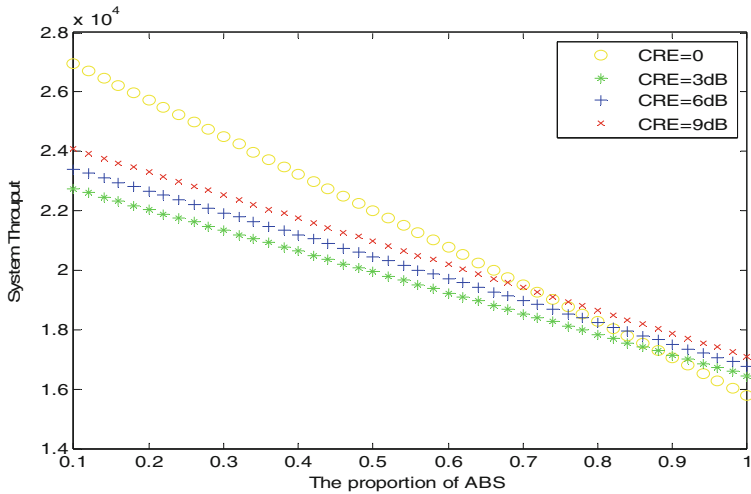


Fig. 4. System Throughput with different proportion of ABS and CRE values.

equals 0.1. When the ABS proportion is 0.1 (the rate that we derived from Eq. (11)), system throughput is maximized. The proposed method proved to be reasonable.

5 Conclusions

In this paper, we get the optimal ABS allocation value that approximately equal to the ratio of the amount of the users within pico CRE area to whole system users by optimizing the system throughput in two-tier HetNet. We conducted simulations to evaluate the performance of the proposed algorithm. The results demonstrate that the proposed method and proper CRE value can maximize the system throughput and the cross-tier interference can be reduced simultaneously. The HetNet comprises of more users can be investigated in future works.

Acknowledgement. This work thanks to the project 61471066 supported by NSFC.

References

1. Chen, N., Rong, B., Zhang, X., Kadoch, M.: Scalable and flexible massive MIMO precoding for 5G H-CRAN. *IEEE Wirel. Commun. Mag.* **24**(1), 46–52 (2017)
2. Damnjanovic, A., et al.: A survey on 3GPP heterogeneous networks. *IEEE Wirel. Commun.* **18**(3), 10–21 (2011)
3. Sun, S., Gong, L., Rong, B., Lu, K.: An intelligent SDN framework for 5G heterogeneous networks. *IEEE Commun. Mag.* **53**(11), 142–147 (2015)
4. Wang, Y., Soret, B., Pedersen, K.I.: Sensitivity study of optimal eICIC configurations in different heterogeneous network scenarios. In: 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, pp. 6792–6796 (2012)
5. Liu, C., Li, M., Hanly, S.V., Whiting, P.: Joint downlink user association and interference management in two-tier hetnets with dynamic resource partitioning. *IEEE Trans. Veh. Technol.* **66**(2), 1365–1378 (2017)
6. Zhou, H., Ji, Y., Wang, X., Yamada, S.: eICIC configuration algorithm with service scalability in heterogeneous cellular networks. *IEEE/ACM Trans. Netw.* **25**(1), 520–535 (2017)
7. Deb, S., Monogioudis, P., Miernik, J., Seymour, J.P.: Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets. *IEEE/ACM Trans. Netw.* **22**(1), 137–150 (2014)
8. Fooladivanda, D., Rosenberg, C.: Joint resource allocation and user association for heterogeneous wireless cellular networks. *IEEE Trans. Wireless Commun.* **12**(1), 248–257 (2013)
9. Wang, H., Zhou, X., Reed, M.C.: Coverage and throughput analysis with a non-uniform small cell deployment. *IEEE Trans. Wirel. Commun.* **13**(4), 2047–2059 (2014)

QoE Evaluation Model for Wireless Video Network Business

Tao Tian^{1(✉)}, Songlin Sun¹, Chenwei Wang², and Xinzhou Cheng³

¹ Beijing University of Posts and Telecommunications, Beijing 100876, China
{tiantao, slsun}@bupt.edu.cn

² DOCOMO Innovations, Inc., Palo Alto, CA 94304, USA
chen.516@gmail.com

³ China Unicom Network Technology Research Institute, Beijing 10048, China
chengxzh@dimpt.com

Abstract. We propose a multi-index evaluation model of quality of experience (QoE) for wireless video, which combines the network quality of service, quality of content, quality of interaction and quality of terminal. In particular, we use the fuzzy analytic hierarchy process (FAHP) to analyze these influencing factors and establish the evaluation system. In addition, we verify the influence of each indicator to QoE for wireless video in network transmission via simulations and subjective evaluation values. The experimental results show that our proposed method is highly correlated with the subjective evaluation, and thus it can more efficiently reveal QoE for the wireless video in the network transmission.

Keywords: Wireless video · Quality of experience
Fuzzy analytic hierarchy process · Quality evaluation

1 Introduction

1.1 Wireless Video Service

The past decade has witnessed the prevalence of wireless video service, due to the development of mobile communication and mobile Internet technologies. In particular, transmitting video to users through the mobile network and mobile terminals becomes increasingly attractive and vital. One requirement of video transmission is to enable data size of transmission to be larger than that of texts and voices. Besides, a new type of video communication is desired, such as live online or on-demand, which require computers or mobile devices to access the video information from a wireless network.

The wireless video transmission mainly relies on mobile streaming media technologies, which provides stable and smooth multimedia programs such as video and audio, and can be easily scalable. For example, streaming transmission technologies can first segment and compress the entire video content into a series of packets, and then transmit them sequentially. At the user side, users do not have to download the entire media file to the local device and can play video (on the terminal device) only

after a very short period. Thus, the video stream files can be downloaded and played nearly at the same time, which can significantly shorten users' waiting time, and also can better release the cache.

The wireless video service is actually a complete system, which includes three parts. The first is content producers, such as wireless video service operators, video creators, websites, portals; the second is network service providers, such as mobile network operators, mobile Internet service providers; the third is terminal equipment, including a variety of playback software and equipment. In practice, system integrators or network infrastructure vendors need to guarantee the connectivity among the above three parts.

1.2 Quality of Experience (QoE)

In the wireless video service ecosystem, the QoE should be considered in each of the three parts. In ITU Telecommunication Standardization Sector, QoE is defined as the degree of delight or annoyance of the user's application. It originates from the fulfillment of the user's expectations with respect to the utility and/or enjoyment of the application. In practice, QoE is evaluated by users, since it describes users' feelings, and each user might have individual intuition on the same service or business. Thus, if we want to better measure the QoE, features of the users should be also considered, such as the users' age, gender, knowledge level, and prior experience.

Other than the users' features, we also need to consider the specific business or service. When evaluating the business-specific QoE, the composition of the business needs to be understood. Similarly, for wireless video, terminal devices and applications need to be considered as well.

The evaluation of QoE depends on a certain objective environment. For the same business, different environments would result in different QoE. For studying the QoE of wireless video service, it is necessary to analyze if the user is in a noisy place and if the whole environment is comfortable.

So far, we have introduced three core components: QoE, user, service, and environment.

1.3 Evaluation Method of QoE for Wireless Video

The QoE evaluation methods for wireless video can be broadly classified into subjective and objective. For the subjective, the object of interest is determined by human subjective consciousness, which can reflect QoE more exactly. The subjective methods mainly include Mean Opinion Score (MOS), Double Stimulus Continuous Quality Scale (DSCQS) and Single Stimulus Continuous Quality Scale (SSCQS). In contrast, the objective method is based on comparing the output to the input, and it includes Full Reference (FR), No Reference (NR) and Reduce Reference (RR).

1.4 Research Status of QoE

Since QoE is a relatively comprehensive and innovative topic, a vast amount of research works have been investigated. We summarize them as follows.

In [1], the authors studied the influence of network, business, and terminal on QoE. In addition, they analyzed the elements of QoE, and the influence factors of network data service on QoE are summarized as delay, jitter, and loss of information. In [2], evaluation methods of QoE were introduced for the QoE evaluation system of statistics, psychology, artificial intelligence, etc. In particular, the Analytic Hierarchy Process (AHP) has a certain instructive effect, which we will also consider in this paper.

Recall that QoE is related to the user's multidimensional perceptual experience. As we mentioned, it involves interdisciplinary studies of information science, psychology, and mathematics. Moreover, different businesses also need corresponding QoE evaluation methods. Regarding the characteristics of QoE models, there are several difficulties, summarized as follows.

- (1) There are very few QoE models of wireless video. In addition, the adaptability and pertinence of wireless video service are not strong enough, due to the lack of analysis of the factors that affect the wireless video.
- (2) Although some studies had established certain QoE models, they only identified a hierarchical relationship and did not quantify the QoE, because there is no analytical solution to calculate QoE.
- (3) The existing QoE models lack the necessary verification.

The rest of this paper is organized as follows. Section 2 describes the system model. Section 3 presents the comprehensive evaluation system of QoE for wireless video. The simulations and numerical results are provided in Sect. 4, and the conclusions of this paper are provided in Sect. 5.

2 System Model

2.1 Fuzzy Analytic Hierarchy Process

Since the transmission quality of the wireless video is affected by many factors, it is difficult to grasp the influence factors on video perception quality. This paper uses fuzzy analytic hierarchy process (FAHP) to calculate the weight value of various factors on the QoE effect on wireless video.

FAHP [3] is essentially a decision-making way, which decomposes the problem into several parts, divides these factors into the orderly hierarchical structure according to the important relationship among them. FAHP is used to establish the evaluation system, and verify the impact of each index by the network simulation implementation and the subjective evaluation.

The core of FAHP is to use 0.1–0.9 scale method to accurately reflect the relative importance of any two factors with respect to specific criteria.

The basic steps for solving problems with FAHP are as follows.

Step 1: Establish prior evaluation matrix $\mathbf{R} = (r_{ij})_{n \times n}$.

The prior evaluation matrix is the matrix of the relative importance weight established by the comparison between the elements in each layer and the upper elements.

As

$$r_{ij} + r_{ji} = 1, 0 \leq r_{ij} \leq 1, (i = 1, 2, \dots, n; j = 1, 2, \dots, n;) \tag{1}$$

For

$$R = (r_{ij})_{n \times n}, \tag{2}$$

If for

$$\forall i, j, k, r_{ij} = r_{ik} - r_{jk} + 0.5 \tag{3}$$

Then prior evaluation matrix \mathbf{R} is known as fuzzy consistent evaluation matrix. In order to quantitatively describe the relative importance of any 2 indexes to a criterion, the nine scale method of Table 1 can be used to give the quantity scale [4].

Table 1. The number of scales.

Scale	Illustrations
0.5	Both elements are equally important
0.6	One element is a little more important than the other
0.7	One element is more important than the other
0.8	One element is much more important than the other
0.9	One element is extremely more important than the other

Step 2: Construct fuzzy consistent judgment matrix. For

$$r_i = \sum_{k=1}^n r_{ik}, i = 1, 2, \dots, n \tag{4}$$

Utilizing consistency, transforming

$$r_{ij} = r_{ik} - r_{jk} + 0.5 \tag{5}$$

To

$$r_{ij} = (r_i - r_j) / 2n + 0.5 \tag{6}$$

Then the fuzzy consistent judgment matrix can be constructed by a prior judgment matrix.

Step 3: Calculate the weight of each index according to matrix R. Assume the weight values of elements a_1, a_2, \dots, a_n are $\omega_1, \omega_2, \dots, \omega_n$, then

$$\omega_i = \frac{\sum_{j=1}^n r_{ij} + \frac{n}{2} - 1}{n(n-1)}, i = 1, 2, \dots, n \quad (7)$$

Step 4: Convert weight values between the layers into the comprehensive weight of the relative target.

2.2 Analyzing Influencing Factors of QoE for Wireless Video

Although QoE is the user's subjective feelings, it is still very important to establish a suitable model and quantify the assessment, which can help service providers and operators quantitatively analyze and predict the acceptability of business. Obviously, each user has different evaluation methods for different services or products, meanwhile, each user's individual needs are also different.

This paper makes a reasonable generalization of the current research and puts forward the quantitative evaluation model of the QoE for wireless video. Assuming that QoE consists of four influencing levels with a certain first level of weight, and each influencing level is composed of several influencing factors with a certain second level of weight.

The four influencing levels [5] of QoE for wireless video are quality of service (QoS), quality of content (QoC), quality of user interaction (QoI), quality of terminal (QoT).

QoS refers to the quality of network services during wireless video transmission, such as packet loss rate, delay, jitter, network bandwidth, etc. QoC refers to the quality of wireless video content, such as video synchronization, block effect, image fuzziness, video coding rate in wireless video. QoI reflects the parallax comfort, motion perception comfort, waiting time, and switching time of users in requesting wireless video. QoT refers to the impact of server terminals and receiving terminals on wireless video quality, such as screen resolution, memory size, CPU performance, operating system, etc. As shown in Fig. 1.

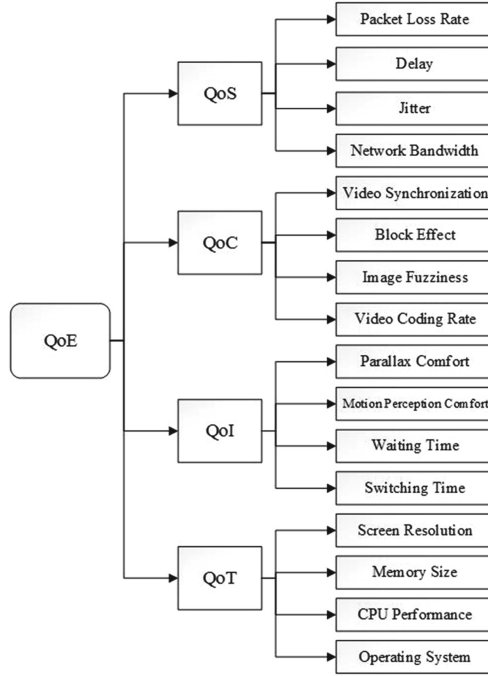


Fig. 1. Influencing factors of QoE for wireless video

3 Comprehensive Evaluation System of QoE for Wireless Video

When using FAHP to solve the decision problem, we should first divide the problem into layers and establish the hierarchy model, which can be divided into 3 layers: target layer, criterion layer, and decision-making level. The elements of the upper layer act as a criterion for the next layer. In the comprehensive evaluation of QoE for wireless video, this paper establishes the QoE hierarchical structure according to Fig. 1, in which the target layer is the intended goal of the problem, that is, the evaluation result of the QoE for wireless video [6]. The criterion layer contains the middle links involved in achieving the goals, which are four factors that affect QoE for wireless video. The decision-making level is a variety of specific influence factors to realize the goal.

QoS, QoC, QoI, QoT, are called key quality indicators (KQI). The evaluation value of the QoE for network wireless video can be expressed as [7]

$$QoE = \omega_1 \cdot QoS + \omega_2 \cdot QoC + \omega_3 \cdot QoI + \omega_4 \cdot QoT \quad (8)$$

ω_1 is weight value. According to the actual test results, the weights of each KQI can be adjusted. Similarly, each KQI depends on different parameters, which is called the key parameter indicators (KPI), KPI can be measured or calculated, then

$$QoX = \sum_{i=1}^n \omega_i KPI_i, X = S, C, I, T \tag{9}$$

It can be seen from the formulas (8) and (9) that QoE can be obtained by establishing the relationship between KPI, KQI and QoE,

$$QoE = \sum_{i=1}^m \sum_{j=1}^n \omega_{ij} KPI_{ij} \tag{10}$$

Since each KPI is obtained by the actual test, which can be converted within the range of (0,1) using the polyline dimensionless method. According to the different factors influencing wireless video quality in Fig. 1, FAHP is utilized to analyze various factors, giving different weights to different factors, finally evaluating the transmission quality of wireless video and calculating estimation value of QoE [8].

This paper calculates KPI of QoC, selecting wireless video synchronization, block effect, image fuzziness, video coding rate as indicators, FAHP is used to calculate the weight of each index. Based on FAHP, the evaluation steps are as follows [9].

Step 1: To obtain prior evaluation matrix.

$$R = (r_{ij})_{4 \times 4} = \begin{bmatrix} 0.5 & 0.7 & 0.8 & 0.9 \\ 0.3 & 0.5 & 0.6 & 0.7 \\ 0.2 & 0.4 & 0.5 & 0.7 \\ 0.1 & 0.3 & 0.3 & 0.5 \end{bmatrix} \tag{11}$$

Step 2: To construct fuzzy decision-making matrix. Transform the matrix R.

$$R = \begin{bmatrix} 0.5000 & 0.6000 & 0.6375 & 0.7125 \\ 0.4000 & 0.5000 & 0.5375 & 0.6125 \\ 0.3625 & 0.4625 & 0.5000 & 0.5750 \\ 0.2875 & 0.3875 & 0.4250 & 0.5000 \end{bmatrix} \tag{12}$$

Step 3: Use the formula (7) to obtain the weight value of each index of QoC.

$$W'_{QoC} = (0.2875, 0.2542, 0.2416, 0.2167) \tag{13}$$

Similarly, the weights of each index relative to the previous level can be obtained

$$W_{QoE} = (0.329, 0.256, 0.223, 0.192) \tag{14}$$

$$W'_{QoS} = (0.2964, 0.2832, 0.2216, 0.1988) \tag{15}$$

$$W'_{QoI} = (0.2774, 0.2652, 0.2116, 0.2458) \tag{16}$$

$$W'_{QoT} = (0.2479, 0.2492, 0.2788, 0.2241) \tag{17}$$

Step 4: According to W_{QoS} , calculate absolute weight values of $W_{QoS}, W_{QoC}, W_{QoI}, W_{QoT}$

$$W_{QoS} = (0.0975, 0.0932, 0.0729, 0.0654) \tag{18}$$

$$W_{QoC} = (0.0736, 0.0651, 0.0618, 0.0554) \tag{19}$$

$$W_{QoI} = (0.0619, 0.0591, 0.0472, 0.0548) \tag{20}$$

$$W_{QoT} = (0.0476, 0.0478, 0.0535, 0.0430) \tag{21}$$

4 Simulation

This article designs a simulation platform to simulate wireless video transmission, in which the network transmission part is based on NS2 to achieve changes of QoS parameter to control network damage; encoding methods of wireless video server is H.264, to simulate the different encoding rate on the impact of QoC terminal; PC display resolution is 1680 * 1050, the mobile phone terminal display resolution is 1280 * 720.

In order to make the subjective score meaningful in statistics, to ensure the reliability and stability, of the data, a total of 48 testers need to have a normal vision and color vision, whose age is between 20 years and 42 years of teachers and students, of which 24 are male, 24 female; 24 people have wireless video technology background, and the rest do not. Setting the video playback 150 frames, but keeping the content unchanged. To evaluate the video subjectively, experiments are divided into 3 groups, each group has 48 testers watching the combination of the original video and the distorted video after the network transmission, according to the degree of the damage to tested video, the quality of the video sequence is scored by five-level standard of the double stimulus continuous quality scale (DSCQS), the video subjective evaluation score standard is listed in Table 2.

Table 2. Video subjective evaluation score standard

Level	Quality	Influence
5	Very good	No damage can be observed
4	Good	Damage can be observed, but is ok
3	Common	Slightly dislike
2	Bad	Dislike
1	Very bad	Extremely dislike

The test process is carried out according to the following steps: first, constructing the simulation experimental environment, set the relevant parameters, such as changing the wireless video encoding rate, adjusting QoS parameters; next, arranging 36 testers to score for wireless video quality by DSCQS, recording and calculating the average value of 36 testers evaluation for a wireless video quality, to obtain the subjective values; then, testing KPI of different experimental conditions, converting values of KPI and

multiplying by the corresponding weight value, to obtain the objective values; finally, comparing subjective and objective values, to verify the effectiveness of the algorithm.

4.1 The Influence of QoS on QoE

Keep QoC, QoI, QoT unchanged, to test QoE for changes of QoS. Ensure that the video is clear and complete, the receiving terminal is the same, changing the delay, jitter, packet loss rate and network bandwidth in both broadband access and wireless access, to test the change QoE. The detail is as follows: the sequence 1 to 4 is the delay factor changes in the range of 30–80 ms; the sequence 5 to 10 is that the jitter changes in the range of 10–60 ms; the sequence 11 to 15 is that the packet loss rate changes from 1% to 8%, in the range of change. Finally, the subjective and objective assessment results are got, as shown in Fig. 2.

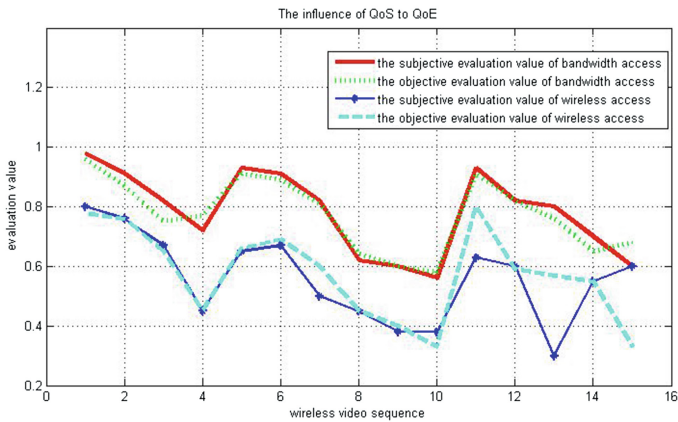


Fig. 2. The influence of QoS on QoE

As can be seen from Fig. 2, the wireless video is very sensitive to the network environment, the better the network transmission environment, the better the video quality will be guaranteed, if the transmission environment deteriorates, the video quality will be affected. QoS has great influence on the quality of the wireless video. But for the same QoS, most of the subjective and objective values change a little. However, due to the instability of wireless bandwidth and the subjective score is influenced by human, environment, etc., resulting in the subjective assessment values on wireless access of sequence 12–15 are volatile.

4.2 The Influence of QoC on QoE

Keep QoS, QoI, QoT unchanged, to test QoE for changes of QoC. Ensure that the receiving terminal and the network environment unchanged, and then change the key influencing factors (video synchronization, block effect, image fuzziness, video encoding rate) of QoC, to test the change QoE.

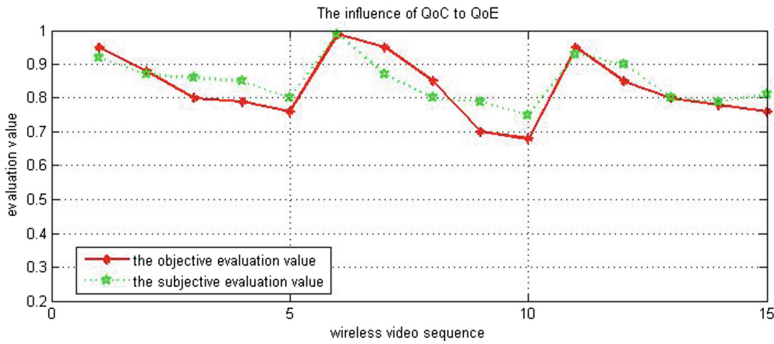


Fig. 3. The influence of QoC on QoE

As can be seen from Fig. 3, in the condition of the change of factors influencing QoC, it can be seen from the overall graph that the subjective and objective values are relatively the same.

4.3 The Influence of QoI on QoE

Keep QoS, QoC, QoT unchanged, to test QoE for changes of QoI. Ensure that the network environment unchanged, the video is clear and complete, then, change the parallax of the wireless video (sequence 1–7), the switching time of the video scene (sequence 8–11), and the waiting time (sequence 12–15), to predict the change of QoE, as shown in Fig. 4.

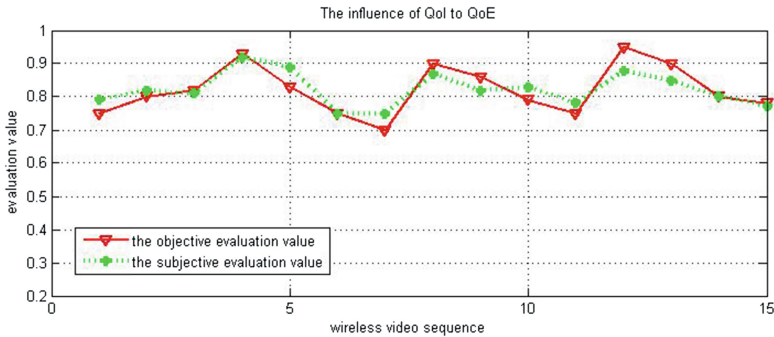


Fig. 4. The influence of QoI on QoE

From Fig. 4, changing the parallax, switching time and waiting time of the wireless video, the subjective and objective evaluation value are affected. The subjective evaluation value changes gently, indicating that QoE is more sensitive to changes of factors influencing QoI, which can more reasonably reflect the changes of wireless video's quality.

4.4 The Influence of QoT on QoE

Keep QoS, QoC, QoI unchanged, to test QoE for changes of QoT. Ensure that the network environment unchanged, the video is clear and complete, then, to test QoE on PC and cellphones. The cellphone and PC are quite different: the resolution is not the same, the resolution of PC is $1680 * 1050$, while the cellphone's is $1280 * 720$; CPU and the memory is different, the speed of data processing has great difference; the operating system of PC is windows, and the cellphone's is Android.

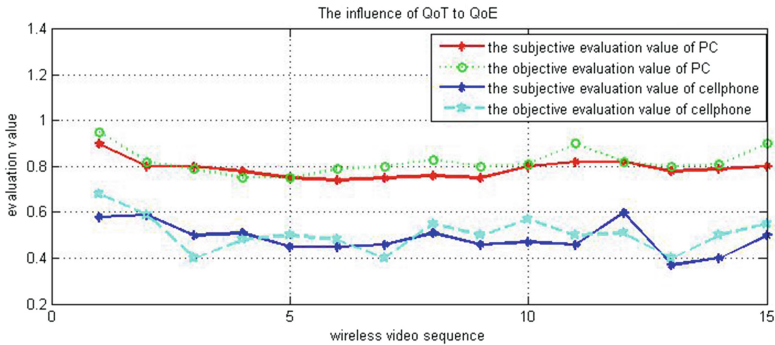


Fig. 5. The influence of QoT on QoE

The better the performance of the terminal platform, the better the quality of the received wireless video, subjective feeling is also better, the terminal performance has a great effect on QoE. From Fig. 5, the quality of wireless video from PC is stable, but the quality of wireless video from the cellphone changes a lot. In addition, the subjective quality of the wireless video received by PC tends to be gentle, indicating that the human's eyes are obtuse to the change of the video's quality, but the objective evaluation value can better reflect the actual change of the QoE of wireless video.

5 Conclusion

By analyzing the key influencing factors of QoE for wireless video in network transmission, this paper models the quality of experience for wireless video by combining the network quality of service, quality of content, quality of interaction and quality of terminal. A multi-index method of evaluating QoE for wireless video is found, and FAHP is utilized to get the value of each index. Then, the QoE evaluation model of the wireless video is verified by network simulation experiment.

References

1. Chen, D.: The QoE measurement system for data services for intelligent terminals. *Comput. Syst. Appl.*, 110–113 (2014)
2. Lin, C.: A review for models and evaluation methods of Quality of Experience (QoE). *Comput. J.*, 1–15 (2015)
3. Zhang, J.: Research on wireless video QoE model and quantitative methods. Donghua University (2016)
4. Li, M., Jiang, C.-W.: QoE-aware and cost-efficient home energy management under dynamic electricity prices. In: Conference 2016, ICUFN, pp. 498–501. IEEE (2017)
5. Chen, Y.: From QoS to QoE: a tutorial on video quality assessment. *IEEE Commun. Surv. Tutor.* **17**(2), 1126–1165 (2015)
6. Aguilera, N.: Adkintun mobile: a study of relation between QoS and QoE in mobile networks. *IEEE Lat. Am. Trans.* **14**(6), 2770–2772 (2016)
7. Sarma, A.: Context aware handover management: sustaining QoS and QoE in a public IEEE 802.11e hotspot. *IEEE Trans. Netw. Serv. Manag.* **11**(4), 530–543 (2014)
8. Wikipedia Homepage: https://en.wikipedia.org/wiki/Quality_of_experience. Accessed 23 July 2017
9. Geng, Z.: A new fuzzy process capability estimation method based on kernel function and FAHP. *IEEE Trans. Eng. Manag.* **63**(2), 177–188 (2016)

A Spectrum Sensing Scheme with Multiple Users

Junsheng Mu^{1,2(✉)}, Xiaojun Jing^{1,2}, Chenchen Sun³, and Jia Li⁴

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
muj_s@bupt.edu.cn

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing Technology and Business University,
Beijing, China

³ Foreign Language School, Beijing Technology and Business
University (BTBU), Beijing, China

⁴ School of Engineering and Computer Science,
Oakland University, Rochester, USA

Abstract. Spectrum sensing has attracted much concern of researchers due to its significant contribution to the spectral efficiency. However, the corresponding work mainly focuses on the sensing event of single primary users within a certain band and the investigation of the effect of PU traffic on the sensing performance is considered rarely. In this paper, a spectrum sensing scheme is presented to explore the co-existence of multiple users in the same frequency band based on subspace filtering. To remove uncertain noise as much as possible, subspace filtering is applied to the received signal of a cognitive radio, where the received signal is decomposed into two parts: noise subspace and signal-plus-noise subspace. Then the closed-form solution of the detection and false alarm probabilities with multiple users is given on the basis of the signal-plus-noise subspace in Rayleigh fading channel. Eventually, simulations are made to validate the proposed scheme.

Keywords: Spectrum sensing · Subspace filtering · Rayleigh fading channel

1 Introduction

With the development of wireless communication technology, scarce spectrum resource catches our attention. However, according to the investigation of Federal Communications Commission (FCC) [1], most of the registered spectrum is unoccupied in time or space. To take full advantage of the vacant spectrum, Cognitive Radio (CR) technology is presented to improve the level of spectrum utilization by allowing some unlicensed (secondary) users to have access in an opportunistic and non-interfering manner some licensed bands temporarily unoccupied by the licensed (primary) users [2].

In a CR network, one aim is to protect the primary user (PU) from the case that the registered user is present within a certain frequency range while the secondary user (SU) also uses the same frequency range for opportunistic wireless transmissions.

To avoid this situation, SU is required to sense and monitor periodically the radio frequency environment for opportunistic occupation within a certain amount of time to decide whether PU exists. Another purpose considered in CR is the achievable throughput for the secondary network. For a higher throughput (spectrum efficiency) in a secondary network, the case should be strictly restricted that SU detects the presence of PU while the frequency band is idle factually. The core technology behind the reuse of spectrum in CR network is Spectrum Sensing (SS). However, the dynamic and unpredictable sensing environment greatly lowers the detection efficiency of SU, resulting in the poor protection of PU and a lower achievable throughput for the secondary network.

In the literature of SS, a number of algorithms have been proposed to identify the presence of the primary signal as well as to improve the detection efficiency by inhibiting the disgusting interference in the sensing environment. Some examples of the existing proposals include Energy Detector (ED) [2], Matched Filtering (MF) [3], Covariance-Based Detection (CBD) [4] and Cyclostationary Feature Detection (CFD) [5]. The proposed solutions devote to optimizing sensing time, complexity or detection capabilities (the necessary prior knowledge of the primary signal or environment) for a possible tradeoff. However, the obtained performance is always at the expense of one or more of the other factors. For example, ED based methods are widely accepted for its low complexity while its detection performance works badly, especially under a low SNR. Although the CBD based method outperforms others on detection performance due to its nearly independence of noise uncertainty, the required sampling frequency is much higher than the normal conditions, leading to a higher complexity of implementation.

Interference from sensing environment almost determines the accuracy of judging whether the given frequency band is used, for most of the SS schemes. Consequently, that how to remove interference of the frequency bands has been significant for the improvement of detection accuracy and applicability to sensing scenarios. In the context of CR, close attention is paid to this issue. Andrea Mariani et al. [6] explored the SNR wall phenomenon caused by uncertain noise for ED based method, addressing the threshold design and giving the conditions for the existence of the SNR wall. Valentin Rakovic et al. [7] proposed an optimization approach for cooperative spectrum sensing utilizing ED with estimated noise power.

As a result, a subspace filtering is firstly applied to the receiver in this paper, where the observed signal of SU is divided into two subspaces: noise subspace and signal-plus-noise subspace. The noise subspace only contains background noise while the signal-plus-noise subspace contains the whole signal and some residual noise [8]. By clearing the noise subspace and noise contribution in the signal-plus-noise subspace, the remainder components mainly consist of the primary signal. Due to the possible removal of environmental interference, the existing interference is dramatically reduced and the noise uncertainty is restrained commendably. Furthermore, detection capability will be greatly heightened for the decrease of background noise.

On the other hand, most of the related works always neglect the influence from the multiple primary users on the spectrum sensing performance of the CR's energy detector. However, in several widely used wireless communication standards, such as long term evolution advanced (LTE-A), WiFi and WiMAX, where code-division

multiple-access (CDMA) is used, users simultaneously operate in the same frequency band [9]. This motivates a consideration of multiple users in the same frequency band. Factually, few researches are made to consider the effect of PU traffic on the sensing performance. In addition, these considerations of PU traffic are based on the uncertain environmental noise, which lowers the sensing efficiency and performance. In this paper, a spectrum sensing scheme with multiple users is proposed based on subspace filtering, where the closed-form solution of the detection and false alarm probabilities is given on the basis of subspace filtering. Some simulations based on MATLAB platform are made to validate the proposed method.

The rest of this paper is organized as follows. Section 2 presents the primary principle of subspace filtering. The main contribution of this paper is shown in Sect. 3, which consists in the derivation of the closed-form solution of detection and false alarm probabilities. Simulation experiments and result analysis are accomplished in Sect. 4. Finally, Sect. 5 concludes this paper.

2 Subspace Filtering

2.1 System Model

Generally, the background noise of a certain radio band we are interested in is modeled as Gaussian, independent and identically distributed random process with symbol $u(t)$. The clean primary signal is denoted as $s(t)$ and $s(t)$ is usually independent of $u(t)$ for convenience to talk about. The received signal is with central frequency f_c and bandwidth W . Then a system model at the receiver for a CR could be formulated as

$$r(t) = u(t) + s(t). \quad (1)$$

To sample the received signal at the frequency of f_s , the signal at the receiver could be described as

$$r(n) = u(n) + s(n). \quad (2)$$

Factually, several primary users may operate in the same frequency band in some CDMA applications. Therefore, $s(n)$ generally exists with another form

$$s(n) = \sum_{i=1}^M s_i(n), \quad (3)$$

where M represents the number of primary users in some a frequency band.

2.2 Subspace Filtering

In this subsection, the primary principle of subspace based filtering is discussed. Subspace filtering is a widely accepted approach in signal processing, where the received signal is decomposed into two orthogonal subspaces: noise subspace and

signal-plus-noise subspace. The main purpose of subspace filtering consists in signal enhancement by the removal of noise subspace and noise contribution in signal-plus-noise subspace. As it has been analyzed on several open literatures, only a flowchart is provided here as follows (Fig. 1).

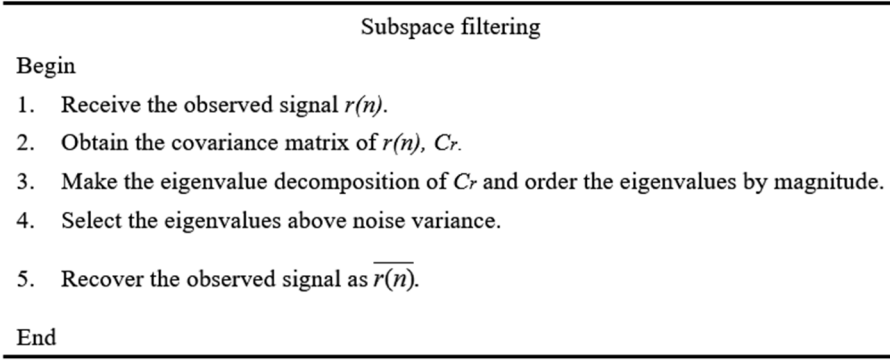


Fig. 1. The general procedure of subspace filtering.

Note that the recovered observation signal after subspace filtering is usually written as

$$\overline{r(n)} = V\Lambda_x(\Lambda_x + \mu\Lambda_u)^{-1}V^{-1}r(n), \quad (4)$$

where V is the eigenvector of C_r , Λ_x is the eigenvalue matrix after eigenvalue selection, Λ_u is the unit matrix with the same dimension as Λ_x and μ is a given Lagrange constant.

3 Multiple Users Based Spectrum Sensing

The energy of the remainders after the removal of noise can be approximatively shown as

$$E_r \approx \sum_{i=1}^M |h_i(n)s_i(n)|^2 = \sum_{i=1}^M a_i(n)|h_i(n)|^2, \quad (5)$$

Assume that $|h_i(n)|$ follows Rayleigh distribution with Probability Distribution Function (PDF)

$$f(x; k) = \frac{x^{k/2-1}}{2^{k/2}\Gamma(k/2)} e^{-x/2} U(x), \quad (6)$$

and Cumulative Distribution Function (CDF) shown as

$$F(x; k) = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)} U(x). \quad (7)$$

Assume that $a_i(n) = (1/M) \sum_{i=1}^M a_i(n) = a$ is a constant value, $|h_i(n)|^2$ follows gamma distribution with the same parameters N and $2\sigma^2/a$, the PDF of E_r in (18) can be rewritten as

$$f_r(z; M, 2\sigma^2/a) = \frac{(2\sigma^2/a)^M z^{M-1} \exp(-2\sigma^2 z/a)}{(M-1)!}, \quad (8)$$

with its CDF denoted as

$$\begin{aligned} F_r(z; M, 2\sigma^2/a) &= \frac{\gamma(M, 2\sigma^2 z/a)}{(M-1)!} \\ &= 1 - \sum_{n=0}^{M-1} \frac{1}{n!} \exp(-2\sigma^2 z/a) (2\sigma^2 z/a)^n. \end{aligned} \quad (9)$$

The corresponding detection and false alarm probabilities are formulated as follows

$$\begin{aligned} P_d &= 1 - F_r(\lambda; M, 2\sigma^2/a|H_1) \\ &= \sum_{n=0}^{M-1} \frac{1}{n!} \exp(-2\lambda\sigma^2/a) (2\lambda\sigma^2/a)^n, \end{aligned} \quad (10)$$

$$\begin{aligned} P_f &= 1 - F_r(\lambda; M, 2\sigma^2/a|H_0) \\ &= 1 - \frac{\gamma(M, 2\sigma^2 \lambda/a)}{(M-1)!}. \end{aligned} \quad (11)$$

4 Simulation and Analysis

Simulations are conducted in this section to verify the performance of the proposed method.

Figure 2 gives ROC (Receiver Operating Characteristic) of different scenarios for the case that $a_i(n) = (1/M) \sum_{i=1}^M a_i(n) = a$ is a constant value according to the formula (10) and (11). In Fig. 2, M represents the number of PU in the same frequency band, var denotes the variance σ^2 in (8) and $a = (i, j)$ indicates the amplitude of H_1 and H_0 under the identical scenarios, where $a = i$ denotes the squared amplitude of residual noise in H_0 case after subspace based filtering and $a = j$ denotes squared amplitude of H_1 case after subspace based filtering.

Figure 2-a mainly discusses how the variance of $|h_i(n)|$ influences the detection performance of a CR for multiple PUs at a certain frequency band. From the simulation results, different variances of $|h_i(n)|$ have few effects on the detection performance, which indicates the detection performance is independent of the variance of $|h_i(n)|$. Figure 2-b exhibits the influence from the signal squared magnitude on the detection performance, where the detection performance improves as the rise of the squared signal magnitude when the noise squared magnitude is invariant. This declares that the signal magnitude is one of the factors to determine the detection accuracy for a set of CR equipment. Similarly, Fig. 2-c shows the influence from the residual noise squared magnitude on the detection performance. Obviously, the ROC increases as the decline of residual noise squared magnitude. Figure 2-b and c indicate that the detection performance of a CR in the case of multiple PUs also suffers from the influence of SNR (signal-to-noise-ratio), where the performance is in positive relation to the variation of SNR. How the number of PU alters the detection performance is investigated in Fig. 2-d. It is emphasized that the detection performance arises with the increase of PU number in the same frequency, which results from the removal of background noise. The removal of background noise is equivalent to a rise of SNR, which will make the sensing performance turn better in return.

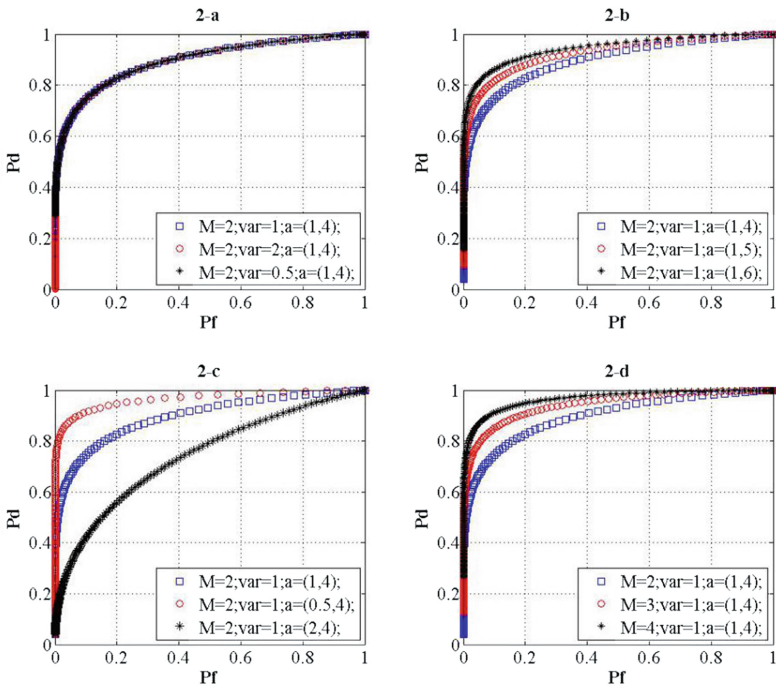


Fig. 2. ROC curve of different situations

5 Conclusions

A multiple users based spectrum sensing scheme is explored on the basis of subspace filtering in this paper. The corresponding closed-form solutions on the detection and false alarm probabilities are given on the assumption that the signal magnitude is identical. The case with different signal magnitude will be discussed for the potential applicability later.

References

1. Sun, S., Ju, Y., Yamao, Y.: Overlay cognitive radio OFDM system for 4G cellular networks. *IEEE Wireless Communications* **20**(2), 68–73 (2013)
2. Mu, J., Jing, X., Huang, H., et al.: Subspace based method for spectrum sensing with multiple users over fading channel. *IEEE Commun. Lett.* **1**(1), 99 (2017)
3. Digham, F., Alouini, M.S., Simon, M.K.: On the energy detection of unknown signals over fading channels. *IEEE Trans. Commun.* **55**(1), 21–24 (2007)
4. Salem, F.M., Ibrahim, M.H., Ali, I.A., et al.: Matched-filter-based spectrum sensing for secure cognitive radio network communications. *Int. J. Comput. Appl.* **87**(18), 41–46 (2014)
5. Hamid, M., Björnsell, N., Slimane, S.B.: Energy and Eigenvalue Based Combined Fully Blind Self Adapted Spectrum Sensing Algorithm. *IEEE Trans. Veh. Technol.* **65**(2), 630–642 (2015)
6. Bkassiny, M., Jayaweera, S.K., Li, Y., et al.: Blind cyclostationary feature detection based spectrum sensing for autonomous self-learning cognitive radios. In: *IEEE International Conference on Communications*, pp. 1507–1511. IEEE (2012)
7. Tandra, R., Sahai, A.: SNR walls for signal detection. *IEEE J. Sel. Top. Sign. Process.* **2**(1), 4–17 (2014)
8. Rakovic, V., Pavlovska, V., Atanasovski, V., et al.: Cooperative spectrum sensing based on noise power estimation. In: *International Symposium on Wireless Personal Multimedia Communications*, pp. 1–5. IEEE (2013)
9. Sun, C., Mu, J.: An eigenvalue filtering based subspace approach for speech enhancement. *Noise Control Eng. J.* **63**(1), 36–48 (2015)
10. Boulogeorgos, A.-A.A., Chatzidiamantis, N.D., Karagiannidis, G.K.: Spectrum sensing with multiple primary users over fading channels. *IEEE Commun. Lett.* **64**(7) (2016)

A New Method of Spectrum Sensing in Cognitive Radio Based on Statistical Covariance Matrix

Zhaocong Sun^{1,2(✉)}, Xiaojun Jing^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
szcsun5@163.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts
and Telecommunications, Beijing, China

³ School of Engineering and Computer Science, Oakland University,
Rochester, USA

Abstract. Spectrum sensing is a significant part of technique in a cognitive radio that detecting the presence of primary users in an authorized spectrum. The method that based on the statistical covariance matrix is one of main spectrum sensing techniques, using the difference of statistical covariance between the received signal and noise. In this paper, the new sensing method we proposed is also based on the statistical covariance. The new method compare to some traditional covariance algorithms has decrease the complexity of algorithm, at the same time, ensured the accuracy of detection. We give the statistics of detection, and we also find the threshold of the method when the probability of false alarm is given. The analysis and derivation process of threshold are provided in behind. Using Matlab for simulation to validate the correctness of the method and making the comparison with some typical detection method.

Keywords: Cognitive radio · Statistical covariance matrix · Energy detection Threshold · Eigenvalue

1 Introduction

With the rapid development of wireless communication technologies and low spectrum utilization rate in many frequency bands, which have increased the demand of usage of spectrum resource. To solve this problem, Cognitive Radio (CR) [1, 2] has is a promising technology that exploit the underutilized spectrum in an dynamic access manner and allow secondary users (SUs) to share the licensed spectrum of PU provided that PU is not harmfully interfered [3].

For spectrum sensing, it is the based technique and the key for cognitive radio. However, there are several challenges that we should to overcome. First, the signal-to-noise ratio (SNR) may be in a low sensing environment. Second, the sensing problem caused by multipath fading and time dispersion of the wireless channels. Third, the noise uncertainty.

Common spectrum sensing strategies include the energy detection [4, 5], and the cyclostationary detection [5], the matched filtering detection [6], spectrum sensing based on the statistical covariance matrix. the detection methods above have their advance and shortage. For instance, the matched filtering detection need the prior knowledge [7] of channel and accurate synchronization. Energy detection does not need any prior knowledge of the signal, but the false estimation of noise power could cause the high probability of false alarm.

In this paper, the method of detection that we propose have the better accuracy than energy detection. Comparing with classical algorithm of statistical covariance matrix like the Eigenvalue-Based spectrum sensing algorithms, the complexity of our algorithm is lower than it. The theoretical basis of the method is to assume that the signal samples are correlated. Making full use difference of correlation of signal between PU and noise. The oversampled signal, the time dispersed propagation channel and the correlated original signal are the key to make signal samples being correlated.

The organization of this paper is as follows. Section 2 presents the system model. In Sect. 3 gives the performance analysis and finds thresholds for the algorithms. Simulation results will be given in Sect. 4. Conclusion are drawn in Sect. 5.

2 System Model

$x(t) = s(t) + \eta(t)$ is the continuous-time received signal, where $s(t)$ is the primary user's signal and $\eta(t)$ is the noise. $\eta(t)$ is assumed to be a stationary process satisfying $E(\eta(t)) = 0$, $E(\eta^2(t)) = \sigma^2$, and $E(\eta(t)\eta(t + \tau)) = 0$ for any $\tau \neq 0$. f_s is our sample rate, where $f_s \geq W$. $T_s = 1/f_s$ is the sampling period. To discretize the continuous signal, we define $x(n) = x(nT_s)$, $s(n) = s(nT_s)$, and $\eta(n) = \eta(nT_s)$. There are two hypotheses: 1, H_0 signal does not exit 2, H_1 , the signal exists.

$$x_i(n) = \begin{cases} \eta_i(n) & H_0 \\ s(n) + \eta_i(n) & H_1 \end{cases} \tag{1}$$

In this formula $x_i(n)$ denote the nth signal sample taken by ith sensing user(second user), $s(n)$ is the receive signal which is effect by path loss and multipath fading. η_i is the receive white Gaussian noise by which the ith SU is interfered and assumed to be i.i.d (independent and identically distributed), and with mean zero and variance σ_η^2 .

Let consider M is the number of SUs, and different SU sample the receive signal from the same primary signal. Then we obtain a matrix.

$$X = [x_1, x_2, x_3, \dots, x_M]^T,$$

As the same:

$$S = [s, s, \dots, s]^T,$$

$$\eta = [\eta_1, \eta_2, \dots, \eta_M]^T,$$

So we can express the matrix as follows.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(N) \\ x_2(1) & x_2(2) & \cdots & x_2(N) \\ \cdots & \cdots & \cdots & \cdots \\ x_M(1) & x_M(2) & \cdots & x_M(N) \end{bmatrix}, \quad (2)$$

Assume s and η are independence. when in situation H_1 , consider the Covariance Matrix in M SUs:

$$R_x = E[XX^H]. \quad (3)$$

The Covariance Matrix in of primary signal after channel is

$$R_s = E[SS^H], \quad (4)$$

then

$$R_x = R_s + \sigma^2 I_M, \quad (5)$$

I_M is the identity matrix, σ^2 is the variance of noise.

Because of the statistical covariance matrix can only be acquired by using a limited number of signal samples, we couldn't get R_x accurately. Since R_x can be approximated by the sample covariance matrix defined as

$$R_x(N) = \frac{1}{N} \sum_{n=0}^{N-1} XX^H. \quad (6)$$

In hypothesis H_0 , $s(n)$ is not exist, ideally as the result $R_s = 0$, $R_x = \sigma^2 I_M$.

$$R_x = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ & \sigma^2 & \ddots & \vdots \\ & & \ddots & 0 \\ & & & \sigma^2 \end{bmatrix} \quad (7)$$

In other hand, in hypothesis H_1 , $R_s \neq 0$. The elements in Upper and down triangle is not zero. So we can use difference of Covariance Matrix when PU is existed or not to make statistics:

$$T_1 = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^M |C_{ij}|, \quad (8)$$

the sum of elements in the matrix divided M

$$T_2 = \frac{1}{M} \sum_{i=1}^M |C_{ii}|, \quad (9)$$

the sum of diagonal divided M

$$T_{CDV} = T_1 - T_2, \quad (10)$$

$T_{CDV} = 0$ when primary signal is not existed. (the noise is white Gaussian noise $T_{CDV} \neq 0$ when primary signal is existed. However, it works in an ideal situation, because of limited number of signal samples. So we need setting the threshold.

3 Threshold Determination

As a good detection method, should have the high P_d and low P_{fa} . P_d and P_{fa} largely determine the choice of threshold γ (P_d is the probability of detection, P_{fa} is the probability of false alarm.). How to set the threshold? We don't have any information of the signal (in real sensing scene we do not even know whether the signal is exiting or not), it is difficult to set the threshold based on P_d . As the result, we choose the threshold based on P_{fa} . First we set an ideal P_{fa} . Then, there are two ways to find the threshold based on P_{fa} , theoretical derivation and computer simulation. In this paper we choose the way of theoretical derivation to find the threshold. It is necessary for us to find the statistical distribution of $T_{CDV} = T_1 - T_2$, which is a hard working. However, we get some references to support statistical distribution.

We can get a conclusion from the reference, when the PU signal does not exist, the expectations of statistics of T_1 and T_2 is:

$$E(T_1) = [1 + (M - 1) \sqrt{\frac{2}{\pi N}}] \sigma^2, \quad (11)$$

$$E(T_2) = \sigma^2 \quad D(T_2) = \frac{2\sigma^4}{N}, \quad (12)$$

To find the threshold based on P_{fa} , we can get probability distribution function:

$$P_{fa} = p_r \{T_{CDV} > \gamma\}$$

$$P_{fa} = p_r \{T_1 - T_2 > \gamma\}$$

When the number of sample is large, we can know from the central limit theorem that T_1 and T_2 follow Gaussian distribution.

$$\begin{aligned}
 P_{fa} &= p_r\{T_1 - T_2 > \gamma\} \approx p_r\left\{[1 + (M - 1)\sqrt{\frac{2}{\pi N}}]\sigma^2 - T_2 > \gamma\right\} \\
 &= p_r\{T_2 < [1 + (M - 1)\sqrt{\frac{2}{\pi N}}]\sigma^2 - \gamma\} \\
 &= p_r\left\{\frac{T_2 - \sigma^2}{\sqrt{\frac{2}{N}}\sigma^2} < \frac{[1 + (M - 1)\sqrt{\frac{2}{\pi N}}]\sigma^2 - \gamma - \sigma^2}{\sqrt{\frac{2}{N}}\sigma^2}\right\} \\
 &= 1 - Q\left[\frac{[1 + (M - 1)\sqrt{\frac{2}{\pi N}}]\sigma^2 - \gamma - \sigma^2}{\sqrt{\frac{2}{N}}\sigma^2}\right],
 \end{aligned} \tag{13}$$

So we figure out the γ

$$\gamma = \sigma^2 \left[(M - 1)\sqrt{\frac{2}{\pi N}} - \sqrt{\frac{2}{N}}Q^{-1}(1 - P_{fa}) \right]. \tag{14}$$

The Q function is: $Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-\frac{u^2}{2}} du$.

From the threshold expression we can get the conclusion that if we set the value of the P_{fa} , and know the variance of noise, we can get the value of threshold.

4 Simulation

In this section, we will give the simulation of new method, and we use the receiver operating characteristic (ROC) curve to show the performance of the new method.

1. The simulation of new method in different SNR situation.

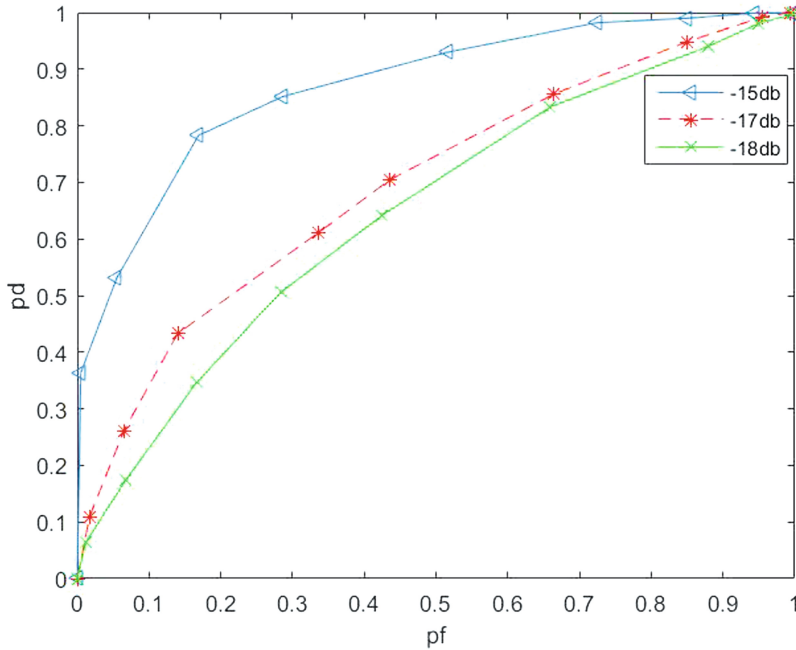


Fig. 1. The simulation of new method in different SNR situation.

From the result of Fig. 1, we can see clearly that the abscissa is the given p_f , the ordinate is the corresponding p_d under the given p_f . The lines respectively represent the simulation is under the SNR of -15 dB, -17 dB, -18 dB. The new method also has the good performance in big SNR situation.

2. The simulation of new method compare to traditional cooperative Energy Detection and Maximum-minimum Eigenvalue Detection.

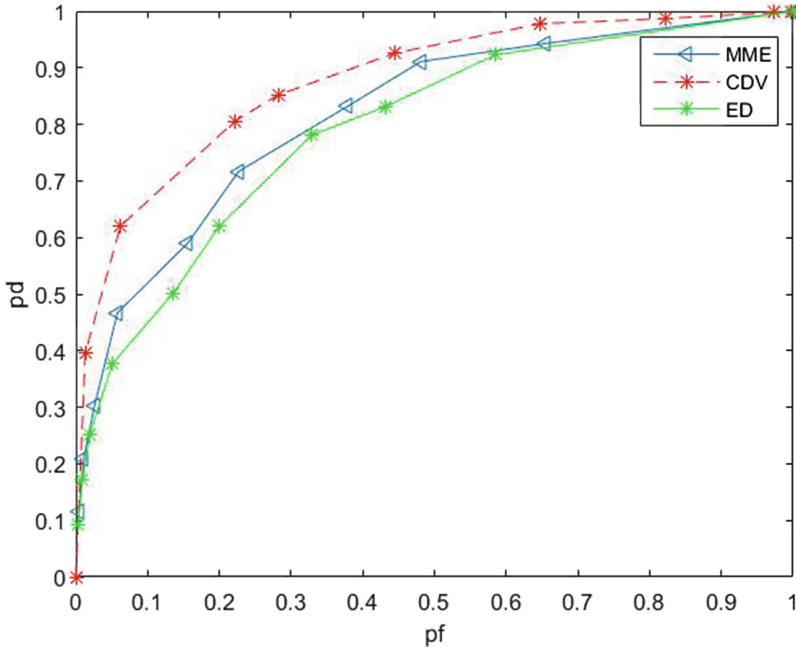


Fig. 2. Compare to the traditional methods

In this simulation Fig. 2, showing the different detected methods. So we can get the result very clearly that the new method is always better than the traditional Energy Detection and Maximum-minimum Eigenvalue Detection. The simulation is under the SNR of -15 dB. The abscissa is the given p_f , the ordinate is the corresponding p_d under the given p_f .

5 Conclusion

In this paper, we proposed a new method of spectrum sensing in cognitive radio based on Statistical Covariance. We give a new Statistical algorithm, based on coherence of signal and noise. Then we analysis distribution of this statistic to get the expression of threshold. Last the simulation under the different SNR situation and compare to the other spectrum sensing methods show our new method performance is well.

Acknowledgement. This works thanks to the project 61471066 supported by NSFC.

References

1. Jacob, P., Sirigina, R.P., Madhukumar, A.S.: Cognitive radio for aeronautical communications: a survey. *IEEE Access* **4**, 3417–3443 (2016)
2. Sun, S., Chen, N., Ran, T., Xiao, J., Tian, T.: A Stackelberg game spectrum sharing scheme in cognitive radio-based heterogeneous wireless sensor networks. *Signal Process.* **126**, 18–26 (2016)
3. Qian, C., Qian, H., Gao, F.: Spectrum sensing and SNR walls when primary user has multiple power levels. In: *IEEE/CIC International Conference on Communications in China*, pp. 1–6. IEEE (2015)
4. Choi, K.W.: Adaptive sensing technique to maximize spectrum utilization in cognitive radio. *IEEE Trans. Veh. Technol.* **59**(2), 992–998 (2010)
5. Hattab, G., Ibnkahla, M.: Multiband spectrum access: great promises for future cognitive radio networks. *Proc. IEEE* **102**(3), 282–306 (2014)
6. Zeng, Y., Liang, Y.C.: Spectrum-sensing algorithms for cognitive radio based on statistical covariances. *IEEE Trans. Veh. Technol.* **58**(4), 1804–1815 (2009)
7. Guo, C., Chen, S., Liu, F.: Polarization-based spectrum sensing algorithms for cognitive radios: upper and practical bounds and experimental assessment. *IEEE Trans. Veh. Technol.* **65**(10), 8072–8086 (2016)

An Enhanced Double Threshold Energy Detection in Cognitive Radio

Xuan Zhou^{1,2(✉)}, Xiaojun Jing^{1,2}, Hai Huang¹, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
zhouxuan@bupt.edu.cn

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ School of Engineering and Computer Science,
Oakland University, Rochester, USA

Abstract. Cognitive radio (CR) is regarded as a perfect technique to cope with the scarcity of spectrum resources. Energy detection is preferred by most of Cognitive Radio researchers, because it is easy to implement and it doesn't need the prior information about primary user's (PU) signals. But the performance of energy detection is poor at the low signal-to-noise ratio (SNR) regime. Double threshold spectrum sensing scheme was proposed to increase the reliability of decision. Under the same SNR, if the sensing time is not long enough, a higher noise variance will make the instance energy level falls below the threshold because of the noise uncertainty. In this paper, an enhanced energy detection (EED) scheme combined with double thresholds is proposed, this scheme can reduce the misdetection caused by noise uncertainty. In the proposed method, we compare the average of last M sensing statistic with the preset threshold. It aims at protecting a channel that are underutilizing from being decided to be idle when an immediate signal energy drop happens. The simulation makes a comparison between the proposed method and the traditional method and proves the effectiveness of the new scheme.

Keywords: Double threshold · Energy detection · Noise uncertainty

1 Introduction

The wireless communication technology develops rapidly, the demand of the wireless spectrum resources also keeps growing, the fixed spectrum selection allocation policy by which the most of spectrum resources are allocated for specific use make the situation even worse. According to Federal Communications Commission (FCC), the actual utilization rate of most band is between 5% and 85%. From the figure, we can find that the efficiency is extremely low.

To cope with the scarcity of spectrum resources, Cognitive Radio (CR) was firstly proposed by Mitola and Maguire [1]. CR is a smart wireless communication system that can improve the utilization of spectrum resources through permitting secondary users (SU) opportunistic use the authorized bands without interfering PU. For the

development of CR, Spectrum Sensing is one of the biggest challenges. There are many methods designed to detect the accessible spectrum resources, such as Matched Filter Detection (MFD), Energy Detection (ED) and Cyclostationary Feature Detection (CFD). Cyclostationary method has a good performance but it takes a longer time with the prior information about PU (Primary User)'s cyclic frequency. Matched filter detection has fast speed but it needs extra prior information about PU's signaling features. With the advantages of simple algorithm, easily implementation and working without priori information, energy detection is prevalent.

Energy detection calculates the energy level of the receiving unknown signal and then compares it with a predefined threshold to conform its presence or absence within a given bandwidth. The performance of different sensing technique is represented by a Receiver Operating Characteristic (ROC) curve which is consisted of p_d (detection probability) and p_{fa} (false alarm probability). p_d represents the probability that the presence of PU is correctly detected while p_{fa} represents the probability that declaring PU presents but actually not. Higher detection probability indicates a better protection of PUs and lower false alarm probability indicates a higher efficient utilization of the spectrum.

The remaining part of this paper consists of following part: Sect. 2 briefly introduces the system model of CR. Section 3 mainly describes the proposed enhanced energy detection method, in Sect. 4, we analyze the theoretical performance of the proposed method. The last section shows the experiment performance to prove the validity of the method.

2 System Model

The problem of spectrum sensing is usually expressed as a binary assumption:

$$H_0 : y[n] = w[n] \quad n = 1, 2, \dots, N, \quad (1)$$

$$H_1 : y[n] = x[n] + w[n] \quad n = 1, 2, \dots, N, \quad (2)$$

where $y[n]$ represents the samples of receiving signal, $x[n]$ represents the sample of authorized user signal and $w[n]$ is the sample of noise. H_0 is the null hypothesis stating that the sensed spectrum is not being used by PU, H_1 is the other assumption indicating the PU is present.

The traditional energy detection measures the energy level on the sensed spectrum during an observation interval and makes a comparison between the energy level and the threshold. If the measured energy level is above the threshold, the sensed spectrum will be marked as busy, otherwise it will be marked as idle.

There are two threshold λ_1 and λ_2 ($\lambda_1 < \lambda_2$) in double threshold scheme, if the measured energy is less than λ_1 , the sensed spectrum will be marked as idle; if the measured energy is greater than λ_2 , the sensed spectrum will be marked as busy; otherwise, no decision will be made.

If N is not big enough, because of the noise uncertainty, the test statistic would change with the immediate variation of the received signal. That means the target band

will be marked as idle although it should be marked as busy which will lead to a misdetection.

3 Proposed Enhanced Double Threshold Energy Detection

This paper proposes a new sensing method by combining EED and double threshold method to alleviate the misdetection caused by the insufficient N . By applying the EED to the situation when measured energy falls between λ_1 and λ_2 , not only the detection probability but also the throughput will be improved.

3.1 Operating Principle

On a primary band, the received energy during a fixed observation interval can be represented by

$$E_i(y) = \sum_{n=1}^N |y[n]|^2, \quad (3)$$

where $E_i(y)$ is the measured energy level of signal y at the i th observing period. Compare the measured energy with two thresholds, the result can be defined as follow:

$$D = \begin{cases} H_1 & E_i(y) < \lambda_1 \\ EED & \lambda_1 < E_i(y) < \lambda_2 \\ H_0 & E_i(y) > \lambda_2 \end{cases} \quad (4)$$

Now we mainly focus on the situation when the measured energy falls between λ_1 and λ_2 . The main difference between EED and traditional energy detection is that EED additionally keeps a list which contains the measured energy level of the last M sensing results. With this updated list, we can compute the average statistic value of M sensing events.

$$E_i^{avg}(E_i) = \frac{1}{M} \sum_{m=1}^M E_{i-M+m}(y) \quad (5)$$

We know that if the sensing time is not long enough, the statistic value $E_i(y)$ will follow the instantaneous energy drop and it will cause a busy channel be declared as idle. So $E_i^{avg}(E_i)$ is used for an additional check when $E_i(y)$ falls below λ_2 . This additional check which can improve the detection probability is aimed at protecting a channel that are underutilizing from being decided to be idle when an immediate signal energy drop happens with an insufficiently sensing period. However, it can also lead to the increasing of false-alarm probability thus degrading ROC. So we have to check the relationship between $E_{i-1}(y)$ and λ_2 when $E_i^{avg}(E_i) > \lambda_2$

It aims

Algorithm 1. Operating principle of double threshold EED

Algorithm

Input: $\lambda_1, \lambda_2 \in \mathbb{R}^+$, $N \in \mathbb{N}$, $L \in \mathbb{N}$

Output: $D_i \in \{H_0, H_1\}$

1: for each sensing event i do

2: calculate energy of N samples: $E_i(y)$

3: if $E_i(y) > \lambda_2$

4: $D_i = H_1$

5: end if

6: if $E_i(y) < \lambda_1$

7: $D_i = H_0$

8: end if

9: if $E_i(y) < \lambda_2$

10: calculate mean of the latest M measured energy: $E_i^{avg}(E_i)$

11: if $E_i^{avg}(E_i) > \lambda_2$ and $E_{i-1}(y) > \lambda_2$ then

12: $D_i = H_1$

13: else

14: $D_i = H_0$

15: end if

16: end if

17: end for

4 Double Threshold EED Theoretical Performance

The observation result can be approximated as Gaussian distribution

$$E_i(y) \sim \begin{cases} \mathcal{N}(N\sigma_w^2, 2N\sigma_w^4) & \mathcal{H}_0 \\ \mathcal{N}(N(\sigma_x^2 + \sigma_w^2), 2N(\sigma_x^2 + \sigma_w^2)^2) & \mathcal{H}_1 \end{cases} \quad (6)$$

$E_i^{avg}(E_i)$ is the average of i.i.d. Gaussian random variables, so it's also Gaussian distributed

$$E_i^{avg}(E_i) \sim \mathcal{N}(\mu_{avg}, \sigma_{avg}^2) \quad (7)$$

in which

$$\mu_{avg} = \frac{K}{M}N(\sigma_x^2 + \sigma_w^2) + \frac{M-K}{M}N\sigma_w^2 \quad (8)$$

$$\sigma_{avg}^2 = \frac{K}{M^2}2N(\sigma_x^2 + \sigma_w^2)^2 + \frac{M-K}{M^2}2N\sigma_w^4 \quad (9)$$

So the p_d and p_f can be represented as

$$\begin{aligned} p_d &= p\{E_i(y) > \lambda_H\}_{H_1} + p\{\lambda < E_i(y) < \lambda_H\}_{H_1} \\ &\quad + p\{\lambda_L < E_i(y) \leq \lambda, E_i^{avg}(E_i) > \lambda, E_{i-1}(y) > \lambda\}_{H_1} \\ &= p\{E_i(y) > \lambda\}_{H_1} + p\{\lambda_L < E_i(y) \leq \lambda, E_i^{avg}(E_i) > \lambda, E_{i-1}(y) > \lambda\}_{H_1} \\ &= p\{E_i(y) > \lambda\}_{H_1} + p\{\lambda_L < E_i(y) \leq \lambda\}_{H_1} \cdot p\{E_i^{avg}(E_i) > \lambda\}_{H_1} \cdot p\{E_{i-1}(y) > \lambda\}_{H_1} \\ &\quad > p\{E_i(y) > \lambda\}_{H_1} = p_d^{conv} \end{aligned} \quad (10)$$

$$\begin{aligned} p_{fa} &= p\{E_i(y) > \lambda_H\}_{H_0} + p\{\lambda < E_i(y) < \lambda_H\}_{H_0} \\ &\quad + p\{\lambda_L < E_i(y) \leq \lambda, E_i^{avg}(E_i) > \lambda, E_{i-1}(y) > \lambda\}_{H_0} \\ &= p\{E_i(y) > \lambda\}_{H_0} + p\{\lambda_L < E_i(y) \leq \lambda, E_i^{avg}(E_i) > \lambda, E_{i-1}(y) > \lambda\}_{H_0} \\ &= p\{E_i(y) > \lambda\}_{H_0} + p\{\lambda_L < E_i(y) \leq \lambda\}_{H_0} \cdot p\{E_i^{avg}(E_i) > \lambda\}_{H_0} \cdot p\{E_{i-1}(y) > \lambda\}_{H_0} \\ &\quad > p\{E_i(y) > \lambda\}_{H_0} = p_f^{conv} \end{aligned} \quad (11)$$

Obviously, both p_d and p_{fa} of EED with DTH is higher than conventional double threshold energy detection. It means that EED algorithm improves the detection probability by sacrificing the false alarm probability. But the degradation of false alarm probability is not significant, so the ROC curve is still better than the traditional double threshold energy detection.

5 EED Experimental Performance

The ROC curve of double threshold EED scheme is shown in Fig. 1. From this figure, we can find that under the same circumstance (SNR = -10, N = 3000), the performance of double threshold EED is better than double threshold with CED which proved the superiority of DTH with EED. Based on the extra check, if the result of last sensing event is idle while the average value of the previous M sensing periods is bigger than the predefined threshold, it is very likely that PU actually presents but the result of last sensing period is below the upper threshold. In such circumstance, the target band should be marked as busy. Meanwhile, if both the result of last sensing period and the average value of the previous M sensing period's test statistic are idle, it indicates that the target channel is really not occupied. So the proposed method can

reduce the probability of misdetections caused by immediate energy changes, which will help to improve the detection results. Figure 2 shows the ROC curve of EED under different SNR. We can see that with the increasing of SNR, the performance of the enhanced double threshold energy detection is also improved.

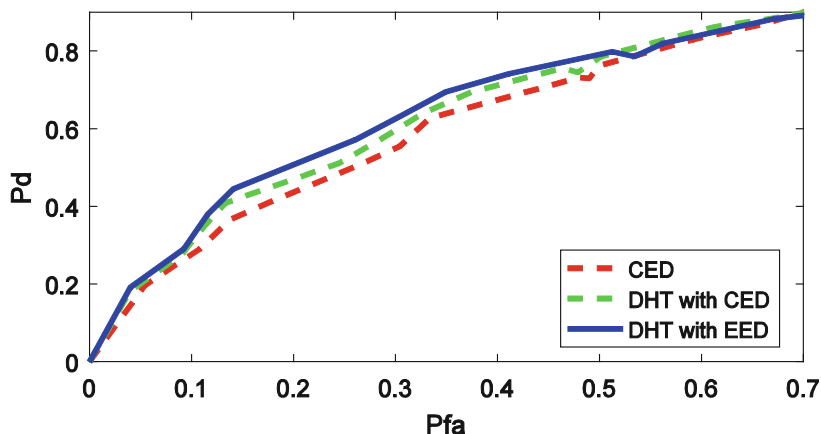


Fig. 1. ROC curve for CED, Double Threshold with CED and Double Threshold with EED ($M = 3$, $\text{SNR} = -10$, $N = 3000$)

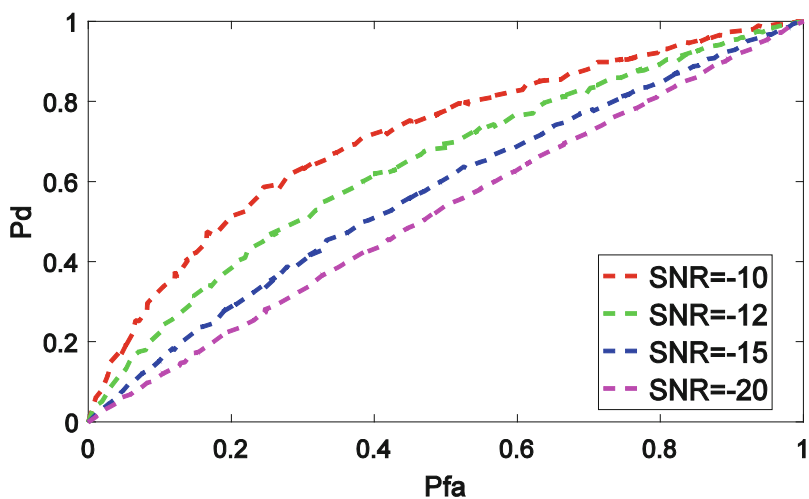


Fig. 2. Performance of DTH with EED under different SNR

6 Conclusions

As we know, the main drawback of energy detection is the limitation caused by noise uncertainty. And many algorithms are proposed to solve the limitation, but the improvement of performance is based on significant computational cost. In this paper, we proposed a better algorithm without increasing the computational work and finished the simulation to prove the effectiveness. The simulation results prove that the proposed method has a better ROC curve than the classical energy detection and classical double threshold energy detection.

Acknowledgement. Project 61471066 supported by NSFC.

References

1. Mitola, J., Maguire Jr., G.Q.: Cognitive radio: making software radios more personal. *IEEE Pers. Commun.* **6**(4), 13–18 (1999)
2. Federal Communications Commission: Spectrum Policy Task Force. Technical report, ET Docket no. 02-135, November 2002
3. Chen, Y.: Improved energy detector for random signals in gaussian noise. *IEEE Trans. Wirel. Commun.* **9**(2), 558–563 (2010)
4. Jafarian, J., Hamdi, K.A.: Sensing-throughput tradeoff in a non-cooperative double-threshold sensing scheme. In: *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, pp. 201–206. IEEE (2013)
5. López-Benítez, M., Casadevall, F.: Performance of spectrum sensing for cognitive radio based on field measurements of various radio technologies. In: *Wireless Conference*, pp. 969–977. IEEE (2010)
6. Liang, Y.C., Zeng, Y., Peh, E.C.Y., Hoang, A.T.: Sensing-throughput tradeoff for cognitive radio networks. *IEEE Trans. Wireless Commun.* **7**(4), 1326–1337 (2009)
7. Bhowmick, A., Nallagonda, S., Roy, S.D., Kundu, S.: Spectrum sensing with censoring of double threshold based cognitive radios in Rayleigh fading. In: *Communications*, pp. 1–6. IEEE (2014)
8. Bagwari, A., Tomar, G.S.: Adaptive double-threshold based energy detector for spectrum sensing in cognitive radio networks. *Int. J. Electron. Lett.* **1**(1), 24–32 (2013)
9. Hu, R.Q., Qian, Y.: An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems. *IEEE Commun.* **52**(5), 94–101 (2014)
10. Li, Q., Hu, R.Q., Qian, Y., Wu, G.: Cooperative communications for wireless networks, in: *techniques and applications in LTE-advanced systems*. *IEEE Wirel. Commun.* **19**(2), 22–29 (2012)

An Energy Detection Based on Coefficient of Variation for Spectrum Sensing in Cognitive Radio

Wenwen Zhang^{1,2(✉)}, Xiaojun Jing^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
zhang.wenwen1@foxmail.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ School of Engineering and Computer Science,
Oakland University, Rochester, USA

Abstract. In the Cognitive Radio system, cognitive users require perceiving the real-time using of the spectrum accurately. At the same time, the cognitive user stations are usually in severe fading or interference. Energy detection is used widely due to the low computation complexity and an effective method under the high SNR, but it is impressionable by the noise. According to these facts, we propose to use the coefficient of variation (CV) of the sampled signals to amend the judgment result and get the blending final result. After the energy detection, if the test statistic of the signal energy is within a specified range and the signal CV is less than a threshold value, it concludes that the spectrum bands are occupied by another user. The simulation results prove that the rectification method can greatly improve the cognitive user's accurate detection performance of spectrum usage in real-time while ensuring the computation complexity.

Keywords: Spectrum sensing · Energy detection · Coefficient of variation

1 Introduction

Today, as the rapid progress and development of modern wireless communications technology, the radio frequency spectrum resources have become more and more strained. Cognitive Radio (CR) technology can raise the utilization and reduce the tension situation of spectrum resources [1]. In CR networks, the Secondary User (SU) is allowed to use the licensed spectrum when the Primary User (PU) don't occupy the bandwidths. With the guaranty of PU in a certain band, the SU needs to continually detect the presence of the PU. Spectrum sensing is the foundation and precondition for the implementation of Cognitive Radio, which also is one of the key techniques in CR [2].

Among the common spectrum sensing methods, energy detection has been used widely because of its simplicity [3]. It is the simplest method to detect the presence of PU, which doesn't require prior information of the PU. The traditional energy detection based schemes typically compare the received energy with the fixed threshold to decide

whether the PU is active or not. But the energy threshold is related to the channel noise and is easy to be affected by the noise power fluctuation. The traditional dual-threshold energy detection algorithm has two thresholds to make the decision, the detected energy values are divided into three areas. But there is a problem of “Confused Region” that is between upper bounds and lower bounds. Once the detected energy value falls into the region, the result of the spectrum sensing would be failing.

The correlation coefficient is used to measure the linear correlation between quantitative variables. Based on the analysis of eigenvalue detection theory, the correlation between PU signal and additive white Gaussian noise (AGWN) is different, and the correlation of PU signal is higher than AGWN [4]. But it is not always a good criterion of the signals’ volatility. Considering traditional energy detection algorithm’s shortages and the influence of the noise uncertainty on system performance, this paper proposes a new method of double-threshold energy detection algorithm which is based on the received signal coefficient of variation (SCVED). This method is applied for the SU between the two thresholds. The advantage of using SCVED instead of energy detection is two-fold: improved the detection probability; reduced the influence of the noise.

The remainder of the paper is organized as follows: the conventional signal energy detection method is briefly reviewed in Sect. 2. Then, we focus on the technical aspects of SCVED in Sect. 3. In Sect. 4, its performance is evaluated by experiments. Finally, we summarize the paper in the last section.

2 The Conventional Signal Energy Detection Method

Energy detection collects and calculates the signal energy in a given time period, compares the result with a pre-set threshold to judge whether PU is present or not. In addition, suppose that there is no signal between the PU and the SU. That is, the SU is completely independent of the PU. A binary hypothesis testing model is shown, which spectrum sensing can be modeled as:

$$y(t) = \begin{cases} n(t), & H_0 \\ n(t) + h \cdot x(t), & H_1 \end{cases} \quad t = 1, 2, \dots, N. \quad (1)$$

In the testing, N denotes the sampling number during a sampling period, $y(t)$ shows the sample of the received signal by the SU, $n(t) \sim N(0, \sigma_w^2)$ is the AWGN, $x(t)$ is the PU signal sample, h indicates the channel gain. In the channel, hypothesis H_0 states that the PU is absent, and hypothesis H_1 states that the PU is present.

For the detection of unknown signals disturbed by AWGN, a conventional energy detector is derived in [5]. This is easily implemented detector for detection of unknown signals in spectrum sensing. Collect the test statistic and compare it to a threshold λ to decide whether the PU exists.

The test statistic is calculated by:

$$Y = \sum_{t=1}^N [y(t)]^2. \quad (2)$$

The 2-gram model of energy detection can be expressed as:

$$\begin{cases} Y < \lambda, & H_0 \\ Y \geq \lambda, & H_1 \end{cases} \quad (3)$$

In AWGN channel, the distribution of the test statistic Y under two hypotheses is [6]:

$$Y \sim \begin{cases} \chi_{2\mu}^2, & H_0 \\ \chi_{2\mu}^2(2\gamma), & H_1 \end{cases} \quad (4)$$

Y follows a central (under H_0) and non-central (under H_1) chi-square distribution with 2μ degrees of freedom. μ shows the time-bandwidth product, γ shows the signal-to-noise ratio of the channel. When N is large enough, the central limit theorem can be employed to approximate the Y as Gaussian.

$$Y \sim \begin{cases} N(N\sigma_n^2, 2N\sigma_n^4), & H_0 \\ N(N(P + \sigma_n^2), 2N(P + \sigma_n^2)), & H_1 \end{cases} \quad (5)$$

where P is the average power of PU signal, then the target probability of false alarm P_{fd} and the probability of detection P_d can be given as [8]:

$$P_{fd} = P_r(P > \lambda | H_0) = Q\left(\frac{\lambda - N\sigma_n^2}{\sqrt{2N}\sigma_n^2}\right) \quad (6)$$

$$P_d = P_r(P > \lambda | H_1) = Q\left(\frac{\lambda - N(P + \sigma_n^2)}{\sqrt{2N}(P + \sigma_n^2)}\right) \quad (7)$$

where $Q(\bullet)$ is the standard Gaussian complementary cumulative distribution function. Based on the Constant False Alarm Rate (CFAR) approach, the threshold λ is set to meet a certain P_{fd} for CR system. Here, λ is got as follows:

$$\lambda = \sigma_n^2 \left(Q^{-1}(P_{fd}) \sqrt{2N} + \sqrt{N} \right) \quad (8)$$

The formula (8) indicates that λ is not only affected by P_{fd} , but also the noise variance σ_n^2 .

3 The Energy Detection Algorithm Based on the Signal CV

3.1 Coefficient of Variation

In probability theory and statistics, the coefficient of variation (CV), also called relative standard deviation, is a standardized measure of dispersion of a probability distribution or frequency distribution. CV can eliminate the effect of dimension and quantity of data. It is usually measured the non-stationary properties of a variable in statistical models. During spectrum sensing, the fluctuation of received signals is very different under two conditions (H_0 and H_1). Thus, this paper introduces CV to the analyses of the spectrum sensing and compared it with the conventional energy detection. In order to make $CV > 0$, it is defined as the ratio of the standard deviation σ_y to the mean μ_y and computed in this thesis as follows [6]:

$$\mu_y = \left| \frac{1}{N} \sum_{t=1}^N y(t) \right| \quad (9-1)$$

$$\sigma_y = \sqrt{\frac{1}{N} \sum_{t=1}^N (y(t) - \mu_y)^2} \quad (9-2)$$

$$CV = \frac{\sigma_y}{\mu_y} \quad (9-3)$$

In order to observe the different degree of received signal CV, the experiments use BPSK signal in AWGN to simulate the actual communication channel [7]. The simulation shows the value of received signals' CV is smaller when PU is the presence. When PU is absent, the values of certain CV are especially high. To better observe the subtle difference between H_0 and H_1 , the CV's maximum value is set to 30. In other words, the value is equivalent to 30 when $CV > 30$. The result of the simulation turns into Fig. 1.

Figure 1 shows the received signals' CV versus times of calculations in SNR = 8 dB when the CV's maximum value is set to 30. In SNR = 8 dB, the values of the CV are higher when PU is absent than present, the CV's maximum value is 0.4996 under H_1 , the CV's minimum value is 3.5069 under H_0 , and all the values of CV under H_0 are greater than it under H_1 .

To better observe the relationship of the SNR and CV, in SNR = 8 dB, we do the same experiments under same conditions. The result of the simulation turns into Fig. 2.

Figure 2 shows the received signals' CV versus times of calculations in SNR = -8 dB when the CV's maximum value is set to 30. In SNR = -8 dB, most of the CVs' values are higher when PU is absent than present, the CV's maximum value is 6.6564 under H_1 , and the CV's minimum value is 3.6141 under H_0 . Compared Fig. 1 with Fig. 2, a conclusion can be got that the volatility of the received signals slightly increase as the increase of the SNR. Whether SNR = -8 dB or SNR = 8 dB, PU is judged to be present if $cv < 3.6$. Set $\gamma = 3.0$, if $cv \leq \gamma$, PU is judged to be present. The 2-gram model of SCVED can be expressed as:

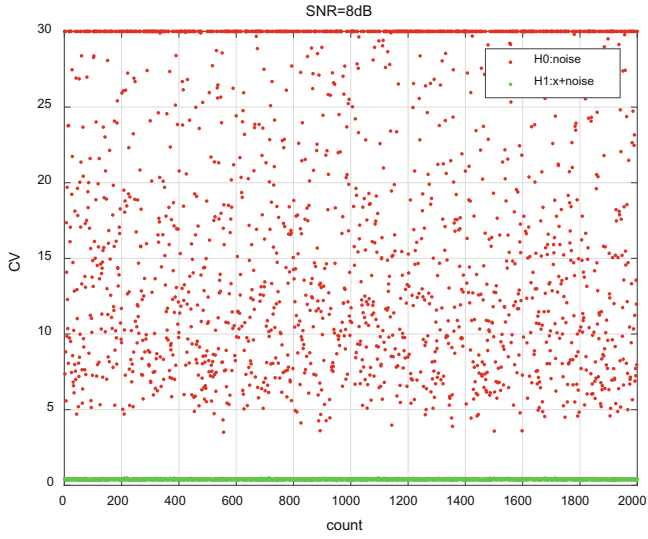


Fig. 1. The signals' CV versus times of calculations in SNR = 8 dB

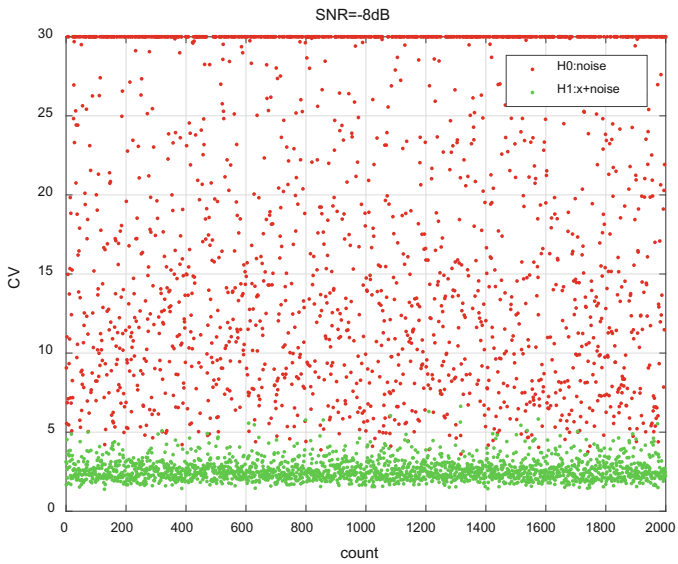


Fig. 2. The signals' CV versus times of calculations in SNR = -8 dB

$$\begin{cases} cv > \gamma, & H_0 \\ cv \leq \gamma, & H_1 \end{cases} \quad (10)$$

3.2 Overview Proposed Algorithm

In the proposed algorithm, the result of energy detection is revised by the signal CV, there is a joint detection of the energy detection and SCVED, and the SCVED is applied for the SU between the two thresholds.

The test statistic of the signal energy Y is collected in a given period of time and calculated by (2), which is compared to the threshold (λ). If $Y \geq \lambda$, it is the case H_1 , which means PU is present; If $Y < \lambda/2$, it is the case H_0 , which means PU is absent; if $\lambda/2 \leq Y < \lambda$, SCVED is conducted. The signals' CV is calculated according to the formula (9-1). If $cv \leq \gamma$, it is the case H_1 . Otherwise, PU is absent.

The procedures of the improved algorithms are as follows (Fig. 3):

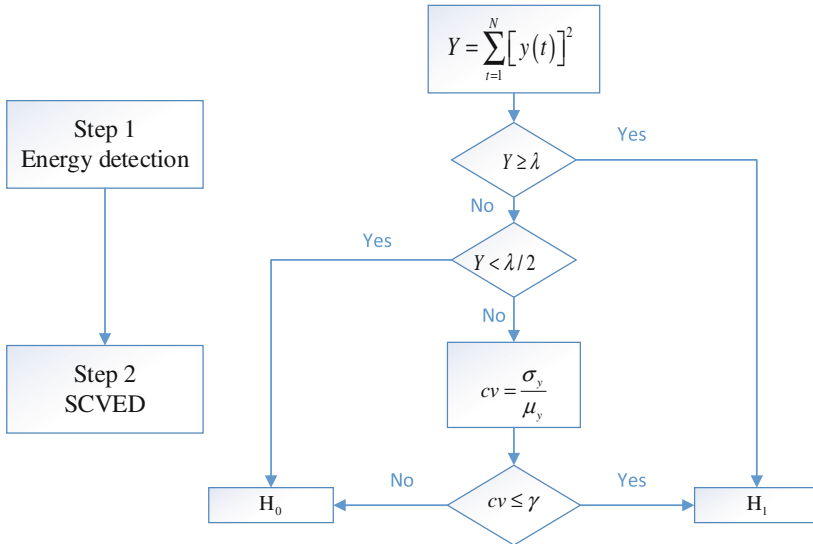


Fig. 3. Proposed algorithm flowchart.

Step 1 Energy detection: Generate a BPSK signal plus white noise or noise in a continuous time and calculate the test statistic Y according to the formula (2). Compare Y with the threshold λ , if $Y \geq \lambda$, PU is present. If $Y < \lambda/2$, PU is absent.

Step 2 SCVED: Calculate the signal CV according to the formula (9-1), if $cv \leq \gamma$, PU is present. Otherwise, PU is absent.

4 Simulation Analysis

The simulation platform uses BPSK signal in AWGN to simulate the actual communication environment [8, 9]. Besides, we vary SNR from -15 dB to 15 dB with a step of 1 dB in order to evaluate the proposed algorithm in different channel environment. And, the probability of false detection is set as 0.1 and 0.01 respectively. The result of the simulation turns into Figs. 4 and 5.

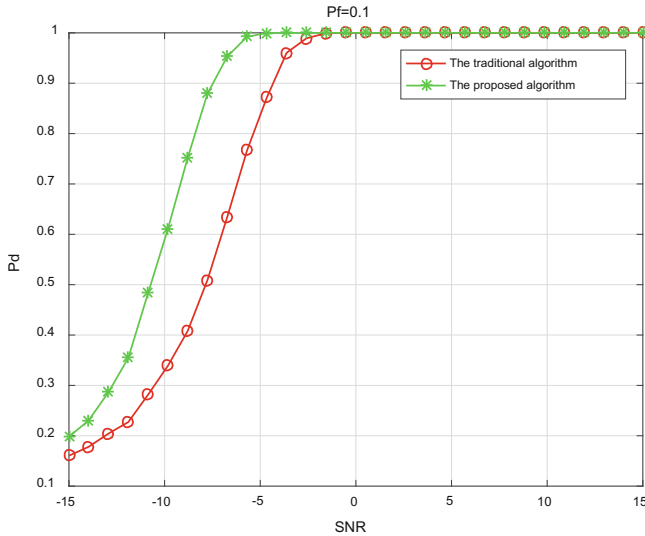


Fig. 4. The detection probability versus SNR at $P_{fd} = 0.1$.

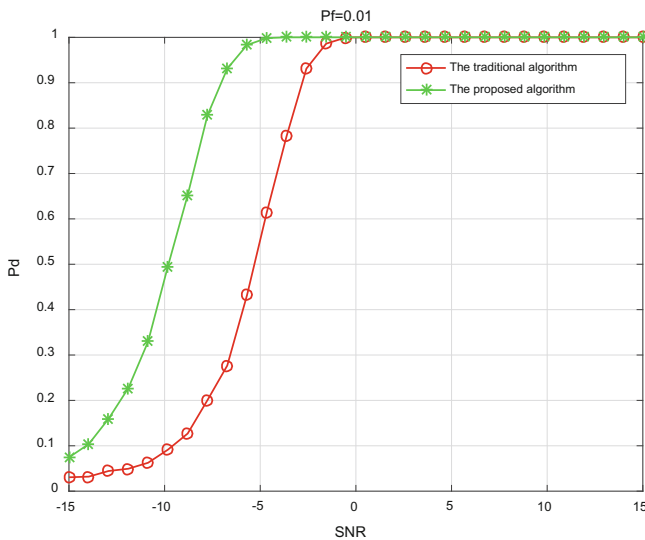


Fig. 5. The detection probability versus SNR at $P_{fd} = 0.01$.

Figure 4 shows the graph of detection probability versus SNR with $N = 128$ at $P_{fd} = 0.1$. The proposed algorithm performs better than the traditional algorithm. In detection performance, SCVED scheme shows 29% of improvement than the traditional under $\text{SNR} = -10$ dB.

Figure 5 shows the graph of detection probability versus SNR with $N = 128$ at $P_{fd} = 0.01$. The proposed algorithm performs remarkably better than the traditional algorithm. In detection performance, SCVED scheme shows 42% of improvement than the traditional under $\text{SNR} = -10$ dB. In the figures, we can see that the proposed algorithm has a greater improvement at low P_{fd} .

5 Conclusions

This paper proposes a new energy detection algorithm based on the variation coefficient of the received signals in the CR system. Based on the signal CV, the SCVED judges the status of PU on special frequencies and perfects the final result. Simulation results show that the joint algorithm can improve the detection probability, and reduce the influence of the noise and SNR. The threshold of step 2 is set based on abundant experiments and has not been proved through the formula. The next work is to investigate the accurate formula of the threshold.

References

1. Haykin, S.: Cognitive radio: brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **23**(2), 201–220 (2005)
2. Haykin, S., Thomson, J., Reed, H.: Spectrum sensing for cognitive radio. *Proc. IEEE* **97**(5), 849–877 (2009)
3. Urkowitz, H.: Energy detection of unknown deterministic signals. *Proc. IEEE* **55**(4), 523–531 (2005)
4. Han, W., Huang, C., Li, J., Li, Z., Cui, S.: Correlation-based spectrum sensing with oversampling in cognitive radio. *IEEE J. Sel. Areas Commun.* **33**(5), 788–802 (2015)
5. Sadhukhan, P., Kumar, N., Bhatnagar, M.R.: Improved energy detector based spectrum sensing for cognitive radio: an experimental study. In: 2013 Annual IEEE India Conference (INDICON), Mumbai, pp. 1–5 (2013)
6. Castellanos-Lopez, S., Cruz-Pérez, F.: Performance of cognitive radio networks under resume and restart retransmission strategies. In: 2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Wuhan, pp. 51–59 (2011)
7. Gardner, F.: A BPSK/QPSK timing-error detector for sampled receivers. *IEEE Trans. Commun.* **34**(5), 423–429 (1986)
8. Pandey, D., Dewangan, N.: Performance analysis of pilot assisted channel estimation in OFDM. In: 2015 International Conference on Computer, Communication and Control (IC4), Indore, pp. 1–6 (2015)
9. AbdelHamid, A., Zong, P.: A new software connector for distributed component simulation platforms. In: 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, pp. 25–28 (2015)

Performance Analysis for User-Centric Cloud Radio Access Network in Millimeter Wave

Yangying Zhang^{1,2}(✉), Hai Huang¹, Xiaojun Jing^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
zyangying@163.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ School of Engineering and Computer Science,
Oakland University, Rochester, USA

Abstract. Millimeter wave (mmWave) and cloud radio access network (C-RAN) are two potential candidates for next generation communication. In this paper, we consider user-centric C-RAN in mmWave with the existence of blockages in urban areas. The remote radio heads (RRHs) are deployed according to a Poisson point process in the circular region \mathcal{D} , of radius R . We employ the stochastic geometry theory to analyze the signal-to-noise ratio (SNR), rate, and outage probability. We emphasize the effect of circular region radius on the performance in this network and evaluate the effect with Monte Carlo simulations. The simulation results show that SNR, rate and outage probability have the same asymptotic trends and have the best performance when replace the circular region \mathcal{D} with the line-of-sight (LOS) circular region.

Keywords: Millimeter wave · Cloud radio access network
Stochastic geometry

1 Introduction

With the ever-increasing high data rates of mobile users, many modern network architectures and transmission technologies have been developed. User-centric cloud radio access network (C-RAN) is one of the promising network architectures [1], where a group of remote radio heads (RRHs) distributed uniformly serve the user U located at center. It saves the wireless resource by employing coordinated transmissions among RRHs. And baseband units (BBUs) pool employs centralized baseband processing for using computational resource efficiently [2]. We can expand serving region and improve the coverage and capacity by adding low cost RRHs to network. In addition, coordinated multipoint processing (CoMP) among RRHs can be deployed to limit the overall interference and improve the network capacity [3].

The millimeter wave (mmWave) communication has been considered as a promising solution to frequency congestion problem because of abundant available spectrum in mmWave. mmWave technology has been recognized as an enabler for next generation networks to achieve higher data rate and low delay [4]. Investigation of the use of mmWave technology is already underway. Due to susceptibility to blockages and severe attenuation in transmission, mmWave is more suitable for short-range wireless communication. C-RAN system with RRHs distributed densely can make full use of mmWave short-range communication characteristic.

Investigation of the user-centric cloud radio access networks is already underway. With explicit backhaul constraints and power constraints in user-centric C-RAN, the beamforming clustering scheme has been investigated [5]. Under the assumption of high signal-to-noise ratio (SNR) and certain path loss exponent, the outage probability has been investigated and closed form of analysis results have been derived in user-centric C-RANs with randomly deployed RRHs [6]. Generally, when the path loss exponent is uncertain in C-RANs, more accurate analytical results of outage probability and rate have been obtained by employing the Gaussian-Chebyshev integration [7]. The downlink outage probability of C-RAN has been investigated, where RRHs are distributed randomly with multiple antenna [3]. The macro base station and the RRHs cooperate and serve the user simultaneously. They employ maximal ratio transmission or transmit antenna selection under three downlink protocols, namely, selection transmission, all RRHs participating, and minimal number of RRHs participating [3].

In [3, 6, 7], the radii of disks, where user is located at the center, are set artificially. RRHs are uniformly distributed in these disks. The effects of blockages have not been considered. Moreover, how to select the radius and the influence of the radius have not been considered. When considering mmWave C-RAN, radius selection is very important for performance metrics such as signal to interference plus noise ratio (SINR), rate and outage probability. If we select larger radius, due to the effect of the blockages, more RRHs, especially including non-line-of-sight (NLOS) RRHs, will serve the user. while the NLOS RRHs will not improve the SNR and rate performance. If we select smaller radius, a lot of line-of-sight (LOS) RRHs will not serve the user, resulting in poor SNR and rate performance.

In this paper, we analyze the SINR, rate, and outage probability in downlink user-centric mmWave C-RAN. We describe the user-centric C-RAN system model, and channel model in Sect. 2. In Sect. 3, we analyze the cumulative distribution function of SINR and rate as well as outage probability. Simulation results are provided in Sect. 4. We summarize and propose the extension directions in Sect. 5.

2 System Model

We consider a downlink user-centric mmWave C-RAN system shown in Fig. 1(a), where a user U , with an omnidirectional antenna, is served by a group of RRHs deployed randomly over a circular region \mathcal{D} , of radius R in Fig. 1(b). The location of RRHs are assumed to obey a homogeneous Poisson point process (PPP) Φ_1 with

density λ_1 . The number of RRHs N in the circular region \mathcal{D} is Poisson random variable whose distribution law is

$$\Pr\{N \text{ RRH in LoS circle}\} = \frac{A^N}{\Gamma(N + 1)} e^{-A}. \quad (1)$$

where A is the intensity and $A = \pi R^2 \lambda_1$. All RRHs employ antenna arrays to achieve higher directional beamforming gain.

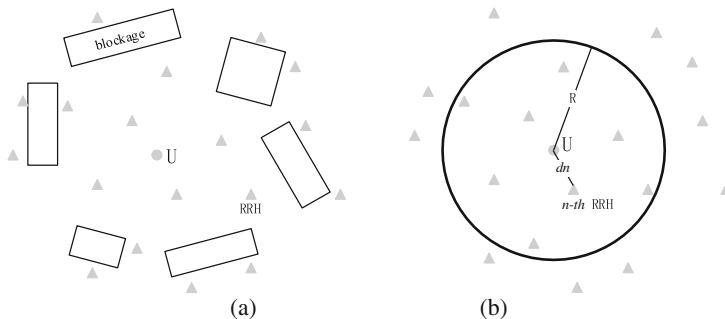


Fig. 1. User-centric mmWave C-RAN

The two fundamental physical differences between the mmWave and (ultrahigh frequency) UHF network are the need for significant directionality and vulnerability to blockages [8]. For tractability, the antenna arrays of RRHs are modeled as sectorized antenna model $G_{M,m,\theta}(\phi)$ [9]. Let M , m are main lobe directivity gain and back lobe gain, respectively. θ is the beam width of the main lobe. We assume the directivity gain between the n -th RRH and the user is M .

The blockages, assumed to be impenetrable, are modeled as Boolean scheme of rectangles [10], whose centers form a homogeneous PPP Φ_2 with density λ_2 , independent of Φ_1 . The lengths L and widths W of blockages are identical independent distributed.

We employ the path loss model $1/(1 + d_n^\alpha)$, which avoids the singularity issue when $d_n \rightarrow 0$. We assume the path loss exponent of LOS propagation is $\alpha = 2$. Assuming Rayleigh fading between the n -th RRH and the user is $g_n \sim \mathcal{CN}(0, \mu)$ and $\mu = 1$ denotes the mean power of the channel. Then $|g_n|$ is a normalized Gamma random variable. The distance d_n between the n -th RRH and the user U is a random variable with distribution function

$$f_{d_n}(x) = \begin{cases} \frac{2x}{R^2} & ; 0 \leq x \leq R \\ 0 & ; \text{else} \end{cases}. \quad (2)$$

Assuming that each RRH serves only a user per unit time, and other users must wait until the RRH finish service. And BBU employs the coordination transmission control, which means that the user U associates with a set of cooperating RRHs and we can ignore the interference between the cooperating RRHs.

3 Performance Analysis

In this section, we analyze the SNR, rate and outage probability under the preceding proposed model. First, we evaluate the effect of the blockages, and approximately replace the circular region \mathcal{D} with the LOS circular region, of radius R_{LOS} . Then we analyze the SNR received by the user U, rate and outage probability.

3.1 Radius of LOS Circular Region

In Fig. 1(a), the LOS region observed by the user U has an irregular shape [10]. The LOS region can be approximated by a circular region which has a certain radius. Only RRHs inside this region are considered as LOS propagation to the user.

The LOS propagation through a link of length x means that there is no blockages cross the link, and the probability of LOS propagation is $p(x) = e^{-\beta x}$ in [10], where $\beta = 2\lambda_2(E[L] + E[W])/\pi$, and λ_2 , $E[L]$, $E[W]$ are the density of blockage, the average length and average width, respectively.

Based on the LOS probability function $p(x)$, the radius of the circular region [10] can be derived as

$$R_{LOS} = \left(2 \int_0^{\infty} p(x) dx \right)^{0.5}. \quad (3)$$

We equivalently approximate the LOS region by a fixed circle $B(U, R_{LOS})$ with user U located at the center, which we define as equivalent LOS circular region [10].

3.2 SNR Analysis

Based on the assumption that one RRH can at most serve one user and the effects of blockages, we can ignore inter-RRHs interference. The SINR received by user can be approximated to SNR. We denote the SNR from the n -th RRH to user as γ_n , the overall SNR received by the user U is given as [7]

$$\gamma = \sum_{n=1}^N \gamma_n. \quad (4)$$

We consider both of small scale Rayleigh fading and path loss. Thus, the received SNR from the n -th RRH to the user U can be written as [6]

$$\gamma_n = \frac{M|g_n|^2}{\sigma^2(1 + d_n^\alpha)}. \quad (5)$$

where σ^2 , α are the noise power received at the user and the path loss exponent, respectively. The overall SNR received by the user U is given as [7]

$$\gamma = \sum_{n=1}^{N_1} \gamma_n = \frac{M}{\sigma^2} \sum_{n=1}^{N_1} \frac{|g_n|^2}{(1 + d_n^z)}. \quad (6)$$

where N_1 is the number of RRHs whose signal can be received by user U. If the radius R is shorter than R_{LOS} , this means that not all LOS RRHs serve the user U. If R is larger than R_{LOS} , there are NLOS RRHs serve the user U while the signal strength from these RRHs come close to zero received by U. Thus N_1 is Poisson random variable.

$$\Pr\{N_1 = k\} = \frac{A^k}{k!} e^{-A}, \quad k = 0, 1, 2, \dots, A = \begin{cases} \pi R_{LOS}^2 \lambda_1, & R \geq R_{LOS} \\ \pi R^2 \lambda_1, & R < R_{LOS} \end{cases}, \quad (7)$$

The cumulative distribution function (CDF) of SNR can be defined as

$$F_\gamma(t) = P\{\gamma < t\} = P\left\{\frac{M}{\sigma^2} \sum_{n=1}^{N_1} \frac{|g_n|^2}{(1 + d_n^z)} < t\right\}, \quad (8)$$

where the t is the threshold. When the radius R is smaller than R_{LOS} , only a part of LOS RRHs provide service with the user U. With the increase of R , the SNR received by U will increase. If R is larger than R_{LOS} , all LOS RRHs and a part of NLOS RRHs serve the user U simultaneously. Due to the blockages, the SNR will not increase apparently. We will evaluate the effect of R on the CDF of SNR in Sect. 4.

3.3 Rate Analysis

The rate distribution is vital for assessing the performance of this network. According to the definition, the rate received by U can be written as $R_U = B \log_2(1 + \gamma)$, where B denotes the bandwidth. The cumulative distribution function of R_U can be defined as

$$F_{R_U}(\tau) = P\{R_U < \tau\} = P\{B \log_2(1 + \gamma) < \tau\}. \quad (9)$$

Based on CDF of the SNR, we can deduce the CDF of the rate according to the definition. The effect of R on the rate has the same trend as the SNR. We will show the simulation results of rate in Sect. 4.

3.4 Outage Probability Analysis

The outage probability is one of key performance metrics. In this section, we qualitatively analyze the influence of radius R on outage probability. We denote R_e as targeted data rate received by U. And the outage probability is defined as

$$\begin{aligned} P_{out} &= P\{R_U < R_t\} = P\{B \log_2(1 + \gamma) < R_t\} \\ &= P\left\{\gamma < 2^{\frac{1}{B} R_t} - 1\right\} = \sum_{n=0}^{\infty} F_\gamma(2^{\frac{1}{B} R_t} - 1 | N_1) P\{N_1 = k\}. \end{aligned} \quad (10)$$

4 Simulation Results

In this section, we present Monte Carlo simulation results, and illustrate impact of the radius R_{LOS} on SNR CDF, rate CDF and the outage probability.

We assume that the RRHs density is $\lambda_1 = 3 * 10^{-2}/m^2$, and the bandwidth allocated to user is $B = 100$ MHz, the antenna gain transmitted from RRHs is $M = 10$ dB. The path loss exponent of LOS propagation is $\alpha = 2$. The density of blockages is $\lambda_2 = 4.4 * 10^{-3}/m^2$, $E[L] = E[W] = 15$ m.

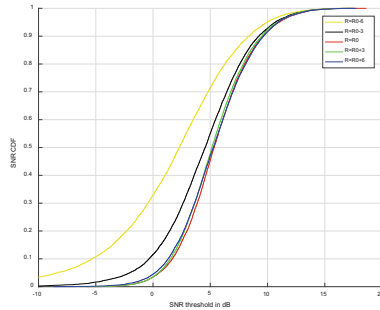


Fig. 2. SNR CDF with different radius. R denotes the radius of circular region. And R_0 denotes the radius R_{LOS} of LOS region corresponding to red curve. The yellow and black curve correspond to the case where R is smaller than R_{LOS} , where the other two is for a larger R .

First, we compare the SNR CDF with different radius of circular region in Fig. 2. With the increase of the radius, the CDF of SNR improve rapidly until the radius approaches to R_{LOS} . When the radius surpass R_{LOS} , the CDF of SNR does not change apparently with the increase of radius. The simulation results demonstrate the analysis in Sect. 3. The RRHs outside of LOS circular region do not improve SNR received by the user.

We present the rate CDF with different circular region radius in Fig. 3.

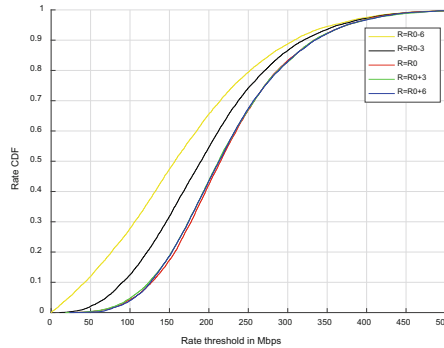


Fig. 3. Rate CDF with different radius.

With the increase of R , the CDF of rate increase until approaching to R_{LOS} . When R is larger than R_{LOS} , the performance do not improve apparently. The CDF of rate has the same trend as the CDF of SNR with the increase of R .

We present the outage probability with different radius and different targeted data rate in Fig. 4.

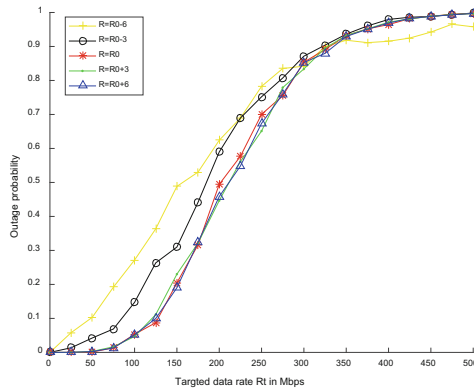


Fig. 4. Outage probability

With the increase of expected data rate, the outage will be more likely to happen. The effect of R on outage probability is same as effect on SNR and rate. Therefore, the system performance will be optimal when the disk radius is equal to the radius of LOS circular region.

5 Conclusions

In this paper, we analyze the SNR, rate, and outage probability based on the stochastic geometry theory in the user-centric mmWave C-RAN considering the blockages. Compared to the existing work, we emphasize the effect of radius on the performance in this network and evaluate the effect with Monte Carlo simulations. This work can be extended in following directions. 3-D antenna gain pattern, which extend the directivity in elevation angles, could be further investigated. The case, where one RRH can serve multi-users simultaneously, could also be investigated in future work.

Acknowledgement. This work is supported by NSFC Project 61471066.

References

1. Chen, N., Rong, B., Zhang, X., Kadoch, M.: Scalable and flexible massive MIMO precoding for 5G H-CRAN. *IEEE Wirel. Commun.* **24**(1), 46–52 (2017)
2. Quek, T., Peng, M., Simeone, O., Yu, W.: *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge University Press, Cambridge (2017)

3. Khan, F., He, H., Xue, J., Ratnarajah, T.: Performance analysis of cloud radio access networks with distributed multiple antenna remote radio heads. *IEEE Trans. Signal Process.* **63**(18), 4784–4799 (2015)
4. Omar, M., Anjum, M., Hassan, S., Pervaiz, H., Niv, Q.: Performance analysis of hybrid 5G cellular networks exploiting mmWave capabilities in suburban areas. In: 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, pp. 1–6 (2016)
5. Dai, B., Yu, W.: Sparse beamforming and user-centric clustering for downlink cloud radio access network. *IEEE Access* **2**, 1326–1339 (2014)
6. Ding, Z., Poor, H.: The use of spatially random base stations in cloud radio access networks. *IEEE Signal Process. Lett.* **20**(11), 1138–1141 (2013)
7. Yang, Z., Ding, Z., Fan, P.: Performance analysis of cloud radio access networks with uniformly distributed base stations. *IEEE Trans. Vehicular Technol.* **65**(1), 472–477 (2016)
8. Andrews, J., Bai, T., Kulkarni, M., Alkhateeb, A., Gupta, A., Heath, R.: Modeling and analyzing millimeter wave cellular systems. *IEEE Trans. Commun.* **65**(1), 403–430 (2017)
9. Hunter, A., Andrews, J., Weber, S.: Transmission capacity of ad hoc networks with spatial diversity. *IEEE Trans. Wirel. Commun.* **7**(12), 5058–5071 (2008)
10. Bai, T., Vaze, R., Heath, R.: Analysis of blockage effects on urban cellular networks. *IEEE Trans. Wirel. Commun.* **13**(9), 5070–5083 (2014)
11. Bai, T., Heath, R.: Coverage and rate analysis for millimeter-wave cellular networks. *IEEE Trans. Wireless Commun.* **14**(2), 1100–1114 (2015)

Millimeter Wave Cloud Radio Access Network Coverage and Capacity

Jinxia Hu^{1,2(✉)}, Xiaojun Jing^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications University, Beijing, China
jinxiahu2016@163.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China

³ School of Engineering and Computer Science, Oakland University, Rochester, USA

Abstract. In this paper, the performance of a cloud radio access networks (C-RAN) where remote radio heads (RRHs) operating at millimeter wave (mm-Wave) range are modeled as a homogeneous Poisson point process has been investigated. In this network, due to the high frequency and small wavelength, we use a LOS circle concept from the former study. Considering downlink transmission, we compared the performance of two different transmission schemes. First is the selection transmission which chooses the best RRH (BR) for transmission. Another scheme is all the RRHs participated (ARP) and transmission the signal together to the typical user. We analysis the outage and capacity of the two schemes and ARP transmission scheme always performed better.

Keywords: Cloud radio access network · Millimeter wave
Remote radio head

1 Introduction

With the explosive growth of mobile traffic demand, the contradiction between spectrum shortage and capacity requirements becomes increasingly prominent. Wireless bandwidth becomes a critical problem for the next generation wireless network, so that recently a lot of research was focused on the high frequency communication. The millimeter wave communication was becoming a promising technology for the future wireless communication system. Due to the high frequency and a small wavelength, mm-Wave communication will suffer from huge propagation loss. To solve this problem, beamforming has been adopted as an essential technique to a directional communication [1]. mm-Wave also experiences penetration loss and sensitive to blockage by obstacles such as building and human body [2]. Therefore mm-Wave communication always considered to be a short distance and directional transmission.

With the increasingly enhanced application scenario of mobile internet, 5G should have the characteristic of higher experiences data-rate and wider bandwidth to support multimedia contents featured by higher definition and living experiences. In recently

5G communication, many scholars has been proposed cell densification aimed to improve the capacity and area spectral efficiency.

The future 5G communication architecture will be an ultra-dense network by increasing the cell base station density to improve network capacity [3]. But at the same time it will cause highly overall interference which resulting in a limit capacity gain. The characteristic of the cloud radio access networks (C-RANs) make it become a candidate architecture in the future 5G network. The rationale behind this is that baseband processing is centralized and coordinated among sites in the centralized BBU pool, and thus it not only achieves significantly higher data rates than conventional cellular networks but also reduces the capital expenditure (CAPEX) and operating expenditure (OPEX) of the networks [4]. The ideal of the C-RANs is to move the baseband units (BBUs) to a central location/data center and connect it to the radio units, also called remote radio heads (RRHs), via optical fibers [5].

Motivated by the aforementioned background, in this study, we proposed mm-Wave C-RANs architecture. We consider the downlink of a C-RANs where RRHs operating at the mm-Wave connected with BBUs via optical fiber link. The BBUs pool in the cloud is established to coordinate the entire network. In our system, the coordinated multipoint (COMP) has adopted to mitigate the inter-RRHs interference and improve the network capacity [6]. We adopted a LOS circle concept from the former study [7] and use different user association to analysis the performance of the network respectively.

The paper is structured as follows: in Sect. 2 we described the system model and we derive the coverage probability and outage probability expression in Sect. 3. Performance analysis simulation result were presented in Sect. 4 and we draw a conclusion in Sect. 5.

2 System Model

2.1 Network Model

As show in the Fig. 1, we assume a typical user is located in the origin associated by the RRHs which distributed randomly over a circular region K with radius R . Due to severe penetration loss in mm-Wave communication, we only consider the RRHs in the circular which called LOS circle. We assumed that as long as R is larger enough, the RRHs which outside the circular regime can be thought NLOS. RRHs operating at mm-Wave range and the location of the RRHs form a homogenous Poisson Point process (PPP) Φ_R with intensity λ_R . Further, we assume that each RRH transmits with the same power (P_R). We use the method of coordinated multipoint processing (COMP) in C-RAN to get the purpose of mitigating the overall interference and improving the performance of the network.

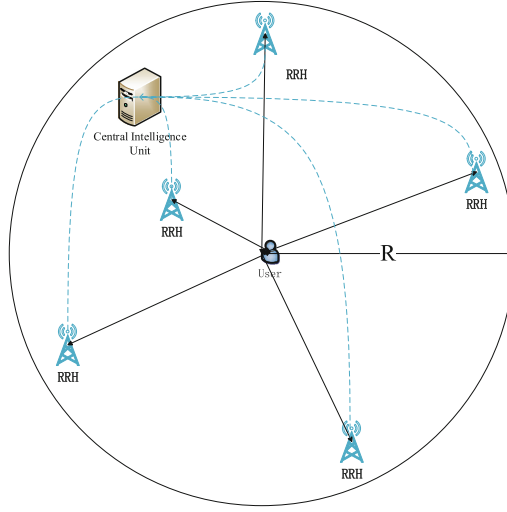


Fig. 1. System model.

2.2 Directional Beamforming

In this network, all RRHs equipped with directional antennas which modelled as a sectorized gain pattern. The antenna gain provide for a typical users was modeled as the following expression [8],

$$G_R(\theta) = \begin{cases} G_{\max} & \text{if } |\theta| \leq \theta_R \\ G_{\min} & \text{otherwise} \end{cases}, \quad (1)$$

where θ is the steering angle and θ_R represents the beam-width or main lobe width. It is assumed that when the antenna beams of intended access link are aligned, then the effective gain which desired access link achieve is G_{\max} [8]. Although the user antenna gain pattern can be modelled in the same manner, we considered the omnidirectional antennas for the user in this network.

2.3 Distance-Dependent Path Loss Model

In our work, the path loss between RRHs or MBS with typical user was estimated as [9]

$$PL(d)_{dB} = \alpha + 10\beta \log_{10}(d) + \zeta_\sigma, \quad (2)$$

where α represents the path loss and β is the path loss exponent. The ζ_σ is modelled as the shadowing effect in dB and σ is the standard deviation of ζ_σ . Motivated by previous works [8, 9], different LOS and NLOS access link experience different propagation environment so that the path loss parameters for the two access link was different. In

this paper, we adopted the equivalent LOS ball that only consider the RRHs in the circle region.

2.4 SINR Model

For best RRH transmission, the receive SINR at a typical user, denoted by γ_{BR} is written as

$$\gamma_{BR} = \frac{P_R G_R(\theta) h_R d^{-\alpha}}{N_0 + I_R}, \quad (3)$$

since coordinated multipoint (COMP) has adopted to mitigate the inter-RRHs interference, the inter-tier interference from RRHs can be ignored thus the SINR can be simplified as

$$\gamma_{BR} = \frac{P_R G_{BR}(\theta) h_{BR} d^{-\alpha}}{N_0}, \quad (4)$$

where P_R is the transmission power at each RRH, $G_R(\theta)$ as a function of θ_R which represents the antenna gain for a typical user, d is the distance from the served RRH to the typical user and $d^{-\alpha}$ represents the path loss gain. $h_R \sim \exp(1)$ is the small-scale fading channel power gain and N_0 is the noise power.

For all RRHs participate in transmission, the SINR can be written as

$$\gamma_{AR} = \sum_{i=1}^N \frac{P_R G_{R_i}(\theta) h_{R_i} d_i^{-\alpha}}{N_0}, \quad (5)$$

where N is a random variables that represents the number of RRHs in the LOS circle which follows the Poisson distribution. Note that the SINR in (3) (4) is a random variable, because the locations of the RRH R_i is randomness, and the small-scale fading h_{R_i} and the directivity gain $G_{R_i}(\theta)$ were depended on it. We Using the aforementioned system model and evaluate the mm-Wave C-RAN coverage and outage probability in the following section.

3 Performance Analysis

In this section, we derive the coverage probability and outage probability expression for the downlink mm-Wave C-RAN system under two different transmission schemes.

3.1 Coverage Probability

The downlink SINR coverage probability is defined as

$$P_C(T, \alpha) = \Pr[\text{SINR} > T], \quad (6)$$

which means that the probability of the a randomly user will attain a target SINR T or the average fraction of users who at any time achieve SINR T [7].

- (1) Coverage of the best RRH transmission scheme can be expressed as

$$\begin{aligned} P_{BR}(T, \alpha) &= \Pr(\gamma_{BR} > T) \\ &= \Pr\left(\frac{P_R G_{BR}(\theta) h_{BR} d^{-\alpha}}{N_0} > T\right). \end{aligned} \quad (7)$$

- (2) Coverage of the all RRHs jointly transmission can be expressed as

$$\begin{aligned} P_{AR}(T, \alpha) &= \Pr(\gamma_{AR} > T) \\ &= \Pr\left(\sum_{i=1}^N \frac{P_R G_{R_i}(\theta) h_{R_i} d_i^{-\alpha}}{N_0} > T\right). \end{aligned} \quad (8)$$

3.2 Outage Probability

- (1) Best RRH for transmission (BR) scheme: In this scheme, we choose the RRH with the best channel for transmission. Since we only consider the LOS propagation, the nearest RRH will be have best channel for transmission. The outage probability is defined as

$$P_{out} = \Pr[\log_2(1 + SINR) < R]. \quad (9)$$

It can be thought of the probability that a user have not achieve a target rate. The outage event will occurs when all the RRH which has the best channel are in outages. Assume that the number of the RRHs in the circular region K is N , the outage probability for the typical user can be expressed as

$$\begin{aligned} P_{BR}(R) &= \Pr(\log_2(1 + \gamma_{BR}) < R) \\ &= \left(1 - \exp\left(-\frac{N_0}{P_{BR} G_{BR}(\theta)} d^\alpha (2^R - 1)\right)\right). \end{aligned} \quad (10)$$

- (2) All RRHs participate in transmission (ARP) scheme: In this scheme, all the RRHs in the circular region K are severed for a typical user. Therefore, the outage events occur when the overall rate from the RRHs is in outage. Assuming that all the N RRHs transmission the signal for the typical, the outage probability can be given as

$$P_{AR}(R) = \Pr(\log_2(1 + \gamma_{AR}) < R). \quad (11)$$

4 Simulation Result

In this section, numerical simulation results are shown to corroborate the derived analytical results. We assume that the mm-Wave C-RAN is operated at the carrier frequency

30 GHz with transmit power $P_R = 30$ dBm and assigned 2 GHz of bandwidth to each user in this paper. The interference between RRHs are ignore due to the COMP adopted to this network. We consider a LOS only downlink transmission and the LOS path loss exponent considered to be $\alpha_{LOS} = 2$.

Figure 2 reveals the probability of coverage of both the best RRH transmission and all RRHs transmission schemes with varying the SINR threshold. It can be seen that the coverage probability of the BR and ARP schemes are decrease with increase SINR threshold. But the ARP scheme has higher coverage probability than the BR scheme. In mm-Wave C-RAN, since the interference between the RRHs can be mitigate due to the COMP technology. It worth to consider how many the number of the RRHs served a user can achieve an optimal effective in the network. We are not only consider the coverage and rate in one user, but also take account of the overall network performance. The density of the RRHs in a user-centric C-RAN would be further study in the future.

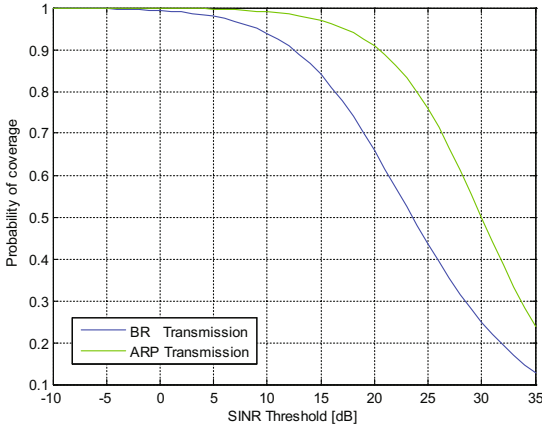


Fig. 2. SINR CCDF when $\lambda = 5 \times 10^{-5} \text{m}^{-2}$.

Figure 3 show the effects of RRH density on coverage probability. We observe when more RRHs are deployed, there is a substantial increase in the coverage probability with both BR transmission scheme and ARP transmission scheme. However, the ARP transmission scheme always achieve higher coverage probability than the BR transmission scheme. In traditional cellular network, with the density of the BSs increased, the coverage probability will through an increased and then rapid due to the inter-tier interference between the BSs. In mm-Wave C-RAN, the interference between the RRHs can be mitigated by COMP, so increase the density can be greatly enhance system performance.

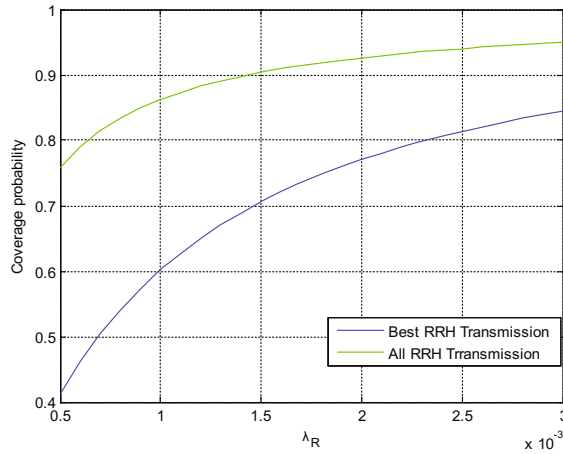


Fig. 3. Effects of density on coverage probability.

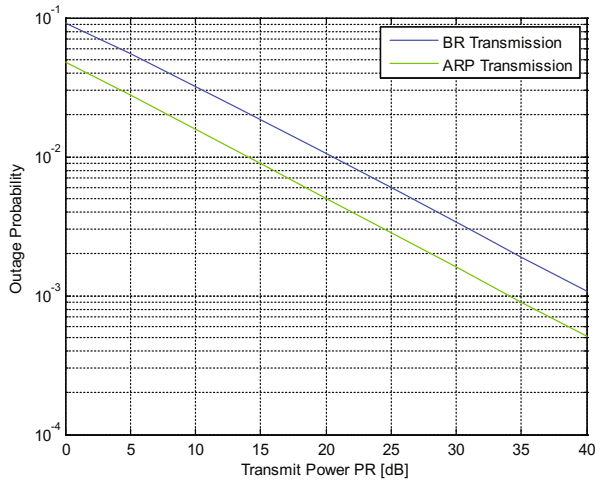


Fig. 4. Outage probability for BR and BRP scheme.

From the Fig. 4, we can see the tendency of outage probability of both transmission schemes with different transmission power. It reveals that the outage probability of both schemes show a decreasing trend with the increase transmission power. The outage probability of the ARP transmission scheme is lower compared to the BR scheme.

Figure 5 show the effect of RRHs density on outage probability. It is obvious that with more RRHs are deployed, there is a substantial decrease in the outage probability and the ARP scheme has a lower outage probability compared to BR scheme.

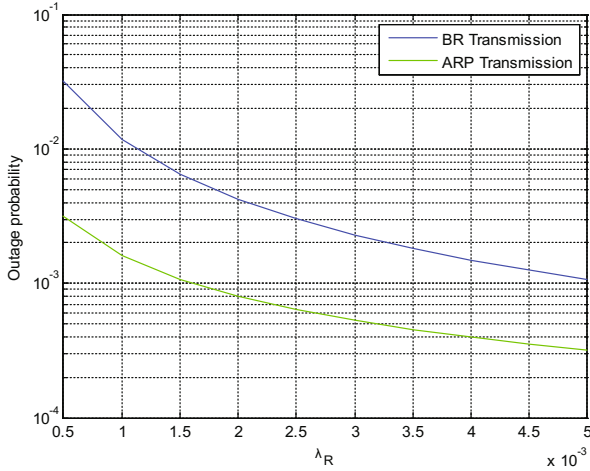


Fig. 5. Effects of density on outage probability.

5 Conclusions

In this paper, the downlink performance of mm-Wave C-RAN with different transmission schemes was investigated. We used a LOS circle concept that only consider the RRHs in the region transmission the information to a user. In this system, the performance of BP and ARP schemes were analyzed. We also derived the analytical expressions for the coverage probability and outage probability and it were validated through numerical simulation. As the result showed, that ignored the interference between the RRHs, the performance of the system will be highly increased and the ARP scheme was outperformed the BR scheme.

Acknowledgement. This works thanks to the project 61471066 supported by NSFC.

References

1. Rappaport, T.S., Sun, S., Mayzus, R., et al.: Millimeter wave mobile communications for 5G cellular: it will work! *IEEE Access* **1**(1), 335–349 (2013)
2. Bai, T., Desai, V., Heath, R.W.: Millimeter wave cellular channel models for system evaluation. In: *International Conference on Computing, Networking and Communications*, pp. 178–182. IEEE (2014)
3. Marsch, P., Raaf, B., Szufarska, A., et al.: Future mobile communication networks: challenges in the design and operation. In: *Mobile Satellite Communication Networks*, p. 196. Wiley (2012)
4. Checko, A., Christiansen, H.L., Yan, Y., et al.: Cloud RAN for mobile networks—a technology overview. *IEEE Commun. Surv. Tutor.* **17**(1), 405–426 (2015)
5. Wang, X., Huang, Y., Cui, C., et al.: C-RAN: evolution toward green radio access network. *China Commun.* **7**(3), 107–112 (2010)

6. Hu, R.Q., Qian, Y.: An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems. *IEEE Commun.* **52**(5), 94–101 (2014)
7. Bai, T., Alkhateeb, A., Heath, R.W.: Coverage and capacity of millimeter-wave cellular networks. *IEEE Commun. Mag.* **52**(9), 70–77 (2014)
8. Singh, S., Kulkarni, M.N., Ghosh, A., et al.: Tractable model for rate in self-backhauled millimeter wave cellular networks. *IEEE J. Sel. Areas Commun.* **33**(10), 2196–2211 (2015)
9. Rangan, S., Rappaport, T.S., Erkip, E.: Millimeter-wave cellular wireless networks: potentials and challenges. *Proc. IEEE* **102**(3), 366–385 (2014)

Optimal Downtilts for 3D Beamforming Based on Greedy Algorithm in Massive MIMO Networks with Imperfect CSI

Kang Zhang^{1,2,3(✉)}, Kwanhak Jon^{4,5}, and Songlin Sun^{1,2,3}

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
1925360967@qq.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ National Engineering Laboratory for Mobile Network Security,
Beijing University of Posts and Telecommunications, Beijing, China

⁴ School of Electronic Engineering, Beijing University of Posts
and Telecommunications, Beijing 100876, China

⁵ School of Communication Engineering, Kim Chaek University of Technology,
Pyongyang 999093, Democratic People's Republic of Korea

Abstract. Massive multi-input multi-output (MIMO) improves networks throughput via using plenty of antennas at the base station (BS) to serve multi-user simultaneously. Considering channel hardening, we directly investigate the approximate expressions for network throughput with conjugate beamforming (CBF) and zero-forcing beamforming (ZFBF) schemes and adaptive downtilts of antenna patterns. Building on this, a new adaptive beam switching tilting strategy based on the greedy algorithm which adopts the optimal downtilts to serve different regions are proposed to enhance the performance. We also study intercell interference and propose a new modified cooperative approach to reduce interference. Simulation results demonstrate that the system throughput has been improved compared with the traditional method.

Keywords: Massive MIMO · 3D antenna pattern · Greedy algorithm

1 Introduction

In 5G, Massive MIMO is considered as a key technology due to various remarkable advantages [1]. Moreover, the orthogonality of random channel vectors proved by theoretic analysis and experimental results can be applied to enhance physical security and weaken inter-cell interference [2, 3]. Considering the horizontal plane, the antenna arrays at the base station (BS) only radiate a fixed downtilt beam only that is considered by most researchers in numerous studies on beamforming algorithms [3, 4]. The main beam can be controlled by 3-dimension (3D) beamforming technique in elevation and azimuth dimensions [5, 6]. But most of these works are applied in regular MIMO networks, not massive MIMO networks. Switched-beam tilting at the system level is

studied in the literature [6], but a specific design method has not been provided yet. Intercell interference mainly limits the achievable throughput of active users especially users who are in a cell-edge region. Multiple base stations (BSs) coordination called multicellular processing can combat intercell interference efficiently [7]. Traditional coordinated beamforming cannot reduce many interference BSs simultaneously in the horizontal plane and solutions for designing beamformers are iterative [7–9].

A new adaptive beam switching cooperation strategy based on the greedy algorithm are proposed to optimize throughput. Considering channel hardening, the approximate expressions for the ergodic rate with ZFBF and CBF are applied to evaluate performance. Next, the greedy algorithm is adopted to search the adaptive tilt to further improve throughput, and also propose one method to suppress interference.

The organization is described as follows. The networks system model introduced in this paper is described in Sect. 2. Section 3 investigates approximate expressions for user ergodic rate with imperfect channel state information (CSI). A novel technique for 3D beamforming is proposed and analytical numerical results in next Section. At last, we conclude this paper.

Notations: conjugate operator is presented as $(\cdot)^*$, $(\cdot)^H$ denotes Hermitian transpose operator, identity matrix is presented as I_N , Euclidean norm is denoted as $\|\cdot\|$.

2 System Model

There are B adjacent cells in the cellular network. They are indexed with $b = 1 \dots B$. Only one BS is located in each cell. The number of antennas employed at BS is N_t . The height applied in BS is h_{bs} . This paper mainly focuses on downlink transmission, assuming universal frequency reuse. For users, the height is denoted as h_u . Users are uniformly distributed and only have one antenna. The network configuration can be seen in Fig. 1.

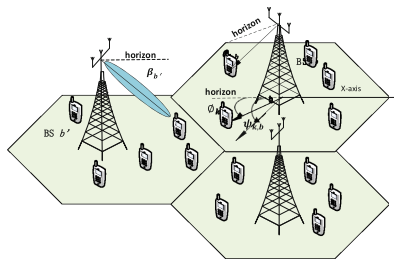


Fig. 1. The system model

2.1 Antenna Radiation Pattern in 3D Channel Model

Active Antenna System (AAS) is employed in the BS antenna array [7]. The authors investigate the antenna gain in [8]. The antenna gain is described as follow:

$$A_{k,b}^{\text{dBi}} = -\min \left\{ \left(\min \left[12 \left(\frac{\vartheta_{k,b} - \varphi_{k,b}}{\vartheta_{3\text{dB}}} \right)^2, \text{SLL}_{\text{m}} \right] + \min \left[12 \left(\frac{\theta_{k,b} - \beta_b}{\theta_{3\text{dB}}} \right)^2, \text{SLL}_{\text{el}} \right] \right), \text{SLL}_{\text{tot}} \right\} \quad (1)$$

For BS b , $\vartheta_{k,b}$ presents angle applied in horizontal measured between user k and x -axis. $\theta_{k,b}$ denotes the tilt measured between user k and horizon planes. This tilt between the peak beam and the horizon is defined as β_b , and the fixed angle is φ_b . The side-lobe attenuation is presented as SLL_{el} . The front-to-back and the total attenuation are respectively SLL_{m} and SLL_{tot} . Considering horizontal and vertical, we respectively present the HPBW as $\vartheta_{3\text{dB}}$ and $\theta_{3\text{dB}}$.

2.2 The Downlink Signal Model and Propagation Environment

From user k to BS b , we write channel vector as $\alpha_{k,b}(\beta_b)h_{k,b}$ where $h_{k,b} \in \mathbb{C}^{N_t \times 1}$ denotes the small-scale fading channel vector [8], and $\alpha_{k,b}(\beta_b) = \text{PL}_{k,b}A_{k,b}(\beta_b)$, where $\text{PL}_{k,b}$ denotes the distant-dependent large-scale fading, and $\alpha_{k,b}$ denotes the antenna gain. $h_{k,b} \sim \text{CNO}, I_{N_t}$, $\forall k, b$ are assume to be i.i.d. We operate the networks in the time division duplexing (TDD) mode, the downlink CSI can be learned via the uplink transmission. $h_{k,b}$ can be described as follow

$$h_{k,b} = \hat{h}_{k,b} + e_{k,b} \quad (2)$$

The error for estimate denotes as $e_{k,b} \sim \text{CN}(0, \sigma_{k,b}^2(\beta_b)I_{N_t})$, $e_{k,b}$ is i.i.d., where $\sigma_{k,b}^2(\beta_b) = 1 / \left(1 + \alpha_{k,b}(\beta_b)P' \right)$ and P' presents the power for the uplink pilot. The estimate channel vector can be described as follow

$$\hat{h}_{k,b} = \frac{\sqrt{P' \sigma_{k,b}^2(\beta_b)}}{P' \sigma_{k,b}^2(\beta_b) + 1} y_{k,b} \quad (3)$$

For user k at BS b , the power for entry of $\hat{h}_{k,b}$ is $\mu_{k,b}^2(\beta_b) = 1 - \sigma_{k,b}^2(\beta_b)$. Data is transmitted via using preceding schemes. Learning the CSI, ZFBF and CBF are both investigated. The beamforming vectors is denoted as $w_{k,b} \in \mathbb{C}^{N_t \times 1}$ and $d_{k,b}$ denotes useful data. We write the received signal as follow

$$y_{k,b} = \sum_{b=1}^B \sqrt{\alpha_{k,b}(\beta_b)} h_{k,b}^H x_b + n_k \quad (4)$$

From BS b , the signal vector denotes $x_b = \sum_{k \in K_{b,a}} w_{k,b} d_{k,b}$ and obey $E[\|x_b\|^2] \leq P$. we present received signal as follow after substituting (2).

$$y_{k,b} = \sqrt{\alpha_{k,b}(\beta_b)} \hat{h}_{k,b}^H w_{k,b} d_{k,b} + \sqrt{\alpha_{k,b}(\beta_b)} e_{k,b}^H w_{k,b} d_{k,b} + \sum_{j \in K_{b,a}, j \neq k} \sqrt{\alpha_{k,b}(\beta_b)} h_{k,b}^H w_{j,b} d_{j,b} + \sum_{b' \neq b} \sum_{l \in K_{b',a}} \sqrt{\alpha_{k,b'}(\beta_{b'})} h_{k,b'}^H x_{b'} \quad (5)$$

AWGN is denoted as $n_k \sim \text{CN}(0,1)$, $p_{k,b}$ denotes the power and vector $\beta = \{\beta_1, \beta_2, \beta_3 \dots \beta_B\}$ presents tilts. We present the expressions for conditional ergodic rate [9]

$$R_k(\beta) = E \left[\log_2(1 + P_k^{\text{AS}}(\beta) + P_k^{\text{MCI}}(\beta) + P_k^{\text{MCI}}(\beta)) \left| \left\{ \alpha_{k,b(\beta_b)} \right\}_{b=1}^B \right. \right] \\ - E \left[\log_2(1 + P_k^{\text{MCI}}(\beta) + P_k^{\text{ICI}}(\beta)) \left| \left\{ \alpha_{k,b(\beta_b)} \right\}_{b=1}^B \right. \right]. \quad (6)$$

3 Optimal Downlink Tilts for 3D MIMO

We directly investigate the approximate expressions for network throughput with conjugate beamforming (CBF) and zero-forcing beamforming (ZFBF) schemes [5].

3.1 The Approximate Expressions

Considering imperfect CSI, the distribution for approximate ergodic rate are presented. We set the power constraint P for each BS. We allocate the power for transmitting among users equally.

- (1) Desired signal: The ZFBF is first considered. The following formula essentially roughly equals the desired signal term [9].

$$P_k^{\text{AS}}(\beta) = \left\| \sqrt{\alpha_{k,b(\beta_b)}} \hat{h}_{k,b}^H w_{k,b} \right\|^2 P_{k,b} \approx \alpha_{k,b(\beta_b)} P_{k,b} (N_t - K_{b,a} + 1) \mu_{k,b}^2(\beta_b) \quad (7)$$

$K_{b,a}$ denotes the number of users who are affected by BS b . Building on the MMSE channel estimation, the approximate expressions for CBF can be described as follow

$$P_k^{\text{AS}}(\beta) = \left\| \sqrt{\alpha_{k,b(\beta_b)}} \hat{h}_{k,b}^H w_{k,b} \right\|^2 P_{k,b} \approx \alpha_{k,b(\beta_b)} P_{k,b} N_t \mu_{k,b}^2(\beta_b) \quad (8)$$

- (2) Intra-cell interference: Considering ZFBF, we present the approximate expressions as follow

$$P_{k,\text{ZFBF}}^{\text{MCI}}(\beta) = E \left(\sum_{j \in K_{b,a}} \left\| \sqrt{\alpha_{k,b(\beta_b)}} \hat{h}_{k,b}^H w_{k,b} \right\|^2 P_{j,b} \right) \approx \alpha_{k,b(\beta_b)} P_{k,b} K_{b,a} \sigma_{k,b}^2(\beta_b) \quad (9)$$

We consider that CBF cannot mitigate intracell interference efficiently, the expressions for all the interference terms and the error estimated term should be both analyzed. Multiuser residual interference term for ZFBF equals the estimation error expressions for CBF can be described as follow:

$$\mathbb{E} \left(\sum_{\substack{j \in \mathbf{K} \\ j \neq k}} \mathbf{b}_{j,a} \left\| \sqrt{\alpha_{k,b(\beta_b)}} \hat{\mathbf{h}}_{k,b}^H \mathbf{w}_{k,b} \right\|^2 P_{j,b} \right) \approx \alpha_{k,b(\beta_b)} P_{j,b} (\mathbf{K}_{b,a} - 1) \mu_{k,b}^2(\beta_b) \quad (10)$$

For CBF, the multiuser residual interference term can be described as follow

$$P_{k,\text{CBF}}^{\text{MCI}}(\beta) = \alpha_{k,b(\beta_b)} P_{j,b} \mathbf{K}_{b,a} \sigma_{k,b}^2(\beta_b) + \alpha_{k,b(\beta_b)} P_{j,b} (\mathbf{K}_{b,a} - 1) \mu_{k,b}^2(\beta_b) \quad (11)$$

(3) Approximate expressions for inter-cell interference is

$$\begin{aligned} P_k^{\text{ICI}}(\beta) &= \mathbb{E} \left\{ \left[\sum_{b' \neq b} \sum_{l \in \mathbf{K}_{b'a}} \left\| \sqrt{\alpha_{k,b(\beta_b)}} \hat{\mathbf{h}}_{k,b}^H \mathbf{w}_{k,b'} \right\|^2 P_{j,b'} \right] \right\} \\ &= \sum_{b' \neq b} \sum_{l \in \mathbf{K}_{b'a}} \alpha_{k,b'(\beta_{b'})} P_{j,b'} \sigma_{k,b'}^2(\beta_{b'}) \end{aligned} \quad (12)$$

3.2 3D Beamforming Optimal Downtilts Based Om Greedy Algorithm

We adopt modified algorithm to globally improve the statistical throughput performance. Considering the wireless channel, there are some differences should be taken into account, the range of angular is about 120° to cover the area in the horizontal plane. But the cell-inter tilt usually is smaller than 27.5° , and the cell-edge tilt is about 6° in the vertical plane. Considering height ($h_b \gg h_u$), we can neglect multipath fading in the vertical plane.

Building on the mention above, the design of beamforming algorithm has separability. The sum-throughput acts as the main performance metric. Considering imperfect CSI, we should find the optimal downtilt β_{s^*} applied in different BSs via solving this problem as follow.

$$\begin{aligned} \beta_{s^*} &= \arg \max_{\beta} \sum_{b=1}^B R_{b,s}(\beta) \\ \text{Subject to } \beta_{\text{minimum}} &\leq \beta \leq \beta_{\text{maximum}} \end{aligned} \quad (13)$$

For BSs, broadcasting electromagnetic waves from a base station has little effect on the electromagnetic waves produced by other BSs. This effect can be neglected. The transmission of electromagnetic waves between base stations can be assumed to be a relatively independent process, but users in the network are the result of the interaction of multiple base stations. Therefore, this optimal problem is an optimal joint problem of multidimensional variables, which can be simplified for multiple one-dimensional joint optimization problems.

Greedy algorithm means that the overall optimal solution of the problem can be achieved through a series of locally optimal choices. This is the first basic element of the greedy algorithm. Greedy algorithm is to search via iterative method, each time algorithm will be asked to simplify the problem as a smaller sub-problem. It is usually

possible to first prove that a whole optimal solution of the problem begins with a greedy choice and that the problem is decoupled to a smaller, sub-problem after a greedy choice.

Building on this, we design the modified greedy algorithm. This algorithm can enhance performance globally and ensure fast convergence. Because of space limitations, the detailed complexity analysis is omitted here.

We first calculate the initial value of the iteration of each BS. As seen in Fig. 2, downtilts have the following characteristics (1) the cell-inter beam downtilt is bound to be greater than the cell-edge beam downtilt (2) the cell edge beam reaches the optimal downtilt faster. We obtain the initial optimal cell-edge downtilts via iterative search method. For simplification, we assume that the initial value of the inner and outer beams at BSs are same.

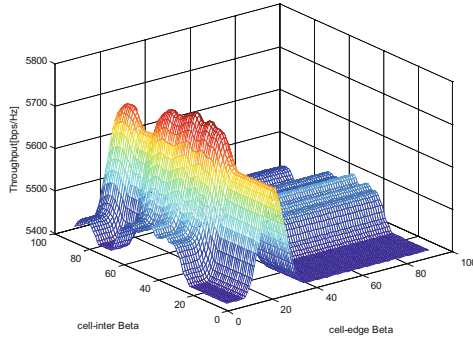


Fig. 2. The throughput under different downtilts

We first set $\beta_{s^*}^* = \theta$ called the initial optimal cell-edge downtilt. We then obtain the optimal tilt of each BS and increment by 1° . If sum throughput increases, the tilts will be kept or ignored. At last, the algorithm stopped when the sum throughput cannot increase anymore. Cooperation system in cellular networks, the intersection regions among the adjacent base stations usually generate the major part of the ICI. Hence, we can significantly combat ICI via using multicell cooperation algorithm [8]. The number of BSs within a cluster can affect the preference of the system and the Superiority of proposed algorithm. The order of complexity and overhead are key parameters for evaluating the algorithm. If increasing the number, the performance can be improved greatly. But the overhead and complexity for CSI cannot be ignored. If reducing the size of the cluster, interference cannot be limited. We weigh the advantages and disadvantages of the size of cluster in this paper. We apply our modified algorithm in this paper via using static clustering method. In this method, we divide our cellular network into small and underlap cluster such that the BSs who generate the dominant interference can be contained in our cooperation cluster [9].

Considering characteristics of interference, we first classify this interference into two kinds and then suppress them: (1) interference caused by the main beam can be suppressed via multiplexing the time slots allocated to serve different vertical regions in

clusters. (2) Considering the small vertical HPBW and the directivity of the antenna patterns in this paper, we suppressed such interference via slightly increasing β cell-edge ($\beta_{\text{cell-inter}} > \beta_{\text{cell-edge}}$).

One new strategy is described as follow: all users only report the ID number of BS which causes the most serious interference to controller central station (CCS). Then CCS ensures that interfering BS cannot simultaneously serve the edge area with other BSs within a cluster.

4 Simulation Results

The proposed schemes and the traditional algorithm are both analyzed. Each cell is divided into three horizontal sectors and two vertical regions, each sector span 120° described in Fig. 3. We present the pathloss factor as $PL_{k,b}$, the reference distance is denoted as D_0 . We present the pathloss exponent as v (Table 1).

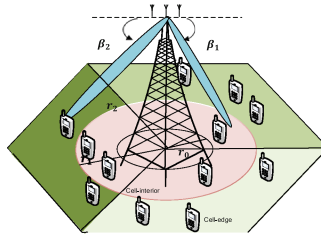


Fig. 3. Two vertical and three horizontal areas

Table 1. Simulation Parameters

The name of Parameters	Assumption
Cellular layout	19 sites
h_b	15 m
h_u	1.5 m
The radius, D	50 m
$PL_{k,b}$	$(\frac{d_{k,b}}{D_0})^{-v}$ $D_0 = 1$ m. $v = 3.76$
The Pathloss breaking distance r_0	1 m
The number of antennas at BS	64
The number of users	60
Power of BS, P	20 w
The antenna pattern of BS	$\theta_{3\text{dB}} = 65^\circ$ $\theta_{3\text{dB},1} = 32$ and $\theta_{3\text{dB},2} = 10$ $SLL_{el} = 50$ dB $SLL_m = SLL_{tot} = 55$ dB

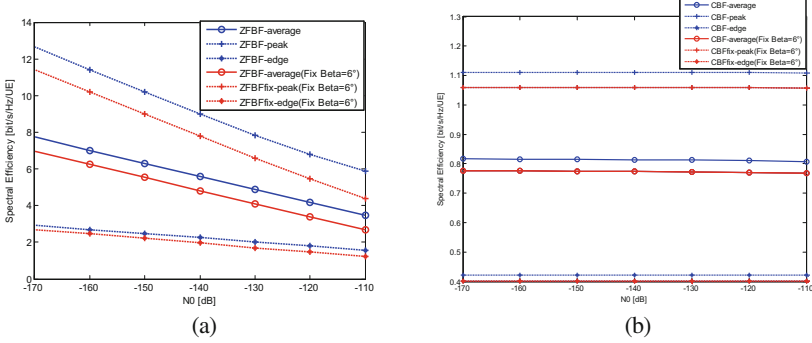


Fig. 4. ZFBF throughput (a) CBF throughput (b)

The proposed and traditional algorithm are both shown in Fig. 4. The edge, average and peak throughput are respectively compared simultaneously. When SNR gets worse, the throughput decreases can be observed clearly. The proposed schemes are chosen to be compared with traditional fixed-downtilt, the improvement of performance can be observed clearly. The proposed schemes compared with fixed tilt applied at BSs beam strategy enhances the edge, average and peak throughputs to 11%, 10% and 10% for ZFBF and 5%, 4% and 5% for CBF. Moreover, the throughput of ZFBF is larger than CBF seen from the simulation because intra-cell interference can be suppressed by ZFBF.

5 Conclusions

In a massive MIMO network, we directly apply approximate ergodic rate expressions with imperfect CSI. For the optimal downtilts, an adaptive switching beam strategy is proposed based on the greedy algorithm to further enhance performance. We investigate interference and present a possible idea to reduce the interference. The superiority of our modified algorithm over the traditional ways is proved by numerical results.

Acknowledgement. This work thanks to the project 61471066 supported by NSFC.

References

1. Larsson, E., Edfors, O., Tufvesson, F., Marzetta, T.: Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.* **52**(2), 186–195 (2013)
2. Chen, J.: When does asymptotic orthogonality exist for very large arrays? In: *IEEE GLOBECOM* (2013)
3. Yang, H., Marzetta, L.: Performance of conjugate and zeroforcing beamforming in large-scale antenna systems. *IEEE J. Sel. Areas Commun.* **31**(2), 172–179 (2013)

4. El-khamy, E., Moussa, K., El-sherif, A.: Performance analysis of massive MIMO multiuser transmit beamforming techniques over generalized spatial channel model. In: 32nd NRSC (2015)
5. Seifi, N., Coldrey, M., Svensson, T.: Throughput optimization in MU-MIMO systems via exploiting BS antenna tilt. In: Proceedings of IEEE Global Communications (GLOBECOM) Workshops, Anaheim, CA (2012)
6. He, S., Huang, Y., Jin, S., Yang, L.: Coordinated beamforming for energy efficient transmission in multicell multiuser systems. *IEEE Trans. Commun.* **61**(12), 4961–4971 (2013)
7. Seifi, N., Heath, R., Coldrey, M., Svensson, T.: Adaptive multicell 3-D beamforming in multi-antenna cellular networks. *IEEE Trans. Veh. Technol.* **65**(8), 6217–6231 (2016)
8. Baracca, P., Boccardi, F., Braun, V., Tulino, A.: Base station selection and per-cell codebook optimization for CoMP with joint processing. In: Proceedings of IEEE International Symposium PIMRC, pp. 2329–2334, Australia (2012)
9. Zheng, Q., Liu, R., Cheng, X., Sun, S., Chen, X.: Optimal downtilts for 3D beamforming in massive MIMO networks with imperfect CSI. In: 16th International Symposium on Communications and Information Technologies (ISCIT) (2016)

Dynamic User Scheduling Algorithms for Massive MIMO Multicast System

Xinran Zhang^{1,2,3}(✉) and Songlin Sun^{1,2,3}

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
zhangxr_wspn@bupt.edu.cn

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ National Engineering Laboratory for Mobile Network Security,
Beijing University of Posts and Telecommunications, Beijing, China

Abstract. In this work we consider the user scheduling problem in the massive multiple-input multiple-output (MIMO) wireless multicast system with heterogeneous structures. The dynamic programming (DP) method and Markov decision process (MDP) model is utilized to describe the system behavior. We use asymptotic results for massive MIMO multicast beamforming to estimate the system capacity for the MDP model. The value iteration (VI) method is adopted to solve the MDP problems. The proposed model can enhance the system performance by solving the optimal MDP policy for user scheduling in an off-line manner with maximized average reward. The numerical results show the behavior of the algorithm and evaluate its performance.

Keywords: Massive MIMO · Wireless multicast
Markov decision process

1 Introduction

The massive multiple-input multiple-output (MIMO) technology has been widely acknowledged to be the key technology in future wireless communication system since it can significantly increase spectrum efficiency and improve system performance. Among the various research fields of massive MIMO, the wireless multicast based on massive MIMO is drawing more and more attention. Recent literatures have shown that under massive MIMO scenario, the asymptotic results of optimal multicast beamformer as well as the achieved optimal SINR have compact forms [1, 2], providing theoretical foundation for further analysis for massive MIMO multicast.

Due to the stochastic behavior of wireless system, in practical scenario the scheduling method is necessary to guarantee and improve system performance

and user's quality of service (QoS). To address the scheduling problem in massive MIMO multicast system, the dynamic programming (DP) and Markov decision process (MDP) models are widely applied for theoretical modeling. In the field of wireless scheduling as well as optimal controlling, recent literatures have been focusing on structural analysis of optimal DP policies [3,5,6], approximate dynamic programming (ADP) [4,7] and aggregation based methods [8], etc. With the rapid development and global focusing of learning-based theories and techniques, it is important to apply these theories to address problems in the field of wireless communication to develop practical algorithms as well as to further explore these theories by modeling in practical complicated wireless communication scenarios.

To address the scheduling problems in massive MIMO multicast system, we use DP and MDP to model the system behavior. First we adopt the asymptotic results from [1,2] and derive the MDP model for the system. By using value iteration (VI) methods, the optimal MDP policy for user scheduling is derived. In numerical result part we test the performances of the algorithm.

The rest of this paper is organized as follows. In Sect. 2 we develop the system model for user scheduling in massive MIMO multicast system. Section 3 provides theoretical analysis for the DP model and derives algorithms to solve the optimal policy. Section 4 presents the numerical results of the proposed algorithms. Section 5 concludes the paper.

2 MDP Modeling in Massive MIMO Multicast HetNet

Consider a heterogeneous network (HetNet) where the base station is coupled with N LPNs, which all locate inside the coverage area of the base station. There are M multicast groups in the HetNet that utilizes multicast beamforming for wireless transmission. We set the purpose of MDP modeling to be providing scheduling strategy when user equipments (UEs) roam across the LPNs inside the base station. Consider the infinite-horizon discrete-state MDP [9]. Without loss of generality, define the state vector of the MDP as follows (Fig. 1):

$$\mathbf{s}(t) = [\mathbf{s}_0(t), \mathbf{s}_1(t), \dots, \mathbf{s}_N(t)], \quad (1)$$

$$\mathbf{s}_i(t) = [s_{i,1}(t), s_{i,2}, \dots, s_{i,M}(t)], \quad i \in \{0, 1, \dots, N\}, \quad (2)$$

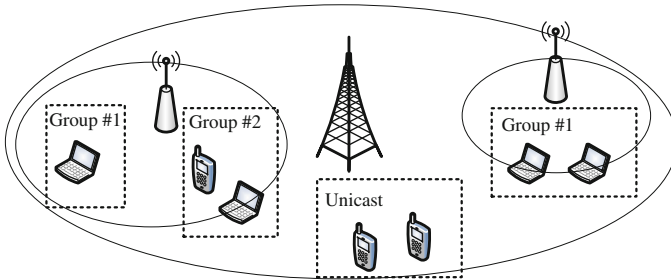


Fig. 1. System model

where $s_{i,m}(t), m \in \{1, 2, \dots, M\}$ denotes the number of connected-state UEs in multicast group m of access point i ($i = 0$ denotes the marco base station). Define the action vector of the MDP as follows:

$$\mathbf{u}(t) = [\mathbf{u}_0(t), \mathbf{u}_1(t), \dots, \mathbf{u}_N(t)], \quad (3)$$

$$\mathbf{u}_i(t) = [u_{i,1}^{AI}(t), u_{i,1}^{AC}(t), u_{i,2}^{AI}(t), u_{i,2}^{AC}(t), \dots, u_{i,M}^{AI}(t), u_{i,M}^{AC}(t)], \quad i \in \{0, 1, \dots, N\}, \quad (4)$$

where $u_{i,m}^{(*)}(t) \in \{0, 1\}, m \in \{1, 2, \dots, M\}$ denote the scheduling policy for the next decision epoch when different type of UEs arrive at access point i and try to subscribe to multicast group m , which is described in Table 1. Let the instances of UE's arrival be the decision epochs. Hence the state vector and action vector can be re-written with the decision epoch index as \mathbf{s}_t and $\mathbf{u}_t, t = 0, 1, \dots$. In order to build the connection between the performance of massive MIMO multicast system and the UE's density, recall the Rayleigh channel model based asymptotic results as follows. Consider down-link multicast transmission, assume the UE is equipped with single antenna, then the received multicast signal can be expressed by the following equation:

$$y_{i,m,k} = \sqrt{\rho_{i,m}} \mathbf{g}_{i,k}^H \omega_{i,m} s_{i,m} + \sum_{n, n \neq m} \sqrt{\rho_{i,n}} \mathbf{g}_{i,k}^H \omega_{i,n} s_{i,n} + z_{i,m,k}, \quad (5)$$

Table 1. Action variable for different types of UE

	0	1	UE's type
$u_{i,m}^{AI}(t)$	Reject	Accept	New Idle-state UE
$u_{i,m}^{AC}(t)$	Reject	Accept	New Connected-state UE

where $s_{i,m} \in \mathbb{C}$ denotes the multicasted complex information with unit power, $\omega_{i,m} \in \mathbb{C}^{N_{t,i} \times 1}$ denotes the multicast beamforming vector with $\|\omega_{i,m}\|^2 = 1$, and $N_{t,i}$ denotes the number of transmission antenna. $\rho_{i,m}$ denotes the power allocation ratio with $\sum_m \rho_{i,m} = 1, \forall i$. $y_{i,m,k}$ denotes the received signal of UE k in multicast group m of access node i . With the channel denoted by $\mathbf{g}_{i,k}$ being Rayleigh fading, the achieved asymptotic optimal SINR under the max-min fairness (MMF) rule is given as follows [1, 2]:

$$\lim_{N_{t,i} \rightarrow \infty} \text{SINR}_{i,m} = \frac{E}{\sigma^2 \sum_k \beta_{i,k}^{-1}}, \quad \forall m, \quad (6)$$

where the beamformer ω and power allocation ratio ρ satisfies:

$$\omega_{i,m} = \alpha_{i,m} \sum_k \frac{\mathbf{g}_{i,k}^H}{\beta_{i,k}}, \quad \forall i, m, \quad (7)$$

$$\alpha_{i,m} = \left(N_{t,i} \sum_k \beta_{i,k}^{-1} \right)^{-\frac{1}{2}}, \quad (8)$$

$$\rho_{i,m} = \frac{\sum_k \beta_{i,k}^{-1}}{\sum_m \sum_k \beta_{i,k}^{-1}}, \quad \forall m. \quad (9)$$

$\beta_{i,k}$ is the large-scale channel attenuation and E is the total transmission power. In this way the Connected-UE's distribution will be mapped directly to the performance of the system. Define $\xi_{i,m}^C(r, \theta)$ and $\xi_{i,m}^I(r, \theta)$ as the density of connected-state UEs and idle-state UEs under polar coordinate. In that case the achieved SINR can be calculated as follows [11]:

$$\text{SINR}_{i,m} = \frac{E}{\sigma^2 \iint_{D_i} K^{-1} r^n \xi_{i,m}^C(r, \theta) dS}, \quad \forall i, m, \quad (10)$$

where K is the constant determined by antenna characteristics, and D_i denotes the coverage area of LPN i . Hence the viable state space can be determined according to the QoS constraints:

$$\mathcal{S} = \left\{ \mathbf{s}_t \mid \mathbf{s}_{i,t} \in \mathbb{Z}_+^{1 \times M}, \text{SINR}_{i,m} \geq Q_m^{\min}, \forall i, m, t. \right\}, \quad (11)$$

where Q_m^{\min} is the minimum SINR required for the multicast signal in group m . Let \mathcal{U} be the action space which encompasses all possible \mathbf{u}_t . Since the state \mathbf{s}_t indicates the number of subscribed UEs, it is obvious that any two states in \mathcal{S} are mutually accessible, i.e., the Weak Accessibility (WA) condition holds. In that case the optimal policy of the MDP problem is stationary [3, 9]. Let $\pi = \{\mu, \mu, \mu, \dots\}$ be the stationary policy where $\mu : \mathcal{S} \rightarrow \mathcal{U}$ is the function of choosing an action for all decision epochs. Let Π be the set of all possible stationary policies. The cost per state is defined by $r : \mathcal{S} \times \mathcal{U} \rightarrow \mathfrak{R}$. In that case the goal of MDP is to maximize the average reward per stage Problem [9]:

$$J^* = \max_{\pi \in \Pi} J_\mu = \max_{\pi \in \Pi} \left\{ \lim_{T \rightarrow \infty} E \left\{ \frac{1}{T} \sum_{t=0}^{T-1} r(\mathbf{s}_t, \mu(\mathbf{s}_t)) \right\} \right\}, \quad (12)$$

where J_μ is the reward function under stationary policy $\{\mu, \mu, \mu, \dots\}$.

3 Dynamic Programming and Optimal Solution

At each decision epoch, the system state transition depends on UE's mobility as well as the action chosen by policy π . In practical scenario the mobility parameters are unknown, which results in difficulty of implementing the MDP-based algorithms. Nevertheless, once the parameters are properly estimated, substantial analytical methods can be introduced to solve the DP problem. For the simplicity of denotation, we use scalar $i, j = 1, 2, \dots, n$ to index each state in \mathcal{S} . The cost is hence denoted by $r(i, u)$. Let $p_{ij}(u)$ be the transition probability from state i to j with action u . The transition matrix is hence denoted by P_u .

Since the Markov chain is a unichain, the optimal average reward J^* satisfies the following equations [9]:

$$J^* + h(i) = r(i, u) + \sum_{j=1}^n p_{ij}(u)h(j), i = 1, 2, \dots, N, \quad (13)$$

where h is a vector with $h(0) = 0$, and u that attains J^* is the optimal policy. Equation (13) is referred to as the Bellman's Equation for unichain policies. To solve the Bellman's equation, define the DP mapping for vector h :

$$\mathcal{T}h = \max_u [r_u + P_u h] \quad (14)$$

Then the optimal reward function can be obtained by the following value iteration (VI) method:

$$J^* = \lim_{k \rightarrow \infty} \frac{1}{k} \mathcal{T}^k h. \quad (15)$$

The optimal policy can be obtained by the well-known policy iteration method, and is hence omitted here.

4 Numerical Results

We experiment the models above using the MDP Toolbox on MATLAB [12]. The experiment is based on Markovian transition hypothesis of the system, and suffers from the curse of dimensionality. For a system with N LPNs and M groups, the set size of compound state-action pair is $|\mathcal{S}| \times |\mathcal{U}|$, and grows exponentially with N and M . As a result, we only set $N = 1$ and $M = 2$, and set the minimum SINR requirement to be sufficiently high, so that the maximum UE's number for each LPN and group is not too large. Let the reward function to denote the penalty of denying UE's access. Define $r(i, u)$ as follow:

$$r(i, u) = \sum_i \sum_m \left\{ R_{AI} \times u_{i,m}^{AI} + R_{AR} \times u_{i,m}^{AR} \right\}. \quad (16)$$

Let R_{AI} be the fixed reference reward set to be 1, and let R_{AR} vary. We evaluate the average reward of the MDP problem under different Idle-state UE density indicated by ξ^I (We assume that $\xi_{i,m}^I$ is unified in the system as the reference parameter in the simulation system. For detailed description of the mobility model, please see [11]). Since with a higher value of R_{AR} , it is obvious to achieve a higher average reward. Hence we also calculate the average reward ratio J^*/R_{AR} as the criterion for performance. The results are shown in the following figures.

As is shown in Fig. 2, the average reward increases for larger value of R_{AR} as expected. Also, with a higher Idle-state UE density, the system traffic becomes heavier and more handover requests are introduced, deteriorating the performance and decreasing average reward, which proves the rationality of the proposed method. On the other hand, in Fig. 3 we observe that with a higher R_{AR}

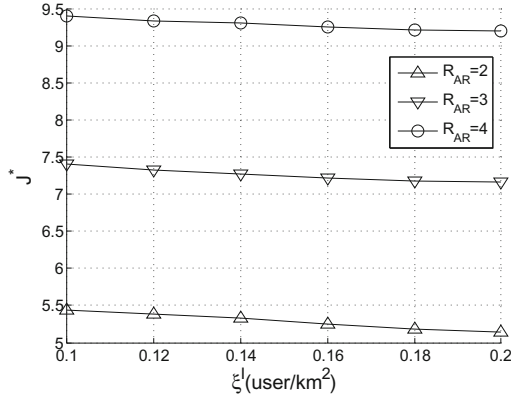


Fig. 2. Average reward with different R_{AR}

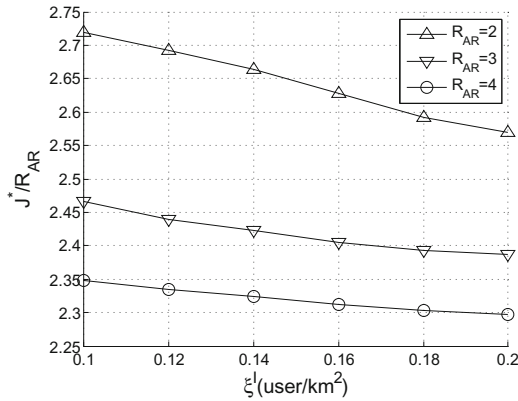


Fig. 3. Average reward ratio with different R_{AR}

the average reward ratio becomes smaller, which indicates that a high value of R_{AR} do not guarantee a high efficiency for the scheduling problem. In practical system, the reward may be the achieved rates of a data service. In that case, for a service with a higher data rate, although the average throughput (average reward) is increased, the unit throughput by introducing this service becomes lower indeed, which suggests lower efficiency of the multicast system.

As is shown in Fig. 4, when we set some specific QoS requirements so that the maximum number of UEs are the same for each LPN and multicast group, similar deterioration tendency can also be observed. Also, with a higher maximum number of users denoted by $SMax$, the system bears more multicast traffic and brings more average reward. Similar to the average reward, we also calculate $J^*/SMax$ in this scenario to evaluate the efficiency. The result is shown in Fig. 5. Similar to the observation of J^*/R_{AR} , a higher maximum user number actually brings lower efficiency, which coincides with previous analysis.

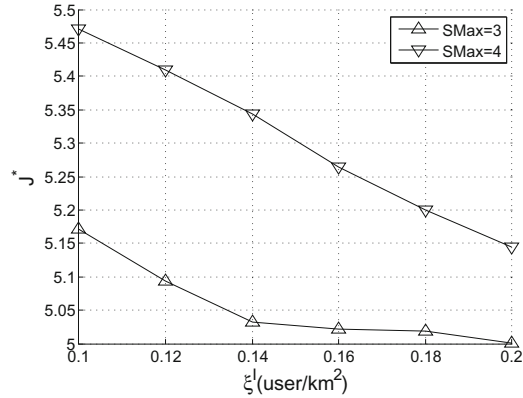


Fig. 4. Average reward with different $SMax$

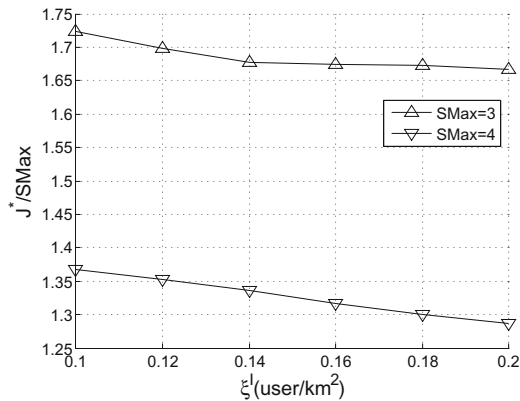


Fig. 5. Average reward ratio with different $SMax$

5 Conclusions

This paper proposes the user scheduling algorithms for massive MIMO multicast system. We use asymptotic results of massive MIMO multicast to build the connection between user density and the achieved SINR. Then the MDP theory is adopted to model system behavior and derive the VI-based method for solving the optimal policy. Numerical experiments testify the rationality of the proposed model, where system behavior for different Idle-state UE density is observed and analyzed. For future work, the deeper researches into the phenomenon observed in the average reward ratio as well as low-complexity algorithm design are of interest.

Acknowledgement. This work is funded by project 61471066 supported by National Natural Science Foundation of China (NSFC).

References

1. Xiang, Z., Tao, M., Wang, X.: Massive MIMO multicasting in noncooperative multicell networks. In: IEEE International Conference on Communications (ICC), Sydney, pp. 4777–4782, June 2014
2. Zhou, H., Tao, M.: Joint multicast beamforming and user grouping in massive MIMO Systems. In: 2015 IEEE International Conference on Communications (ICC), London, pp. 1770–1775 (2015)
3. Zhou, B., Cui, Y., Tao, M.: Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks. *IEEE Trans. Wirel. Commun.* **15**(9), 6284–6297 (2016)
4. Zhang, X., Jin, H., Ji, X., Li, Y., Peng, M.: A separate-SMDP approximation technique for RRM in heterogeneous wireless networks. In: Proceedings of IEEE Wireless Communications Networking Conference WCNC (WCNC 2012), pp. 2087–2091, April 2012
5. Djonin, D.V., Krishnamurthy, V.: Q -learning algorithms for constrained markov decision processes with randomized monotone policies: application to MIMO transmission control. *IEEE Trans. Sign. Process.* **55**(5), 2170–2181 (2007)
6. Sun, S., Dong, M., Liang, B.: On stochastic feedback control for multi-antenna beamforming: formulation and low-complexity algorithms. *IEEE Trans. Wirel. Commun.* **13**(9), 4731–4745 (2014)
7. Chandrashekar, L., Bhatnagar, S.: Approximate dynamic programming with (min; +) linear function approximation for markov decision processes. In: Proceedings of 53rd IEEE Conference on Decision and Control, Los Angeles, CA, pp. 1588–1593 (2014)
8. Jia, Q.: On state aggregation to approximate complex value functions in large-scale markov decision processes. *IEEE Trans. Autom. Contr.* **56**(2), 333–344 (2011)
9. Bertsekas, D.P.: *Dynamic Programming and Optimal Control*, vol. 2, 3rd edn. Athena Scientific, Belmont (2011)
10. Zheng, K., Ou, S., Yin, X.: Massive MIMO channel models: a survey. *Int. J. Antennas Propag.* **2014** (2014). Article ID 848071
11. Zhang, X., Sun, S.: Dynamic scheduling for wireless multicast in massive MIMO HetNet. *J. Phys. Commun.* (under reviewing process)
12. Chades, I., Chapron, G., Cros, M.J., Garcia, F., Sabbadin, R.: MDPtoolbox: a multi-platform toolbox to solve stochastic dynamic programming problems. *Ecography* **37**, 916–920 (2014). <http://www7.inra.fr/mia/T/MDPtoolbox/>

A Bat Algorithm Based on Centroid Strategy

Siqing You^{1,2}, Tongjuan Liu², Fei Xue^{1,2(✉)}, Hongjie Liu¹,
and Zhaoqun Qi¹

¹ Beijing Advanced Innovation Center for Future Internet Technology,
Beijing 100124, China

xuefei2004@126.com

² School of Information, Beijing Wuzi University, Beijing 101149, China

Abstract. The standard bat algorithm is slow convergence and low precision. To overcome this shortcoming, we propose a novel variant of bat optimization algorithm based on centroid strategy. The proposed algorithm has a better global searching capability because the centroid strategy can effectively prevent it falling into a local optimum. Two typical test functions are employed to test its performance. Simulation results show our proposal is both effective and efficient than three other comparison algorithms. Moreover, for high-dimensional function optimization, our proposed algorithm also has excellent approximation performance.

Keywords: Bat algorithm · Centroid strategy · Function optimization

1 Introduction

Many heuristic optimization algorithms [1] have been proposed to solve complex computational problems (e.g., nonlinear problems, high-dimensional optimization). Currently, these swarm intelligence algorithms are also widely applied in many fields [2–4] (e.g., engineering optimization, scientific computing, biological information) due to the flexibility and superior ability to deal with different problems. Moreover, the no free lunch theorem (NFL) conveys that there is no heuristic algorithm can cope with all optimization problems [5]. That is to say, the existing algorithms have achieved satisfactory results in solving some problems, but not all of them. Therefore, an increasing number of heuristic intelligent algorithms have been proposed by researchers.

Bat algorithm (BA, Yang in 2010) [6] is a novel swarm intelligence algorithm. This algorithm is inspired by the bats' echolocation behavior that allows bats to capture prey and avoid obstacles quickly. The potential parallelism and robustness of bat algorithm attracted many researchers. Xie et al. employed differential operators and Levy flights to enhance the searching ability of bat algorithm [7]. Cai et al. introduced Gaussian walk into BA to solve high-dimensional numerical problems [8]. Furthermore, bat algorithm is widely applied in many applications, for example, engineering optimization [9], multi-objective optimisation [10], constrained optimization tasks [11], etc.

Although the basic bat algorithm has powerful search performance, it still has some deficiencies such as no global convergence, low accuracy. Therefore, to address this question, a novel bat algorithm based on centroid strategy is proposed to make bat

algorithm more effective. The proposed algorithm has a better global searching capability because the centroid strategy can effectively prevent it falling into a local optimum. Two typical test functions are employed to test its performance. Simulation results show our proposal is both effective and efficient than three other comparison algorithms. Moreover, for high-dimensional function optimization, our proposed algorithm also has excellent approximation performance.

The remaining part of the paper is organized as follows: Sect. 2 outlines the formulation of bat algorithm and centroid strategy. Section 3 gives the design framework of bat optimization algorithm based on centroid strategy, named CBA. Section 4 illustrates the experimental evaluations. Section 5 concludes this paper.

2 Bat Algorithm and Centroid Strategy

2.1 Bat Algorithm

The echolocation behavior of microbats allows bats to capture prey and avoid obstacles quickly. For bat algorithm, Yang et al. idealized three rules to formulate this behavior as follows [6]:

- (1) All bats can detect the distance from the target in search space;
- (2) Each bat adjust its velocity and position according to the loudness and frequency of ultrasonic, which varies with the distance from the target;
- (3) The loudness decreases with the increase of the iteration (from A_{\max} to A_{\min}).

In simulations, firstly, the position and velocity of bat i are $\vec{x}_i(t)$ and $\vec{v}_i(t)$ at time t in the search space, respectively, as well as the loudness $A_i(t)$, frequency f_i and rate of emission pulse $r_i(t)$. Then the bats adjust its velocity and position according to the loudness and frequency of ultrasonic, which varies with the distance from the target; For bat i at time $t+1$, its new velocity $\vec{v}_i(t+1)$ and position $\vec{x}_i(t+1)$ is given by

$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t+1) \quad (1)$$

$$\vec{v}_i(t+1) = \vec{v}_i(t) + (\vec{x}_i(t) - \vec{p}_i(t)) * f_i \quad (2)$$

$$f_i = f_{\min} + (f_{\max} - f_{\min}) * \beta \quad (3)$$

where $\vec{P}(t)$ is the global best position at time t , $\beta \in [0, 1]$ is a random number which follows the normal distribution, the f_{\min} and f_{\max} is the minimum and maximum of frequency, respectively.

For the local search, bat algorithm employs following strategies that a new position is generated as follow:

$$\vec{x}_i(t) = \vec{p}_i(t) + \varepsilon A_{mean}(t) \quad (4)$$

where $\varepsilon \in [-1, 1]$ is a random number, the mean $A_{mean}(t)$ is the average value of all loudness at time t .

Furthermore, the biological characteristics of bats (the A decreases and r increases as they approach their target), are formulated as follows:

$$A_i(t+1) = \alpha A_i(t) \quad (5)$$

$$r_i(t+1) = r_i(0)[1 - \exp(-\gamma t)] \quad (6)$$

This is inspired by the characteristic of bat predation. $r_i(0)$ is the initial value of pulse rate, γ and α are constant.

2.2 Centroid Strategy

Centroid is a hypothetical point in the material system that is believed to be concentrated in mass. Unlike the center of gravity, the centroid does not have to be in a system with a gravitational field. It is worth noting that unless the gravitational field is uniform, the centroid and the center of gravity of the same material system are usually not at the same hypothesis. This paper will focus on an object in the object position uncertainty region is called the centroid of the distribution of instances, the reaction can be instances of the centroid which the concentration distributes in an uncertain region.

The centroid of k points x_1, x_2, \dots, x_k in a D -dimensional space can be given by

$$C = \frac{\sum_{i=1}^K w_i x_i}{\sum_{i=1}^K w_i} \quad (7)$$

Where w_i is the weight of point i , in this paper, it is the fitness of bat i . This point is the arithmetic mean of all points, which is the uniform position. This strategy enables all bats to share information better.

3 A Bat Algorithm Based on Centroid Strategy

In this paper, the basic bat algorithm based on the introduction of information exchange mechanism with the historical experience of individual bats bat populations controls the bat's flight speed and guides bat flight positioning, and so the bat sonic frequency linear adaptive increases, which reinforces a certain extent. The ability can explore algorithms to escape from local optima, and be more in line with nature. This paper introduces the centroid method of decision-making group, which predicts the optimal position to enhancement algorithm optimization.

For the current position of the individual bats can be calculated based on the principle of a centroid which is the potential for the most advantages. Specific steps are as follows:

a, more compatible value than the current value of the individual's choice bat, and calculate the number of individuals m ;

b, If $m > 1$, the centroid is calculated according to the formula $x_m^{t+1} = \frac{1}{m} \sum_{i=1}^m x_i^t$,

otherwise go to step a;

c, the current instance is pulled bat centroid position;

d, processed to determine whether all individuals, if not then transferred to a; Centroid calculation formulas and illustrated as follows:

$$x_m^{t+1} = \frac{1}{m} \sum_{i=1}^m x_i^t \tag{8}$$

The improved algorithm steps bat as follows:

Step 1 Objective function and basic parameters initialization.

Step 2 Initialization all bats position x_i^0 and velocity v_i^0 , according to the equation $f_i = f_{\min} + (f_{\max} - f_{\min}) * \beta$ generating each bat pulse frequency v_i^0 and the initialize pulse r_i^0 calculating loudness adaptation emissivity values.

Step 3 new parameters are added to the speed of the weight parameter w , according to the formula $w = 0.9 - 0.8 \frac{i-1}{\max_length-1}$, w linearly decreases according to the formula $v_i^{t+1} = wv_i^t + (x_i^t - x_*)f_i + (x^* - x_i^t)$ and $x_i^t = x_i^{t-1} + v_i^t$, update the velocity v_i^t and position of bats x_i^t .

Step 4 For each bat, generates a random number rand1 , if $\text{rand1} > r_i^t$, according to the equation $x_i^t = x_* + \varepsilon \overline{A^t}$, and regenerate the optimal disturbance instance in the vicinity of the current position.

Step 5 generates a random number rand2 , If $\text{rand2} < A_i^t$ and fitness is better than the optimal value, then accept the new solution, and according to the formula $r_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)]$ and $A_i^{t+1} = \alpha A_i^t$, adjusting r_i^0 and A_i^0 .

Step 6 updates the global optimum position.

Step 7 determines whether the end condition is met, and outputs a result of the algorithm is terminated, otherwise, go to Step 3.

4 The Simulation and Analysis

4.1 Benchmarks

In this paper, to test the performance, two typical test functions are employed. Their formula are given as follows:

(1) Sphere Model function: $f_1(x) = \sum_{j=1}^{30} x_j^2$

where $|x_j| \leq 100.0$, and $f_1(x^*) = f_1(0, 0, \dots, 0) = 0.0$

(2) Rosenbrock function:

$$f_3(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2],$$

Where $-30 \leq x_i \leq 30$, and $\min(f_3) = f_3(1, \dots, 1) = 0.0$.

4.2 Parameter Settings

To verify the performance of CBA, we compare it with three other algorithms: particle swarm optimization (PSO), cuckoo search algorithm (CS) and the variant of PSO with time-varying acceleration coefficients (TVAC) [12].

For PSO and TVAV, the parameter setting are the same with [12]. For CS, the parameter setting are the same with [13], the parameter α is 1.0, and the parameter p is 0.25. For BA, The coefficients β and ε are both set to a random number uniformly distributed -1 to 1 , as well as in CBA. For BA and CBA, the parameter α setting is 0.95, and, the parameter γ setting is 0.9. The population size of all algorithms are all 100. The dimensionality is 30 and 300 for Sphere Model and Rosenbrock. Each algorithm run independently 30 times for the test suite. Each algorithm stops when the number of function evaluations reaches the maximum $50 \times \text{dimension}$.

4.3 Single-Modal Functions

Comparison results at Sphere Model function with several other algorithms are presented in Table 1. In this table, the Dim is the dimension, Mean represents the average of the 30 results and STD means standard variance. Best is the optimal solution, and Worst is the worst solution. From Table 1, we can see that our improved algorithm has achieved best results dimension 30 and 300. Especially for the dimension 300, CBA performs better.

Table 1. Comparison results on Spher Model function

Dim	Algorithm	Mean	STD	Best	Worst
30	PSO	3.2681e - 10	4.3787e - 10	1.6485e - 11	1.7625e - 09
	TVAC	4.0616e - 29	2.0212e - 28	1.5868e - 36	1.1055e - 27
	CS	8.8770e - 04	3.3410e - 04	4.1654e - 04	1.7673e - 03
	BA	2.5775e + 03	1.1415e + 03	5.2983e + 02	6.4569e + 03
	CBA	1.0615e - 05	1.8552e - 05	3.7951e - 09	7.1932e - 05
300	PSO	3.3946e + 00	1.9783e + 00	9.6284e - 01	8.5833e + 00
	TVAC	4.4760e - 01	1.5790e + 00	2.0208e - 03	8.7546e + 00
	CS	1.4655e - 05	4.6657e - 06	6.7384e - 06	2.6744e - 05
	BA	2.2559e + 05	2.2651e + 04	1.7827e + 05	2.7778e + 05
	CBA	5.9077e - 06	7.0108e - 06	1.9500e - 115	2.7263e - 04

To express more intuitively, the dynamic convergence of comparison is presented for dimension 30 and 300 in Figs. 1 and 2. As we can see, CBA achieves faster convergence rate on the Sphere Model function.

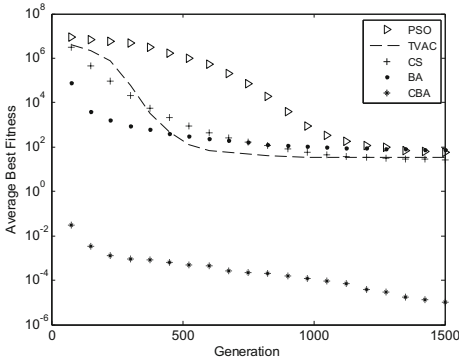


Fig. 1. Comparison of Spher Model function on 30 dimensions

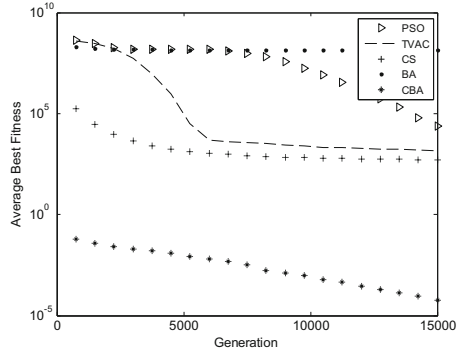


Fig. 2. Comparison of Spher Model function on 300 dimensions

This benchmark is regarded as a classic functions to optimize. For this model, the performance of CBA always superior than others at dimension 300. In addition, TVAC performs well for low dimension (30) among PSO, CS, BA and CBA.

4.4 Multi-modal Functions

Rosenbrocks is a multi-modal function, and this benchmark is one of the most complicated functions to optimize. This is due to that its local minimum value increases exponentially with the increase of dimension. As we can see from Table 2 and the dynamic convergence of comparison Figs. 3 and 4, our proposal CBA always best, whether in low dimension or high dimension.

Table 2. Comparison results on Rosenbrock Model function

Dim.	Algorithm	Mean	STD	Best	Worst
30	PSO	5.6120e + 001	4.3675e + 001	1.2918e + 001	1.9647e + 002
	TVAC	3.3219e + 001	4.1450e + 001	3.2026e + 000	1.7872e + 002
	CS	2.7398e + 001	1.1070e + 000	2.4852e + 001	2.9285e + 001
	BA	7.4706e + 001	1.2442e + 002	2.2049e + 001	5.9988e + 002
	CBA	2.7426e + 001	1.0128e + 000	1.0128e + 000	2.8668e + 001
300	PSO	2.3308e + 004	1.9727e + 004	9.9654e + 003	7.3200e + 004
	TVAC	1.4421e + 003	3.4372e + 002	1.2302e + 003	2.3770e + 003
	CS	5.0309e + 002	8.7603e + 001	4.1376e + 002	7.3538e + 002
	BA	1.3205e + 008	3.0832e + 007	8.7909e + 007	1.9654e + 008
	CBA	2.9891e + 002	7.0818e - 002	2.9878e + 002	2.9899e + 002

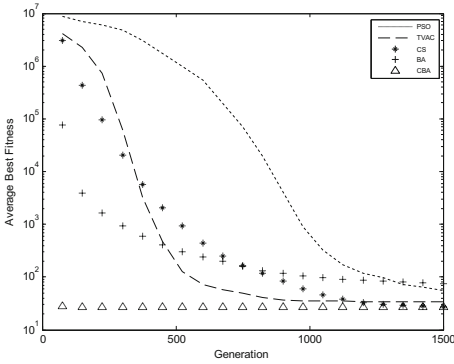


Fig. 3. Comparison of Rosenbrock Model function on 30 dimensions

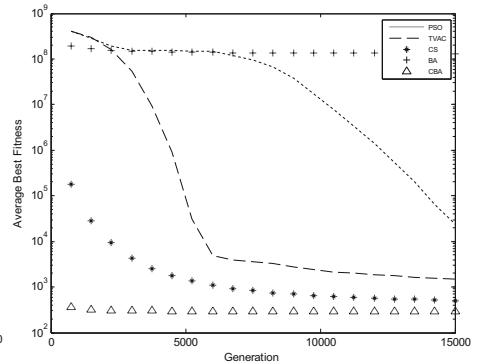


Fig. 4. Comparison of Rosenbrock Model function on 300 dimensions

5 Conclusions

This paper proposed a novel variant of bat algorithm based on centroid strategy. The proposed algorithm has a better global searching capability, which can effectively get out from a local optimum. Simulation results on two typical benchmark functions show our proposal is more effective when compared with PSO, TVAC, CS and BA. Moreover, for high-dimensional function optimization, our proposed algorithm also has excellent approximation performance.

References

1. Kennedy, J.: Swarm intelligence. In: Handbook of Nature-Inspired and Innovative Computing, pp. 187–219. Springer (2006)
2. Shi, Y., et al.: Particle swarm optimization: developments, applications and resources, pp. 81–86. IEEE (2001)
3. Karaboga, D., Akay, B.: A survey: algorithms simulating bee swarm intelligence. *Artif. Intell. Rev.* **31**(1–4), 61–85 (2009)
4. Cui, Z., Sun, B., Wang, G., et al.: A novel oriented cuckoo search algorithm to improve DV-Hop performance for cyber-physical systems. *J. Parallel Distrib. Comput.* **103**, 42–52 (2017)
5. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**(1), 67–82 (1997)
6. Yang, X.: A new metaheuristic bat-inspired algorithm. In: Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), pp. 65–74 (2010)
7. Xie, J., Zhou, Y., Chen, H.: A novel bat algorithm based on differential operator and Lévy flights trajectory. *Comput. Intell. Neurosci.* **2013** (2013)
8. Cai, X., Wang, L., Kang, Q., et al.: Bat algorithm with Gaussian walk. *Int. J. Bio-Inspired Comput.* **6**(3), 166–174 (2014)
9. Yang, X., Hossein, G.A.: Bat algorithm: a novel approach for global engineering optimization. *Eng. Comput.* **29**(5), 464–483 (2012)

10. Yang, X.: Bat algorithm for multi-objective optimisation. *Int. J. Bio-Inspired Comput.* **3**(5), 267–274 (2011)
11. Gandomi, A.H., Yang, X., Alavi, A.H., et al.: Bat algorithm for constrained optimization tasks. *Neural Comput. Appl.* **22**(6), 1239–1255 (2013)
12. Cai, X., Cui, Y., Tan, Y.: Predicted modified PSO with time-varying accelerator coefficients. *Int. J. Bio-Inspired Comput.* **1**(1–2), 50–60 (2009)
13. Yang, X., Deb, S.: Cuckoo search via Levy flights, pp. 210–214. *IEEE* (2009)

Algorithms Optimization and Implementation

An Optimized Daisy-chain Topology for Multi-load Interconnection in High-speed and High-density Electronic Systems

Fengrui Guo¹, Xingming Li^{2(✉)}, Shanqing Hu¹, and Yanyan Qin¹

¹ Beijing Key Laboratory of Embedded Real-Time Information Processing Technology, Beijing Institute of Technology, Beijing 100081, China

² Tsinghua University, Beijing 100081, China
lxm11544@163.com

Abstract. With the development of the electronic system towards high-speed and high-density, the influence of the interconnection topology becomes more and more significant. There are three factors deteriorating the signal: non-ideal effect of the transmission line, heavy load effect and the discontinuous impedance at the link via. These factors frequently cause signal integrity problems, and seriously restrict the realization of the high-speed and multi-load interconnection. In this paper, the limitation of the conventional daisy-chain topology is analyzed, and a novel three-dimensional daisy-chain topology is proposed by considering print circuit board (PCB) as a three-dimensional (3D) structure. The proposed topology can provide an effective method to design the multi-load interconnection of the higher speed and complexity PCB. As is demonstrated in the case study, the proposed topology can effectively reduce the non-ideal effect of multi-load branch lines and vias, which will greatly increase the noise margin of the loads. It can be seen from the results that, the proposed topology makes the received eye height of the load U3 which is the worst affected optimized for 234.8 mv (76.8%).

Keywords: Daisy-chain topology · High-speed circuit
· Multi-load interconnection · Signal integrity

1 Introduction

With the continuous development of the informatization, high-speed and high-density have become trends in modern and future electronic system designs. However, when the system's operating frequency reaches above 1 GHz, the performance of the entire system will be seriously affected by the interconnection between components. The most important factors are the signal integrity problems such as reflection and crosstalk caused by the non-ideal effect of the transmission link. And it can result in a great distortion of the signal in the transmission process. In the field of high-speed and multi-load interconnection design, the conventional daisy-chain topology is well accepted, to improve the signal quality of multi-load interconnections, as shown in Fig. 1. This is especially used in multi-load integrated design of high-speed DDRX, where the number of PCB

layers is small and the thickness is thin, and the length of the non-ideal branch line is short. In this situation, the designers generally consider PCB as a two-dimensional planar structure, and ignore the non-ideal effect.

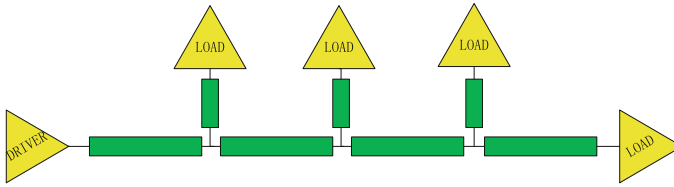


Fig. 1. The diagram of the conventional daisy-chain topology

However, with the rising complexity of PCB, the non-ideal effect of signal via becomes no longer negligible, which will greatly deteriorate the quality of the signal of the multi-load interconnections, such as multi-DDRX interconnections. Therefore, it's urgent to find an effective strategy to consider the non-ideal effect of signal via, and to attenuate this effect.

2 Design and Optimization of New 3D Daisy-chain Topology

In order to improve the signal integrity for the multi-load DDRX design on high-complexity PCB, a novel 3D daisy-chain topology for high-speed and multi-load DDRX interconnection is proposed. The characteristic of the proposed topology includes the interconnection distribution in the PCB layers and the optimization for the transition of the signal via.

2.1 Optimization of the Interconnection Distribution in the PCB Layers

Since the PCB is considered as a three-dimensional structure, the proposed topology offers a strategy to ensure the signal traces are close to the top and bottom layers of the PCB. This strategy requires the branch lines of via to be as short as possible, which means to keep the most of the via cascaded in the main link. Take an implementation of the DDR3's 1-to-8 address signal on a 22-layer PCB as an example.

As shown in Fig. 2, U1 is the driver, U2-U9 are the load receivers. R is the matching termination resistor, and VTT is the matching termination voltage. The signal is sent by the surface driver U1, going through via A to the third or fourth trace layer (near the PCB's surface) and flowing through the transmission line TL1 to via B. And the signal then reaches the load U2 through the branch H1 of via B and passes through the branch H2 of via B to the transmission line TL2 of the 19th or 20th trace layer (near the bottom of the PCB). Then the signal reaches the load U3 through the branch H3 of via B, and so on. The signal flows through the loads in close to the PCB's surface and bottom layer, to ensure that the length of the non-ideal branches H1 and H3 are short enough, while the part H2 of these vias are connected in series to the main link.

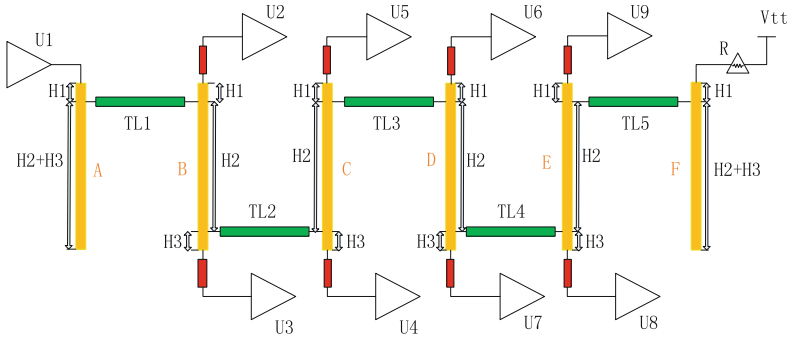


Fig. 2. An illustration of the optimized interconnection distribution

Alternatively, the interconnection distribution can be optimized as shown in Fig. 3. U1 is the driver, U2-U9 are the load receiver. R is the matching termination resistor, and VTT is the matching termination voltage. The signal is sent by the surface driver U1, going through via A to the 19th or 20th trace layer (near the PCB’s bottom), and flowing through transmission line TL1 to via B. The signal then reaches the load U2 through the branch H3 of via B and passes through the branch H2 of via B to the transmission line TL2 of the third or fourth trace layer (near the surface of the PCB). And the signal reaches the load U3 through the branch H1 of via B, and so on. Still, the signal flows the loads in close to the PCB’s surface and bottom layer, to ensure that the lengths of the non-ideal branches H1 and H3 are as short as possible, while the part H2 of these vias are connected in series to the main link.

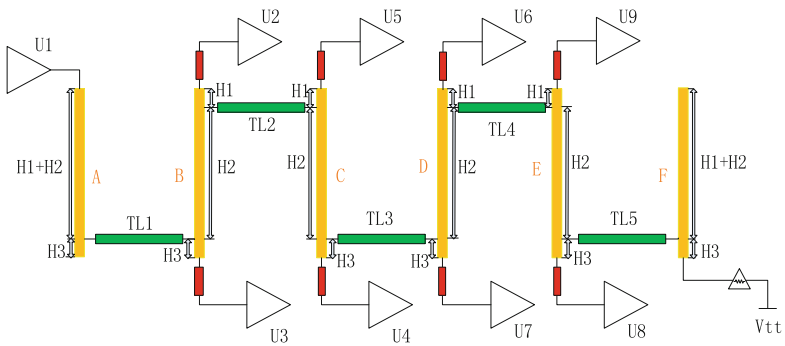


Fig. 3. The illustration of the alternative interconnection distribution

It should be noted that, in both of the two distribution rules, the same group of parallel bus should has uniform distribution rules, and the two distributions should not be mixed to avoid the timing disorder of the grouped signals.

2.2 Optimization for the Transition of the Signal via

Since the design of higher-speed DDRX, the non-ideal effect of via will be further strengthened, two methods are proposed to weaken the non-ideal effect of via, including backdrill and keep out optimization.

As shown in Fig. 4, the via stub is shorten by the backdrilling to weaken non-ideal effect.

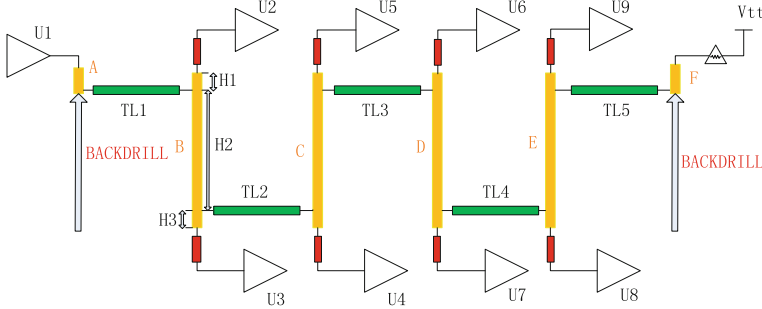


Fig. 4. The illustration of the backdrill

As shown in Fig. 5, the diameter of the keep out is increased without affecting the layout and routing, which is determined by the simulation. This improvement can suppress the parasitic capacitance effect between via and the reference plane, resulting in reducing the impedance discontinuity.

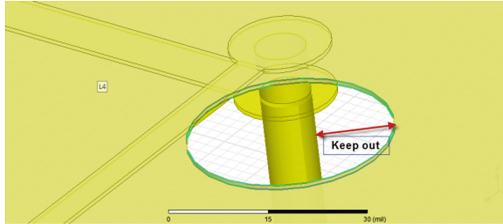


Fig. 5. “keep out” at the link via

3 Analysis of the Proposed 3D Daisy-chain Topology

3.1 3D Daisy-chain Topology Modeling

The HFSS from Ansoft is employed to model the address signal interconnection in a high-complexity 22-layer PCB.

The PCB integrates TI’s C6678 processor (TMS320C6678) and external memory DDR3 (MT41K128M16-16Meg * 16 * 8 banks), resulting in a 1-to-8 address signal topology. According to the proposed topology, the DDR3 address signal interconnection’s parameters are shown in Fig. 6, H1 = 10 mil, H2 = 100 mil, H3 = 10 mil, the

length of transmission line TL1 = 2000 mil, TL2-TL5 = 750 mil, the length of backdrill stub = 110 mil, the length of each load's surface branch TLN = 200 mil, and the matching termination resistor $R = 50 \Omega$.

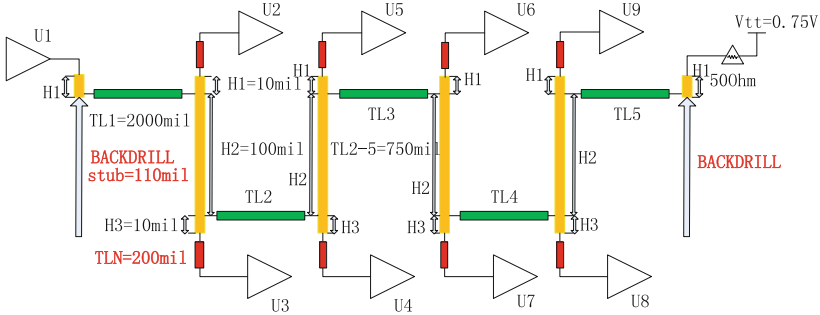


Fig. 6. The parameters of the demonstrated 3.1 3D daisy-chain topology

In the other hand, the parameters are illustrated in Fig. 7 by using conventional daisy-chain topology. The signal is generated by the surface driver U1, reaching 10th or 11th layer (intermediate routing layer) through the branch of via A. Then the signal passes the middle line layer, and reaches via B–via F in turn. Another way, the signal reaches the load U2-U9 and 50 Ω termination resistor. In this case, H1 = H2 = 60 mil, the transmission line TL1 = 2000 mil, TL2-TL5 = 750 mil, the length of each load's surface branch TLN = 200 mil.

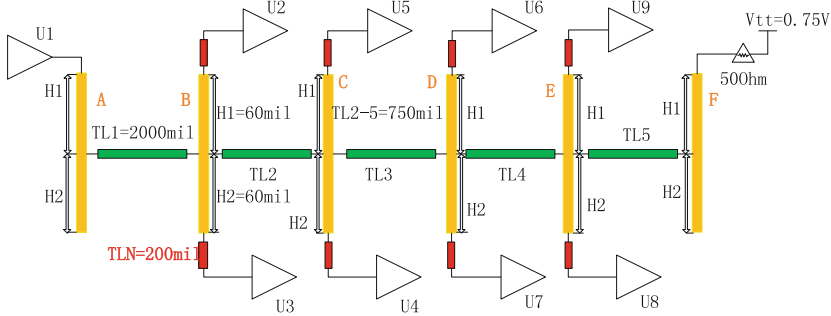


Fig. 7. Parameter diagram for conventional topology

The conventional daisy-chain topology and the new 3D daisy-chain topology are modeled by HFSS. According to the simulation in Si9000, the surface line's width, the stripe line's width TL1 and the width TL2–TL5 are 5 mil, 8 mil and 4 mil. Then build the model according to the parameters (in Figs. 6 and 7), as shown in Fig. 8.

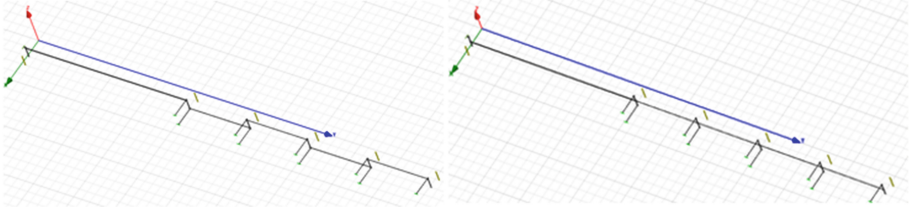


Fig. 8. New daisy-chain topology model and conventional daisy-chain topology model

And the wave port and radiation boundary are then set. The solution frequency range is set as 2 GHz to obtain high-accuracy S parameters, and 0 MHz-1 GHz is set as the sweep range and 10 MHz as the frequency sweep step.

3.2 Simulation Results and Analysis

With the resulted S parameter data of the two daisy-chain topology links, the entire interconnection including driver and receiver ibis models (DSP C6678 tms320c6678 4 2 1 and DDR3 MT41J512M16HA) is setup in the Hyperlynx9.0 provided by Mentor Graphics as illustrated in Fig. 9. A PRBS sequence is excited at a rate of 0.8 Gbps, and the resulted eye diagram at the U2 are shown in Figs. 10 and 11.

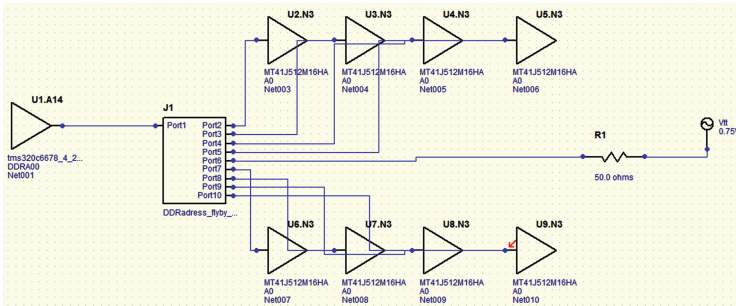


Fig. 9. The entire interconnection including driver and receiver ibis models in Hyperlynx

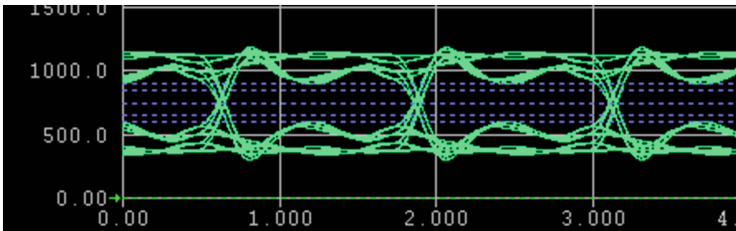


Fig. 10. Eye diagram of conventional daisy-chain topology at U2

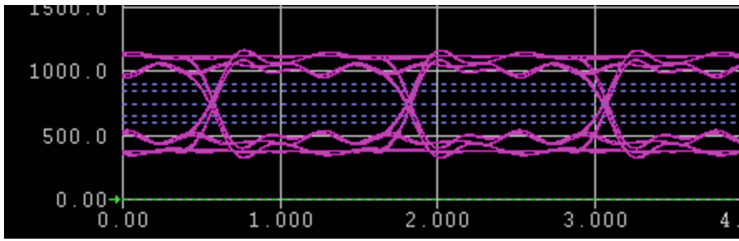


Fig. 11. Eye diagram of proposed daisy-chain topology at U2

The detailed results of the eye diagram parameters including eye height and eye width are listed in Tables 1 and 2, respectively.

Table 1. Eye height's contrast of two topologies

Eye height's contrast (mv)				
	U2	U3	U4	U5
Normal	305.6	305.6	334.7	331.4
Optimized	466.5	540.4	459.9	473.9
Rate	52.70%	76.80%	37.40%	43%
	U6	U7	U8	U9
Normal	444.5	450.9	624.2	621.5
Optimized	455.6	515.5	620.7	598.6
Rate	2.50%	14.30%	-0.50%	-3.70%

Table 2. Eye width's contrast of two topologies

Eye width's contrast (ns)				
	U2	U3	U4	U5
Normal	1.185	1.183	1.206	1.206
Optimized	1.21	1.216	1.216	1.224
Rate	2.10%	2.70%	0.80%	1.50%
	U6	U7	U8	U9
Normal	1.191	1.192	1.23	1.229
Optimized	1.228	1.217	1.229	1.225
Rate	3.10%	2%	-0.08%	-0.30%

It can be seen from the results that, the proposed topology makes the received eye heights of the loads U2, U3, U4, and U5 optimized for 160.9 mv (52.7%), 234.8 mv (76.8%), 125.2 mv (37.4%), and 142.5 mv (43%), compared with the conventional daisy-chain topology. Especially, the optimization does not sacrifice the eye widths of each load.

Therefore, the proposed daisy-chain topology can effectively reduce the non-ideal effect of the multi-load branches, and greatly increase the noise margin of the loads in the high-speed and multi-load interconnection design.

4 Conclusion

With the great demand for optimization in the high-speed and multi-load interconnection design, a novel daisy-chain topology is proposed as a prominent improvement to the conventional daisy-chain topology. The characteristic of the proposed topology includes the interconnection distribution in the PCB layers and the optimization for the transition of the signal via. The former weakens the branch effect of the signal via, while the latter weakens the stub effect of the signal via. In this paper, the proposed topology makes the received eye height of the load U3 which is the worst affected optimized for 234.8 mv (76.8%). The novel topology offers an effective means for the multi-load interconnection in the higher-speed and higher-complexity PCB.

Acknowledgments. This work was supported in part by the Chang Jiang Scholars Program under Grant T2012122, in part by the Hundred Leading Talent Project of Beijing Science and Technology under Grant Z141101001514005.

References

1. Signal integrity and PCB layout considerations for DDR2–800 Mb/s and DDR3 Memories, Virtex-6 FPGA Select IO Resources UG361 (v1.6), November 2014
2. Xilinx Virtex-6/Spartan-6 FPGA DDR3 Signal Integrity Analysis and PCB Layout Guidelines, Syed Bokhari (Fidus System Inc.) and Romi Mayder (Xilinx Inc.) WP420 (v1.0), June 2012
3. Chang, W.-Y., See, K.-Y., Hu, B.: Characterization of component under DC biasing condition using an inductive coupling approach. *IEEE Trans. Instrum. Meas.* **59**(8), 2109–2114 (2010)
4. Sampath, M.K., Atout, N.: Signal integrity validation of de-embedding techniques using accurate transfer functions. In: *Consumer Electronics (ISCE)*, pp. 1–4, 4–6 June 2012
5. Guo, W.-D., Lin, J.-H., Lin, C.-M., Huang, T.-W., Wu, R.-B.: Fast methodology for determining eye diagram characteristics of lossy transmission lines. *IEEE Trans. Adv. Packag.* **32**(1), 175–183 (2009)
6. Antonini, G., Scogna, A.C., Orlandi, A.: Equivalent network synthesis for via holes discontinuities. *IEEE Trans. Adv. Package* **25**(4), 175–183 (2009)
7. Hall, S.H., Hall, G.W., McCall, J.A.: *High-Speed Digital System Design: A Handbook of Interconnect Theory and Design Practices*. Wiley, New York (2000)
8. Kim, J.-H.: Analysis of via in multilayer printed circuit boards for high-speed digital systems. In: *Proceedings of 3rd International Symposium on Electronic Materials and Packaging Conference*, pp. 382–387, November 2001
9. Connor, S., Archambeault, B., Diepenbrock, J.C.: The impact of external RF energy on high-speed differential signal quality of long cables. In: *IEEE International Symposium on Electromagnetic Compatibility, EMC 2008*, pp. 1–4, 18–22 August 2008

Realization and Optimization of Pulse Compression Algorithm on OpenCL-Based FPGA Heterogeneous Computing Platform

Jiacheng Yu¹, Xingming Li²(✉), Shanqing Hu¹, and Yuwei Wang¹

¹ Beijing Key Laboratory of Embedded Real-Time Information Processing Technology, Beijing Institute of Technology, Beijing 100081, China

² Tsinghua University, Beijing 100081, China

lxm11544@163.com

Abstract. The development of modern radar signal processing technology put forward higher requirements for processor performance. However, Moore's law encounters bottlenecks, the computational performance of general-purpose processors is constrained and can not meet application requirements. The high-performance and low-power features of FPGA make them recently become of interest in research as a heterogeneous computing platform together with CPU. Pulse compression algorithm is widely used in the field of radar signal processing, which contains a large number of floating-point computing, the processing effect largely depends on the performance of the processor. Based on Open Computing Language (OpenCL), we first evaluated the Fast Fourier Transform (FFT) of various sample sizes on Arria10 FPGA board and FPGA achieve up to 33.5 times the performance improvement compared to DSP C6678 on processing different sample size of FFT. Then we realize a $4\text{ K} \times 8\text{ K}$ size pulse compression processing using kernel channel. The results show that the core computation implemented on Arria10 FPGA through OpenCL is approximately 10x faster than DSP C6678 for $4\text{ K} \times 8\text{ K}$ size pulse compression processing.

Keywords: FPGA · OpenCL · Heterogeneous computing · FFT
Pulse compression algorithm · Kernel channel

1 Introduction

In recent years, there has been a significant shift towards parallel computing architectures, the primary reason is that with the development of chip technology to the 10 nm stage, Moore's Law encountered bottlenecks, the computing performance and energy efficiency of generic processor can not meet the ever-increasing and diverse data processing needs [1]. At present, it has become the academic and industrial research hotspot to make the GPU, DSP, FPGA and other coprocessors together with the CPU to form a heterogeneous processing platform to achieve the improvement of computing performance. FPGAs have many advantages over other processors, such as high parallelism and flexibility, low power, customized circuits, etc. In particular, the performance to power ratios of FPGA make them more attractive for many applications.

At the same time, FPGA is also facing a series of challenges. The first and foremost is that applications in FPGA are designed using low level Hardware Description Languages (HDL), the traditional process involves the establishment of data access, setting the state machine, calling the underlying IP and solving the problem of timing closure, which requires developers have sufficient knowledge of the internal structure of FPGA and HDL design experience, which greatly limits the application and promotion of FPGA as a heterogeneous processor.

OpenCL is a cross-platform parallel programming model based on C/C++ for heterogeneous computing platforms and it has come to light as a solution to the issues above. As a cross-platform programming language, OpenCL provides a new development method for FPGA, which has the advantages of short development cycle, high level of abstraction, and high portability.

This paper is organized as follows. Firstly, we provide a brief introduction on the hardware platform and the programming model used in this research. Secondly, we evaluate the performance of different points of FFT on Arria10 FPGA. Next, we provide an overview of pulse compression algorithm, and describe an implementation of this algorithm in OpenCL. We then discuss how this code is optimized for FPGA. Finally, we compare the best results achieved with DSP C6678 and present concluding remarks.

2 FPGA Heterogeneous Computing Platform

2.1 OpenCL Programming Model

As shown in Fig. 1, the application is composed of two sections: the host program, and the kernel. The host program is the serial portion of the application, which is written in standard C/C++ and can run on a variety of types of microprocessors, such as hard ARM processor embedded in FPGA, or external x86 processor, and the host program is responsible for managing data and control flow of the algorithm [2]. The kernel program is the highly parallel part of the application to be accelerated on a device such as a multi-core CPU, GPU, or FPGA.

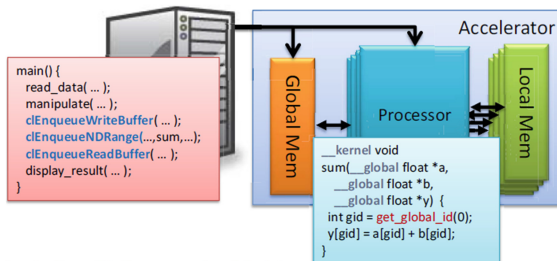


Fig. 1. OpenCL programming model [1]

2.2 Introduction of the Hardware Platform

Our research is based on Intel Arria10 FPGA. As shown in Fig. 2, the external x86 processor work as the host of the heterogeneous platform, the CPU interacts with the FPGA via the PCIe Gen3 bus. And the two groups of external DDR3 memory work as a global memory.

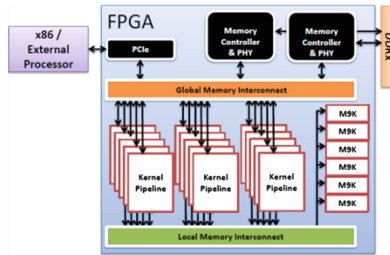


Fig. 2. The hardware platform [2]

Using Intel SDK for OpenCL, we can map the OpenCL kernel code on the FPGA, the compiler automates the integration, layout, timing analysis, and generates the FPGA binary configuration file (.aocx). For example, consider a simple vector addition kernel shown in Fig. 3a. The kernel describes the vector addition of two arrays, and each thread is assigned an ID (`get_global_id(0)`) to mark the data processed by each thread. The threads are executed in parallel, running the same code but processing different thread. The translation to hardware will result in the high level circuit structure shown in Fig. 3b, including three main units: data loading, vector addition and data storage.

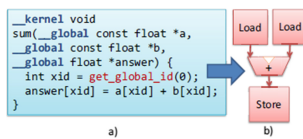


Fig. 3. Vector addition FPGA implementation [2]

The most important concept behind the FPGA parallel acceleration is the notion of pipeline processing parallelism. In order to simplify the description, assuming that the kernel is mapped to a three-stage pipeline on the FPGA. As shown in Fig. 4, on the first clock cycle, thread 0 is loaded into the load module. On the second clock cycle, thread 1 is loaded, at the same time, thread 0 has been read from the external memory and entered into the next register for addition operation. On the third clock cycle, thread 2 is loaded, thread 1 has completed reading, thread 0 has completed the addition operation and enter the storage module. It is evident that in the steady-state, all parts of the pipeline are active, with each stage processing a different thread.

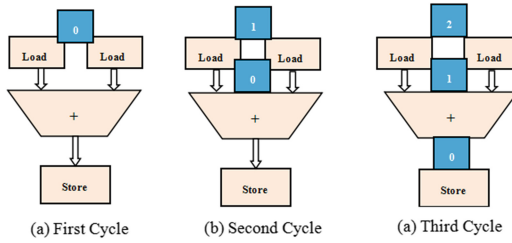


Fig. 4. Example of pipeline parallelism

2.3 Kernel Communication Mechanism

As shown in Fig. 5, assume that a task contains two sub-tasks: kernel 1 and kernel 2, and there is a production-consumption relationship between the two kernels, that is, the output of kernel 1 is then input to kernel 2. In standard OpenCL model, the data between the different kernel need to go through the global memory for interaction. First, kernel 1 read the data from the global memory and the results after processing will be written back into the global memory, then kernel 2 read the results of kernel 1 from the global memory for further processing, the resulting input and output delay will reduce the performance of the system.

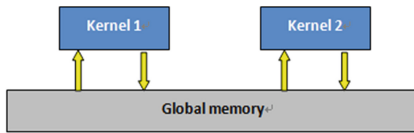


Fig. 5. Standard OpenCL model

Kernel-channel is a kind of kernel communication mechanism provided by Intel FPGA, which allows different kernels interact with each other directly through the FIFO. As shown in Fig. 6, the current thread of data output from kernel 1 is sent directly to kernel 2 by the kernel channel for further processing, while kernel 1 can process the next thread. The pipeline is in the steady-state after a number of clock cycles, kernel 1 and kernel 2 run in parallel processing different thread at the same clock cycle.

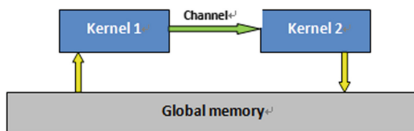


Fig. 6. Kernel communication mechanism

3 Performance of FFT on FPGA Based on OpenCL

Fast Fourier Transform (FFT) is widely used in a variety of radar signal processing algorithms, which contains a large number of floating-point processing and can be used to effectively evaluate the performance of a processor. In this paper, we implement a radix-4 FFT engine capable of processing eight data points per clock cycle on Arria10 FPGA based on OpenCL and evaluate the performance for different sample size of FFT.

The evaluation results are shown in Table 1. We list the best performance of FFT achieved by our research group on Arria10 FPGA and DSP C6678 and calculate the floating-point processing performance (GFLOPS) achieved by FPGA for various sample size of FFT.

Table 1. FFT evaluation results

FFT size	Number of iterations	Arria10 FPGA average time (us)	DSP C6678 (us)	FPGA GFLOPS	
512	10000	0.75	5.2	30.7	
1 K		1.38	10.0	37.1	
2 K		2000	2.25	23.0	50.1
4 K			5.7	52.1	43.1
8 K			10.9	225.4	48.8
16 K	1000	21.9	472.0	52.3	
32 K	500	53.6	1800	45.8	

We use the method of calculating the average time of calculating multiple sets of FFT. From the table, we can conclude that FPGA achieve up to 33.5 times the performance improvement compared to DSP C6678 on processing different size of FFT algorithm. We also find that FPGA has no advantage over DSP when calculating only one set of FFT, however, for the average time of calculating multiple sets of FFT, FPGA is much shorter than that of the DSP. This is mainly because all parts of the pipeline of FPGA will be fully active only when calculating multiple sets of FFTs.

The method of calculating GFLOPS is given by Eq. (1), where N is the sample size of FFT and T is the average time of calculating N sets of FFTs.

$$GFLOPS = 5 \times N \times (\log N / \log 2) / T \quad (1)$$

Depending on the end application and the available FPGA resources, more instances of this engine can be instantiated for higher performance, subject to the memory bandwidth available on the OpenCL board.

4 Pulse Compression Algorithm

For radar system, the pulse width is proportional to the radar energy and the detection distance, and is inversely proportional to the distance resolution. Using pulse compression technology, we can achieve a larger detection range while achieving a higher distance resolution. Pulse compression technology is widely used in a variety of radar signal processing algorithms. Pulse compression processing involves a large number of floating point operations such as FFT, multiplication and IFFT, which can be used to effectively evaluate the performance of the processor and also has a reference significance for practical application.

The pulse compression technique needs to match filter the echo signal of the transmitted pulse to compress the echo into a narrow pulse, thus improving the signal-to-noise ratio and the distance resolution of the received signal. The algorithm processing flow is shown in Fig. 7.

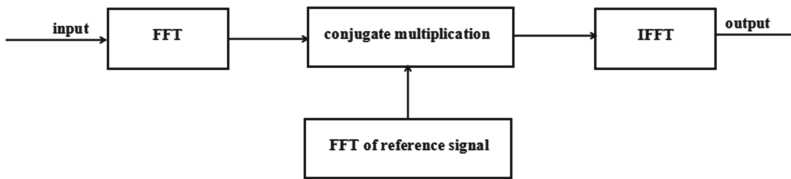


Fig. 7. Pulse compression algorithm processing flow

As can be seen from the figure, the pulse compression processing includes three sub-tasks with production-consumption relationship: FFT, conjugate multiplication and IFFT, and the data to be processed is independent for multiple groups of PRT, which is suitable for FPGA to do pipeline processing.

5 Realization and Optimization of Pulse Compression Algorithm

Based on the previous evaluation of the FFT algorithm, we realize $4\text{ K} \times 8\text{ K}$ size pulse compression processing, we simulated 4096 groups of PRT, each PRT contains 8 K sampling points. We mapped three kernel on the Arria10 FPGA: 8 K point FFT, conjugate multiplication, and 8 K point IFFT. In order to evaluate the performance of kernel channel, we used two kinds of programs to realize pulse pressure algorithm for comparative analysis. The working mode of program 1 is described in Fig. 8, three sub-tasks work in serial way, and data interaction between kernels go through the global memory under the management of the Host processor. For program 2, we add a kernel-channel between FFT and conjugate multiplication for optimizing, which send the data output from FFT module directly to conjugate multiplication module for further processing. The working mode is shown in Fig. 9.

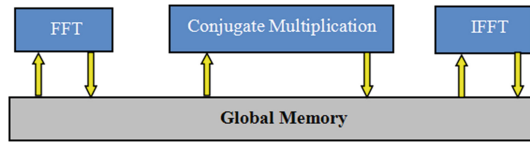


Fig. 8. Working mode of Program 1

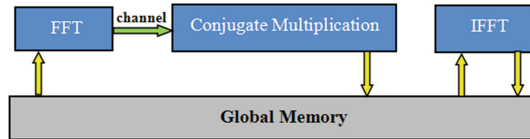


Fig. 9. Working mode of Program 2

We implemented the two programs on Arria10 FPGA and tested the kernel execution time, the result is shown in Table 2.

Table 2. Result of pulse compression for two programs

	Program 1 (ms)	Program 2 (ms)
FFT	41.2	41.6
Conjugate multiplication	135.4	
IFFT	79.3	79.4
Kernel execution time	255.9	121.0

From the test results can be seen, the use of channel-kernel for optimization effectively shorten the kernel execution time. We compare the best results achieved on Arria10 FPGA and DSP C6678 for $4\text{ K} \times 8\text{ K}$ size pulse compression processing and the result is shown in Table 3.

Table 3. Results of pulse compression for Arria10 FPGA and DSP C6678

	Arria10 FPGA	DSP C6678
Execution time (ms)	121	1200

As can be seen from the result in Table 3, Arria 10 FPGA achieve approximately 10 times increase in performance compared to the DSP C6678 for $1\text{ K} \times 4\text{ K}$ size pulse compression. FFT is the main process of pulse compression algorithm, so the execution time of the pulse compression algorithm mainly depends on the capability of calculating FFT. And we optimize our code by using kernel channel on Arria10 FPGA to shorten the data latency between different kernels, thus improving overall performance of FPGA. However, The interaction with the memory consumes a lot of time on DSP.

6 Conclusion

In this paper, we introduce the results of the FFT and pulse compression algorithm Arria10 FPGA based on OpenCL. compared with DSP C6678, we find that for different FFT algorithms, FPGA achieves up to 33.5 times performance improvement; for $4\text{ K} \times 8\text{ K}$ particle size pulse compression algorithm, FPGA achieves 10 times the performance improvement. It is suitable for the use of FPGA to accelerate for the tasks that can be decomposed into several sub-tasks with production-consumption relationship, and it is effective for the use of kernel communication mechanism to optimize the task to shorten the kernel execution time. In fact, the OpenCL flow automates many complexities of external interfacing such as DDR and PCIe. The OpenCL design flow is quite simple to comprehend and produces an implementation.

Acknowledgments. This work was supported in part by the Chang Jiang Scholars Program under Grant T2012122, in part by the Hundred Leading Talent Project of Beijing Science and Technology under Grant Z141101001514005.

References

1. Chen, D., Singh, D.: Fractal video compression in OpenCL: an evaluation of CPUs, GPUs, and FPGAs as acceleration platforms. In: Design Automation Conference. IEEE, pp. 297–304 (2013)
2. Chen, D., Singh, D.: Invited paper: using OpenCL to evaluate the efficiency of CPUS, GPUS and FPGAS for information filtering. In: International Conference on Field Programmable Logic and Applications. IEEE, pp. 5–12 (2012)
3. Ibrahim, M., Khan, O.: Performance analysis of fast fourier transform on field programmable gate arrays and graphic cards. In: International Conference on Computing, Electronic and Electrical Engineering. IEEE, pp. 158–162 (2016)
4. Jacquin, A.E.: Image coding based on a fractal theory of iterated contractive image transformations. *IEEE Trans. Image Process.* **1**(1), 18–30 (1992)
5. Ayat, S.O., Khalil-Hani, M., Bakhteri, R.: OpenCL-based hardware-software co-design methodology for image processing implementation on heterogeneous FPGA platform. In: IEEE International Conference on Control System, Computing and Engineering. IEEE, pp. 36–41 (2016)
6. Muslim, F., Liang, M., Roomez, M., et al.: Efficient FPGA implementation of OpenCL High-Performance Computing applications via High-Level Synthesis. *IEEE Access* **5**, 2747–2762 (2017)
7. Waidyasooriya, H.M., Takei, Y., Tatsumi, S., et al.: OpenCL-Based FPGA-Platform for stencil computation and its optimization methodology. *IEEE Trans. Parallel and Distrib. Syst.* **28**(5), 1390–1402 (2017)
8. Arcas-Abella, O., Ndu, G., Sonmez, N., et al.: An empirical evaluation of high-level synthesis languages and tools for database acceleration. In: International Conference on Field Programmable Logic and Applications. IEEE, pp. 1–8 (2015)
9. Waidyasooriya, H.M., Hariyama, M.: FPGA-based deep-pipelined architecture for FDTD acceleration using OpenCL. In: IEEE International Conference on Computer and Information Science. IEEE, pp. 1–6 (2016)

10. Hill, K., Craciun, S., George, A., et al.: Comparative analysis of OpenCL vs. HDL with image-processing kernels on Stratix-V FPGA. In: International Conference on Application-Specific Systems, Architectures and Processors. IEEE, pp. 189–193 (2015)
11. Firmansyah, I., Yamaguchi, Y., Boku, T.: Performance evaluation of Stratix V DE5-Net FPGA board for high performance computing. In: International Conference on Computer, Control, Informatics and its Applications. IEEE (2017)

The Research of SAR Processing Performance Based on Multi-core GPU

Yuwei Wang¹, Xingming Li²(✉), Shanqing Hu¹, and Jiacheng Yu¹

¹ Beijing Key Laboratory of Embedded Real-Time Information Processing Technology,
Beijing Institute of Technology, Beijing 100081, China

² Tsinghua University, Beijing 100081, China
lxm115444@163.com

Abstract. With the characteristics of large data volume, high algorithm complexity and large computational complexity, Synthetic Aperture Radar (SAR) technology which makes the signal processing system have to be improved continuously in the aspects of real-time, storage capacity, data throughput and computing capability. As a kind of multi-core architecture, Graphics Processing Unit (GPU) take the advantages of powerful computing capability and efficient storage bandwidth to meet the urgent need in scalability, computing capability and storage bandwidth for large-scale data parallel applications. In this paper, the first thing is to evaluate the FFT performance of the NVIDIA Tesla M6 GPU, which achieves an average 41x speedup ratio compared to TI's TMS320C6678 DSP. Then, the RD (Range Doppler) algorithm which is the most classical SAR imaging algorithm is implemented on the platform of CPU + GPU using CUDA language, and execution time of the SAR algorithm for 4 K × 8 K point is shortened by 1.18 s and the result shows that GPU achieve 1.9x the performance improvement compared to DSP C6678 on RD-SAR algorithm.

Keywords: Multi-core · GPU · SAR · Performance evaluation

1 Introduction

As the high-resolution imaging radar, with its superior performance, Synthetic Aperture Radar (SAR) is widely used in military and civilian areas. With the continuous development of SAR technology, processing data scale is increasing so fast that the requirements of signal processor computational complexity, imaging accuracy and many other factors are getting higher and higher. In terms of the initial single-core processor, the way to enhance the performance of the processor is to improve the processor frequency, but by the chip production process constraints, the benefits of the power consumption are covered up by the power consumption and yield problems. As an effective method to enhance the system processing capacity further, multi-core parallel processing and multi-processor parallel processing appeared.

Currently, the mainstream digital signal processors include DSP, FPGA and GPU. The most widely used processors among them, for instance, the peak processing performance of DSPC6678 is 160GFLOps, and Intel Arria10 SoC FPGA is 1.5TFLOps,

while the peak processing performance of Tesla M6 GPU can reach 3.2TFlops. According to this, as a multi-core architecture processor, GPU which has strong floating-point computing capability is a general-purpose processor that used to achieve high performance parallel operation in the ultra-multi-stream processor platform. Compared to other processors, GPU has obvious advantages in terms of processing capability and memory bandwidth. At the same time, the mainstream GPU manufacturer NVIDIA launched a computing architecture CUDA in 2006 which is combined with a comprehensive software platform to break through the hardware programmability and development constraints, using a kind of language that is easy to understand like C language, and do not need to use the graphical API. So it will give the full exert to the GPU's powerful computing capability when building the high-performance applications. CUDA calculation model is working in the CPU + GPU heterogeneous mode, the CPU is the Host, and the GPU is the secondary processor or Device. With the standardization of the language, using CUDA for GPU software development has been widely used in high-performance computing.

First of all, this paper evaluates the FFT performance of the Tesla M6 GPU; Then, introduces the basic principle of SAR algorithm, studying the RD algorithm flow in the classical SAR algorithm, and analyzing the characteristics of the RD algorithm in detail; Finally, the RD_SAR algorithm is implemented on the CPU + GPU platform using CUDA language, and comparing the characteristics of the multi-core processing platform and DSPC6678 parallel processing platform in RD algorithm processing.

2 FFT Performance Evaluation of GPU

Fast Fourier transform (FFT) is often used in digital signal processing to obtain characteristics of the signal in the frequency domain. FFT is often used to evaluate the performance of a processor due to the features of computationally intensive and time-consuming.

In this paper, we implement a radix-4 FFT on Tesla M6 GPU based on CUDA and use different size of FFT to evaluate the processor's performance. Using the CUFFT math library in CUDA's official release, we can process one-dimensional, two-dimensional or three-dimensional fast Fourier transforms of multi-batch in parallel, and we can process fast Fourier transforms of multiple batches at the same time. Thus the multi-batch fast Fourier transform in complex domain is realized. And then use the average time of multiple calculations' results in different points FFT to get the final time.

The final test results are listed in Table 1 and the comparison chart with the processing results of DSPC6678 is shown in Fig. 1. Wherein, the comparison contains three sets of data: Group A, Group B and Group C.

Group A: The result is based on TI's official library dsplib.ae66, using eight cores of DSPC6678 to implement FFT in parallel;

Group B: The result is based on a new algorithm, VLFFT, which is designed for large points FFT, has a certain improvement over the efficiency of Group A, using eight cores of DSPC6678 to implement FFT in parallel;

Group C: The result is based on Tesla M6 GPU to implement FFT.

Table 1. The FFT results of DSP and GPU

Processing time(us)			
FFT points	Group A	Group B	Group C
512	5.2		0.09
1024(1 K)	10.0		0.18
2048(2 K)	23.0		0.34
4096(4 K)	52.1	32.0	0.66
8192(8 K)	225.4	51.7	1.32
16384(16 K)	472.0	85.2	5.40
32768(32 K)	1800.0	162.1	10.80

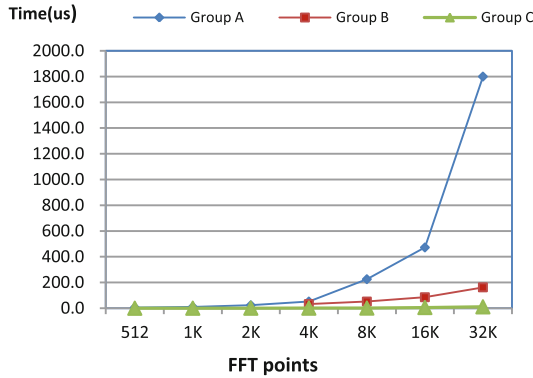


Fig. 1. The FFT results of DSP and GPU

It can be seen from the data above, GPU (C group) is 56x–578x faster than DSPC6678 (A group), is 15x–48x faster than DSPC6678 (B group). If we take two sets of DSP to achieve the faster time, then the GPU can achieve an acceleration ratio of about 15 to 67 times compared to the DSP C6678, that is, an average of 41 times acceleration ratio.

3 Research on SAR Processing Performance of Multi-core GPU

This research is based on a high-performance processing system board (Fig. 2) which consists of a CPU + GPU architecture and used in the OpenVPX platform. The main part of this platform is composed of NVIDIA Tesla M6 GPU module and Express-SL7 i68- E22 ComE CPU module. CPU is responsible for the master work, including bus management and data distribution, while GPU, the key module of data processing, is mainly on deal with the relatively large amount of data in parallel processing, helping CPU to process data together. The communication of CPU and GPU is the PCIE protocol based on the VITA specification.

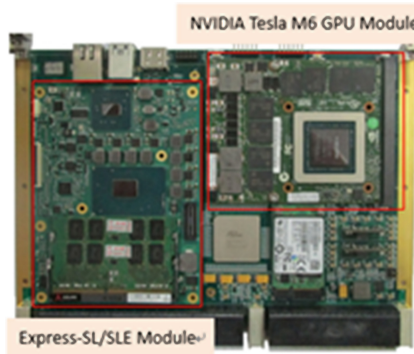


Fig. 2. The CPU + GPU board

3.1 Synthetic Aperture Radar (SAR) Algorithm

SAR processing is to extract the two - dimensional distribution of the scattering coefficient in the target area from the received echo data. It is essentially a two-dimensional data processing, the usual method is to decompose the two-dimensional data into the distance (Y-axis) and azimuth (X-axis) system, so the imaging process is essentially a two-dimensional matched filtering process. This research will study on RD (Range Doppler) algorithm which is one of the classical SAR imaging. The idea of the RD algorithm is to convert the two-dimensional imaging processing of the synthetic aperture radar echo data into a one-dimensional processing by the two matched filtering operations of the range direction and the azimuth direction. The typical data processing flow of RD algorithm is shown in the Fig. 3.

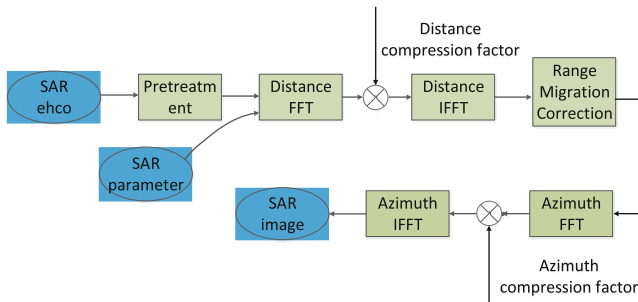


Fig. 3. RD algorithm diagram

Firstly, from the algorithm diagram above, RD algorithm has the characteristics of large amount of data and large amount of computing, thus it is suitable for the processing structure of parallel flow type. As a kind of processor with powerful parallel calculating ability, GPU have more transistors for data processing, causing it is very suitable for the realization of RD algorithm.

Secondly, there are strict dependencies between the key steps of the RD algorithm, the input data of each processing module is the previous module's output data. Therefore, the processing modules can't be separated and distributed separately in different processing cores of the GPU. And it has to put all the computing resources into the current module, when the module is processed before the next operation.

Finally, the FFT and IFFT which are the main step of RD algorithm are widely used in this algorithm, Therefore, the efficiency of FFT on CUDA platform is an important affecting factor of the whole program's performance.

3.2 The Performance Analysis of SAR Algorithm Based on GPU

According to the RD algorithm above, this research is based on 8 K point FFT which is used most widely in practice, adopting $4\text{ K} \times 8\text{ K}$ complex points of the echo data to realize RD algorithm in the GPU parallel processing system, that is, the total calculation is the 4000 groups of 8 k points one-dimensional FFT. Finally, we can get a complete image (Fig. 4) of the two sets of echo data after the processing of RD algorithm completed, it shows a clear river.

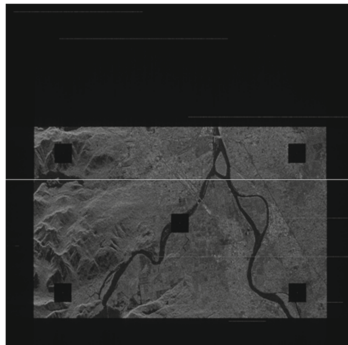


Fig. 4. RD SAR Image

Through the description of the function above, the main parts of the radar signal processing include the pulse compression process of the radar echo data on range direction, the FFT on azimuth direction process, the matched filtering on azimuthal direction and quantization. In the realization of the SAR algorithm, the optimization is mainly from the following two aspects.

On the one hand, as we all know, the transmission bandwidth between Intel CPU and DDR3 is approximately 25 GB/s, however, the transmission bandwidth between Nvidia GPU and GDDR5 can reach 200 GB/s, simultaneously, the communication of CPU and GPU is the PCIE protocol and its bandwidth is 16 GB/s in theory and it can reach 9.6 GB/s in reality through the program test results. This shows that the rate of PCIE is far from meeting the GPU processing requirements, and it is the slowest part of the GPU program. So, in the implementation of the program, the optimization principle

is reducing the transmission between host data and device data as far as possible and the allocation, operation and release of the process data can operate directly on the GPU.

On the other hand, the optimization of the kernel function is the key part to achieve high-performance GPU program, the optimization methods generally start from the following two aspects: memory access optimization and instruction optimization. This program has been carefully optimized for the matrix transpose kernel, and it has a greater impact on overall performance. We use Shared Memory for memory optimization in the process of matrix transpose, but the shared memory of NVIDIA GPU is generally small, we need to divide the data into blocks according to the size of shared memory, and this process is commonly called tile operation. At the same time, in order to avoid bank conflict when using the shared memory, the size of the two-dimensional tile is generally expressed as $[TILE_DIM] [TILE_DIM+1]$, which $TILE_DIM$ is the one-dimensional size of tile. Through the thread index settings, the design of the transfer kernel function using shared memory can achieve double transposition, when the location of the tile in the entire input data transpose adjustment, the transpose operation is also going on inside the tile. The transpose module is shown in the Fig. 5.

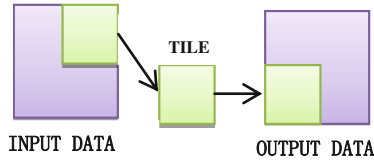


Fig. 5. Transpose kernel module

After testing precisely, the results of the two different processing systems are compared as follows (Table 2 and Fig. 6). Wherein, the comparison contains two sets of data: Group A and Group B.

Table 2. RD_SAR Processing Time of DSP and GPU

Processing time(ms)		
RD algorithm	Group A	Group B
1. SarRangeComp ()	1199.83	241.10
2. SarAziFFt ()	306.94	203.52
3. SarAziComp ()	537.48	524.39
4. ResultDisplay Raw ()	452.73	343.55
5. Total Time	2496.98	1312.56

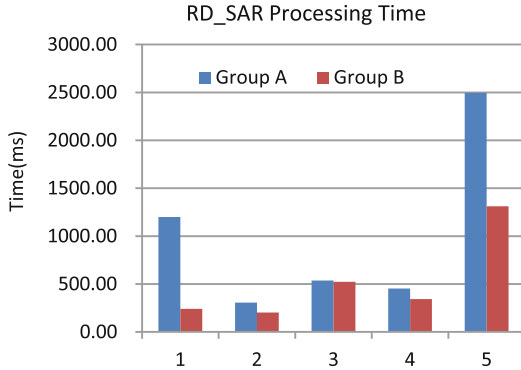


Fig. 6. RD_SAR processing time of DSP and GPU

Group A: The result is based on using eight cores of DSPC6678 to implement RD_SAR in parallel;

Group B: The result is based on using Tesla M6 GPU to implement RD_SAR.

As we can see from the results, GPU has an obvious achievement on the acceleration of the RD algorithm compared with DSPC6678. And the total processing time of SAR algorithm for $4\text{ K} \times 8\text{ K}$ point is shortened by 1.18 s compared with the existing DSPC6678 processor. The result shows that the execution of the algorithm on GPU is approximately 1.9x faster than DSPC6678.

Simultaneously, GPU is 1x–5x faster than DSP in each main part of the RD algorithm, especially in the step of the distance pulse compression, GPU obtain 4.98 times' speed ratio than DSP. Then analyzing the main steps of pulse compression, the following Table 3 gives the results in detail.

Table 3. Times of the distance pulse compression

Distance pulse compression	Time(ms)
4 K × 8 K FFT (echo data)	5.42
4 K × 8 K FFT (reference data)	5.41
4 K × 8 K FFT Complex multiplication	49.18
4 K × 8 K IFFT	5.43
Normalization	93.94

From Table 3, FFT, IFFT is the main process of the pulse compression on range direction. According to the analysis of the previous part, GPU has a substantial performance improvement in the FFT calculation comparing to the DSP, the results of the RD algorithm are the strong proof of this conclusion.

4 Conclusion

With the development of SAR real-time imaging system towards high precision, high real-time and high data throughput, while optimizing the algorithm, it also has higher

and higher requirements on the hardware structure. So choosing the right processor to handle large amounts of data becomes more and more important. In this paper, GPU multi-core processor is the main researching object, and the RD_SAR algorithm is implemented on the CPU + GPU platform using the CUDA calculation model. It effectively demonstrated GPU has powerful computing ability in the high-performance area and it also proved that GPU has a very significant effect in the acceleration of the RD algorithm. GPU can satisfy the real-time requirements of radar signal processing better.

Acknowledgments. This work was supported in part by the Chang Jiang Scholars Program under Grant T2012122, in part by the Hundred Leading Talent Project of Beijing Science and Technology under Grant Z141101001514005.

References

1. Songm, M.C., Liu, Y.B., Zhao, F.J.: Processing of SAR data based on the heterogeneous architecture of GPU and CPU. In: Radar Conference 2013, IET International, pp. 1–5. IET (2013)
2. Tang, H., Li, G., Zhang, F.: A spaceborne SAR on-board processing simulator using mobile GPU. In: IGARSS 2016, 2016 IEEE International Geoscience and Remote Sensing Symposium, pp. 1198–1201. IEEE (2016)
3. Baier, G.: GPU-based nonlocal filtering for large scale SAR processing. In: Geoscience and Remote Sensing Symposium, pp. 7608–7611. IEEE (2016)
4. Frey, O., Werner, C.L., Wegmuller, U.: GPU-based parallelized time-domain back-projection processing for Agile SAR platforms. In: Geoscience and Remote Sensing Symposium, pp. 1132–1135. IEEE (2014)
5. Peternier, A., Defilippi, M., Pasquali, P.: Performance analysis of GPU-based SAR and interferometric SAR image processing. In: Synthetic Aperture Radar, pp. 277–280. IEEE (2014)
6. Alvarezsalazar, O., Hatch, S., Rocca, J., et al.: Mission design for NISAR repeat-pass Interferometric SAR. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 9241, pp. 92410C–92410C-10 (2014)
7. Zhang, F., Hu, C., Li, W.: A deep collaborative computing based sar raw data simulation on multiple CPU/GPU platform. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **10**(2), 387–399 (2017)
8. Zhang, F., Hu, C., Li, W.: Accelerating time-domain SAR raw data simulation for large areas using multi-GPUs. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7**(9), 3956–3966 (2014)
9. Otten, M., Vlothuizen, W., Spreeuw, H.: Real-time processing of multi-channel SAR data with GPUs Radar Conference. IEEE (2017)
10. Yao, X., Hu, C., Zhang, F.: Atomic-free optimization on GPU based SAR raw data simulation. In: IGARSS 2016, 2016 IEEE International Geoscience and Remote Sensing Symposium, pp. 645–648. IEEE (2016)
11. Que, R., Ponce, O., Baumgartner, S.V.: Multi-mode real-time SAR on-board processing. In: Eusar (2016)
12. Ammar, M.A., Hassan, H.A., Abdel-Latif, M.S.: Performance evaluation of SAR in presence of multiplicative noise jamming. In: National Radio Science Conference (2017)

Ship Detection in Optical Satellite Images Based on Sparse Representation

Haotian Zhou^{1,2}, Yin Zhuang^{1,2}, Liang Chen^{1,2(✉)}, and Hao Shi³

¹ Radar Research Lab, School of Information and Electronics,
Beijing Institute of Technology, Beijing 100081, China
chenl@bit.edu.cn

² Beijing Key Laboratory of Embedded Real-Time Information Processing Technology,
Beijing 100081, China

³ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Abstract. Ship detection in remote sensing imagery has been widely applied in military and citizen applications, such as fishery management, vessel surveillance or marine safety and security. With the development of optical satellite, optical satellite imagery ship detection has caused a lot of attention. In this paper, we propose an offshore ship detection method based on sparse representation. First we employ histogram of oriented gradient (HOG) as the feature descriptor, then the HOG feature are extracted from training dataset. After feature extraction, all of samples are used to adaptively train a dictionary. Next, we encode HOG feature description of patches from test image by the dictionary. Finally, the sparse code and support vector machine (SVM) classification are employed in ship target validation and false alarms elimination. Experiments have shown better detection performance and stronger robustness of our method compared with other methods.

Keywords: Remote sensing · Ship detection · Sparse representation

1 Introduction

Ship detection, which is widely used in military and citizen applications such as fishery management, vessel surveillance or marine safety and security, has always been a popular research area of remote sensing image processing. In last few decades, ship detection mainly focused on synthetic aperture radar (SAR) images due to its capacity of all-weather and all-time. However, speckled noise interference and low resolution restrict applications of SAR imagery. With the development of optical satellite imaging technology in recent years, optical satellite imagery has been gradually attracted extensive attention because it has more visible and detail characters than SAR imagery.

In this paper, we focus on offshore ship detection since land areas can be removed by sea-land segmentation or geographic information. At the early stage, image segmentation, visual saliency and shape analysis methods are employed in optical ship detection methods. Zhu et al. proposed a ship detection method using image segmentation and shape analysis to extract candidates [1]. Bi et al. proposed a saliency detection based

hierarchical ship detection method [2]. Shi et al. proposed a ship detection method based on anomaly detection descriptor and local shape feature [3]. With the development of machine learning techniques, many methods apply machine learning algorithm in ship detection, and achieve good results. Feature extraction algorithms such as local binary pattern (LBP) or histogram of oriented gradient (HOG), and classification algorithms such as support vector machine (SVM) or Adaboost have ever been applied in ship detection. Qi et al. proposed a detection method using descriptor named S-HOG [4]. Shi et al. achieve the goal of verification using HOG feature and Adaboost classification [3]. The algorithms such as convolutional neural network (CNN) and sparse representation, which can automatically extract abstract semantic features, have been gradually applied in remote sensing target detection. Zou et al. proposed an automatic feature extraction method based on SVDNet to detect ship targets [5]. Zhang et al. proposed a S-CNN ship detection method based on CNN model [6]. Yokoya et al. proposed a ship detection method which combines generalized Hough voting and sparse representation [7]. Regards to these contributions, an offshore ship detection method based on sparse representation has been proposed in this paper. In our method, we adaptively train a dictionary based on HOG feature extraction, and use the dictionary to encode patches from test image. Then binary classification based on SVM is employed to classify the sparse code and judge whether a patch contains a ship target or not. Experiments have shown better detection performance and stronger robustness. The remainder of this paper is organized as follows. The proposed method is described in Sect. 2. Experiment results and analysis are presented in Sect. 3. The conclusions are presented in Sect. 4 and acknowledgements are illustrated in this chapter.

2 Proposed Method

In this paper, an offshore ship detection method based on sparse representation has been proposed, which mainly includes adaptive dictionary learning, sparse coding and binary classification. The flow of the method is in Fig. 1.

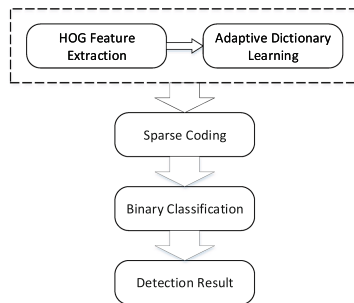


Fig. 1. The ship detection method work flow

The Fig. 1. shows the work flow of the ship detection method. The individual steps of the proposed method are briefly presented in the following discussions.

2.1 Dictionary Learning and Sparse Coding

The goal of our task is to detect ship targets in a given image. In our method, sparse representation has been used for this purpose. Patches from a given image can be represented as a linear combination of atoms from a trained dictionary, and can be modeled as Eq. (1):

$$Y = DX \quad (1)$$

where Y is the input signal dataset $\{y_i\}_{i=1}^m$, D is the previously trained dictionary in which each column represents an atom, and X means sparse coefficient vectors. To find the approximate solutions of D and x , dictionary learning can be represented as Eq. (2):

$$\min_{D, x_i} \sum_{i=1}^m \|y_i - Dx_i\|_2^2 + \lambda \sum_{i=1}^m \|x_i\|_1 \quad (2)$$

The first item of (2) aims to get the most appropriate D and x_i , which limits by the least reconstruction error. The second item of (2) aims to ensure sparsity of the coefficients.

To obtain a satisfactory solution, we use an alternate optimization method to solve this problem. In the first step, we assume D is fixed and solve x_i . Orthogonal matching pursuit (OMP) method [8] is employed in this step, which can get a sparse coefficient vector x_i for the input signal y_i . In the second step, the dictionary D is updated by the coefficient vectors changing caused by the underdetermine system, which can be modeled as:

$$\min_D \|Y - DX\|_F^2 \quad (3)$$

To solve this problem, K-singular value decomposition (K-SVD) method [9] is employed to update the dictionary. Equation (3) can be rewritten as:

$$\begin{aligned} \min_D \|Y - DX\|_F^2 &= \min_{d_i} \|Y - \sum_{j=1}^m d_j x_j\|_F^2 \\ &= \min_{d_i} \|(Y - \sum_{j \neq i} d_j x_j) - d_i x_i\|_F^2 \\ &= \min_{d_i} \|E_i - d_i x_i\|_F^2 \end{aligned} \quad (4)$$

When the i -th column of the dictionary is updated, we assume the rest columns are fixed, then E_i is fixed. To minimize formula (4), singular value decomposition (SVD) is used to get the solution d_i . After the iterative process mentioned above, the dictionary update is completed.

2.2 Adaptive HOG-Dictionary Construction

In our proposed method, we need to construct an over-completed dictionary to represent patches from a given image. For the sparse dictionary construction, HOG feature description method [10] is employed. The description capability of HOG feature has been proved in computer vision application field. In the proposed method, HOG features is employed to be the atoms of the dictionary. During HOG feature extraction, image is divided into small cells, then gradient histogram of each cell is extracted, and all histograms are cascaded. In this paper, we extract HOG feature of each sample from the training dataset (Fig. 2).

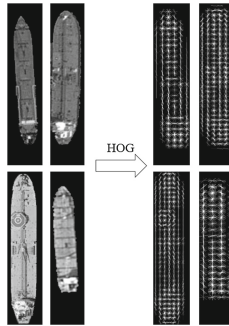


Fig. 2. HOG feature extraction

After feature extraction, the dictionary is initialized with two parts. The former half of the atoms are HOG features of positive samples which represent ship targets. And the latter half of atoms are from negative samples which represent backgrounds without ship targets. In this way, when a positive sample is used for dictionary update, the non-zero items of the sparse code are centralized in former half of the dictionary. Then the dictionary will adaptively update the former atoms of the dictionary and less update the latter atoms. Similarly, when a background sample is used for update, the former atoms are less updated. That’s why we call it adaptively dictionary learning. Due to this training strategy, the coefficient vectors encoded by the dictionary can be more discriminative, by which we can separate ship target and background easily (Fig. 3).

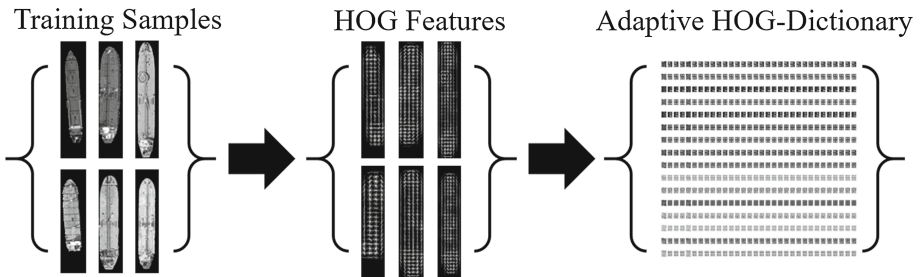


Fig. 3. Dictionary Learning

2.3 Ship Detection Based on Sparse Representation

After adaptive dictionary learning, the dictionary is employed to detect ship targets in test images. First, we use sliding-window method to get patches from the test images. When patches are obtained, the HOG feature of each patch is extracted and sparse code is calculated [11]. As we mentioned above, because the former half of the dictionary corresponds to positive atoms that represent ship targets, and latter half corresponds to negative atoms that represent backgrounds, when the encoded patch contains a ship target, most of the non-zero items of the code corresponds to positive atoms. And when the patch doesn't contain ship targets, there has little non-zero items that corresponds to positive atoms. Therefore, these sparse codes are more discriminative between patches that contain a ship target or not. Figure 4 shows the difference between positive and negative sample patches. Horizontal axis in Fig. 4. represents atoms in the dictionary and vertical axis in Fig. 4. represents the sparse codes of patch. After sparse code of each patch is obtained, the task of ship detection has become a binary classification problem. To solve this problem, SVM is employed to achieve binary classification. A linear SVM classifier is trained by the codes of positive and negative training samples. Then, this classifier is employed to classify sparse codes of test samples. In this way, offshore ship detection is achieved.

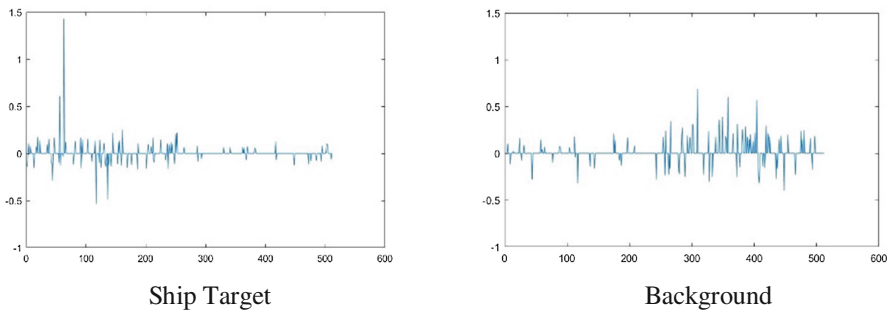


Fig. 4. Sparse codes of ship target and background

3 Experiments

In experiment, to validate the effectiveness of the proposed method, optical remote sensing images from Google Earth were employed to demonstrate the proposed method. The spatial resolution of the image is 1 m. The proposed method was implemented in Matlab R2016a, Windows 10. Configuration of the computer is Intel® Core(TM) i7-6500U CPU @ 2.50 GHz, 4.00 GB RAM.

The experiment is divided into the following steps. First, each training sample is transformed into gray-scale, and the size of each sample is adjusted into 160*160. Then HOG feature of each sample is extracted. During HOG extraction, the size of cell is 8*8, and block is overlapped. During dictionary learning, we first use 256 positive samples and 256 negative samples to initialize the former and the latter part of the dictionary

respectively. Then training dataset was employed to update atoms of the dictionary. The update method is K-SVD, iteration cycle is 1000 times, and learning rate set as 0.15. when dictionary learning is completed, we use the dictionary to encode the positive and negative samples respectively. The sparse codes are used to train SVM classifier, and the kernel function of SVM is linear. During ship detection, first we use multi-scale sliding window method to obtain patches from the test image. Then the size of patches is adjusted to 160*160 in HOG feature extraction. Then the dictionary is employed to encode each patch. After that, the sparse codes are classified by SVM classifier. If the result is positive, the patch is labeled in the test image. Figure 5. has shown the detection results.

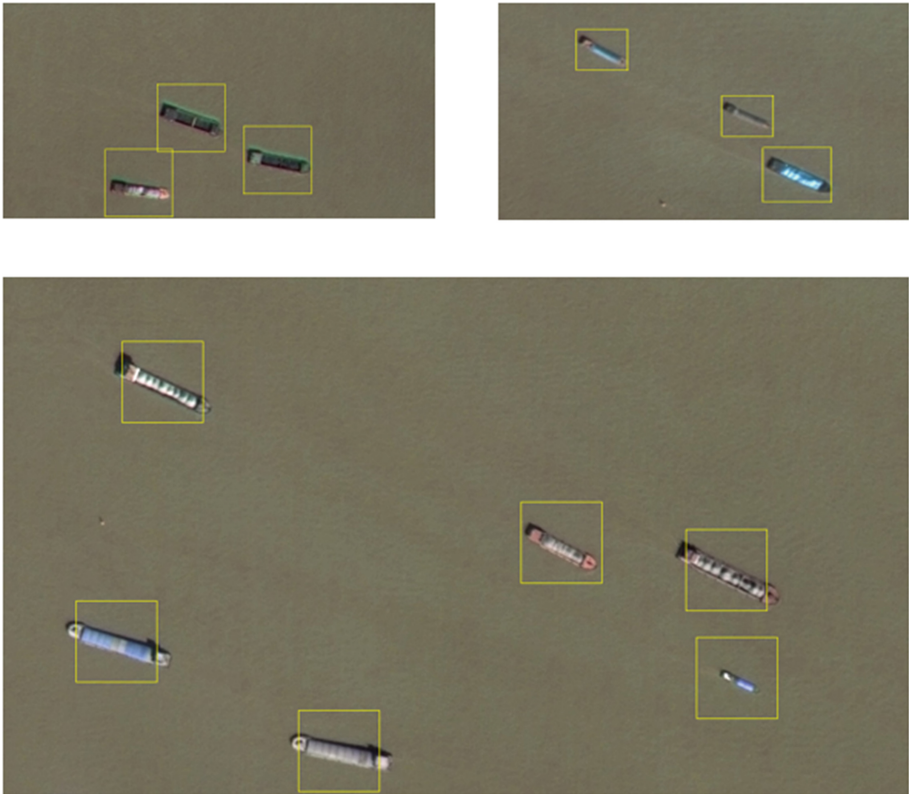


Fig. 5. Ship detection results

The experiment also compared the proposed method with other methods, such as S-HOG [4], SVDNet [5] and Bayesian based method [12]. To evaluate the effectiveness of these methods, precision and recall are employed as performance metrics, defined as

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (5)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (6)$$

The detailed results are shown in Table 1. Table 1 demonstrates the better performance of the proposed method compared with other methods. This is because the S-HOG method use artificial feature to detect targets, which is sensitive with interfering orientation information in background. The SVDNet method is based on simplified deep learning model, which needs massive training data to train a robust model. In the proposed method, although atoms of the dictionary are initialized by artificial HOG features, it become abstract semantic features after dictionary update. Therefore, the feature description strength of the dictionary is stronger than artificial features. In addition, during model training, the proposed method takes background information into account, and this point is ignored by other methods. Hence, the proposed method has good performance.

Table 1. Ship detection results of different methods

Method	Precision	Recall
Proposed method	94.2%	93.1%
S-HOG	93.4%	92.8%
SVDNet	92.0%	87.0%
Bayesian based method	69.2%	82.1%

4 Conclusion

In this paper, an optical imagery ship detection method based on sparse representation was proposed, which mainly includes adaptive dictionary learning, sparse coding and binary classification. Dictionary learning extracts advanced semantic features from the samples, and overcomes the disadvantages of artificial features. Experiment results has shown good performance of our method. In the future, we will take the inshore ship detection into account, and try to solve the inshore ship detection problem.

Acknowledgments. This work was supported in part by the Chang Jiang Scholars Program under Grant T2012122, in part by the Hundred Leading Talent Project of Beijing Science and Technology under Grant Z141101001514005, and in part by the National Natural Science Foundation of China under Grant 91438203.

References

1. Zhu, C., Zhou, H., Wang, R., et al.: A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geosci. Remote Sens.* **48**(9), 3446–3456 (2010)
2. Bi, F., Zhu, B., Gao, L., et al.: A visual search inspired computational model for ship detection in optical satellite images. *IEEE Geosci. Remote Sens. Lett.* **9**(4), 749–753 (2012)

3. Shi, Z., Yu, X., Jiang, Z., et al.: Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. *IEEE Trans. Geosci. Remote Sens.* **52**(8), 4511–4523 (2014)
4. Qi, S., Ma, J., Lin, J., et al.: Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images. *IEEE Geosci. Remote Sens. Lett.* **12**(7), 1451–1455 (2015)
5. Zou, Z., Shi, Z.: Ship detection in spaceborne optical image with SVD networks. *IEEE Trans. Geosci. Remote Sens.* **54**(10), 5832–5845 (2016)
6. Zhang, R., Yao, J., Zhang, K., et al.: S-CNN ship detection from high-resolution remote sensing images. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B7, pp. 423–430 (2016)
7. Yokoya, N., Iwasaki, A.: Object localization based on sparse representation for remote sensing imagery. In: *Geoscience and Remote Sensing Symposium 2014*, pp. 2293–2296. IEEE (2014)
8. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 40–44. IEEE (2002)
9. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006)
10. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 886–893. IEEE Computer Society (2005)
11. Mairal, J., Bach, F., Ponce, J., et al.: Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res. arch* **11**(1), 19–60 (2010)
12. Yang, G., Li, B., Ji, S.: Ship detection from optical satellite images based on sea surface analysis. *IEEE Geosci. Remote Sens. Lett.* **11**(3), 641–645 (2014)

Application of Back Propagation Neural Network with Simulated Annealing Algorithm in Network Intrusion Detection Systems

Chen Chang¹(✉), Xuebin Sun¹, Dianjun Chen¹, and Chenwei Wang²

¹ Beijing University of Posts and Telecommunications, Beijing, China
changchenwork@126.com, {sunxuebin, djchen}@bupt.edu.cn

² DOCOMO Innovations, Inc., Palo Alto, CA, USA
chen.516@gmail.com

Abstract. In this paper, we apply the back propagation neural network (BPNN) into the network intrusion detection system (NIDS). To overcome the training speed and local optimality, we propose a new algorithm of simulated annealing back propagation (SABP), incorporating BPNN with simulated annealing algorithm (SAA). The simulations results show that our proposed SABP outperforms the original BPNN in terms of the training speed.

Keywords: Network intrusion detection systems (NIDS)
BP neural network (BPNN) · Simulated annealing algorithm (SAA)

1 Introduction

The traditional firewall keeps the network security by identifying the IP, MAC and protocols included in the communication packages, which means it is difficult for the traditional firewall to defend the attacks such as IP spoofing and DNS spoofing. To solve this problem, NIDS can be used [1]. NIDS relies on classifying detection, mapping the input data to the output. The classification result is always non-linear mapping. For example, it has been demonstrated that the 3-layer BPNN where the sigmoid function was taken as the activation function can approximate any nonlinear continuous function [2]. Thus, in this paper we choose BPNN as the algorithm of NIDS.

Since BPNN is essentially a gradient descending method, the weights of the neurons are readily updated, which makes the cost function to easily fall into the local minimum and lead to failure trainings [3]. On the other hand, when the neuron output is close to zero, the weight error is also very small so that the training speed is significantly slow. To tackle with this problem, we resort to SAA to optimize the BPNN. Compared to the climbing algorithm, SAA introduces a random factor in the search process. In particular, it has a certain probability to jump out of the local minimum [4], so that it is able to find out the globally optimal solution. Also, the random factors could be used to speed the training process.

2 Back Propagation Neural Network

BPNN consists of an input layer, hidden layers and an output layer [5]. While the number of neurons in the input layer and output layer is determined by the training data set, the number of hidden layers is usually no more than 3 (contrary to deep neural networks), and the number of hidden neurons can be half of the sum of the number of neurons in the input layer and output layer.

The training process of BPNN can be divided into two parts: the forward propagation process and the error back propagation process. During the former one, the output of each neuron is based on the output values and corresponding weights of all the neurons in the previous layer, the bias factor, and the activation function. The output of the i -th neuron is given by the following equations:

$$S_j = \sum_{i=0}^{m-1} w_{ij}x_i + b_j, \quad (1)$$

$$x_i = f(S_i), \quad (2)$$

Where we usually choose $f(x)$ to be the sigmoid function, b_j is the bias factor, and m is the number of neurons in the previous layer, each with the input x_i and the corresponding weight w_{ij} . During the back propagation process, based on the forward propagation, we compare the result of the calculation with the expected result, and then inverses the error, and correct the weights and bias.

The difference between the actual output of the j -th neuron in l -th layer and the expected output is represented by δ_j^l , and the process of the back propagation can be described as follows:

1. Do the forward propagation operation and get the activation value of each layer.
2. For the output layer (set to n_1 layer), calculate the δ_j^l of the neurons in output layer:

$$\delta_j^l = a_j^l - y_i \quad (3)$$

3. Calculate the δ_j^l of the neurons in hidden layers:

$$\delta_i^l = \left(\sum_{j=1}^l w_{ji}^l * \delta_j^{l+1} \right) * [a_i^l * (1 - a_i^l)] \quad (4)$$

4. Calculate the final partial derivative value:

$$\nabla_{w^l} J(W, b; x, y) = \delta^{l+1} (a^l)^T \quad (5)$$

$$\nabla_{b^l} J(W, b; x, y) = \delta^{l+1} \quad (6)$$

5. Initialize the all-zero matrix, all-zero vector, and assign them after calculating:

$$\Delta w^l = \nabla_{w^l} J(W, b; x, y) \quad (7)$$

$$\Delta b^l = \nabla_{b^l} J(W, b; x, y) \quad (8)$$

6. Update weight and bias:

$$w^l = w^l - \alpha \left[\left(\frac{1}{m} * \Delta w^l \right) + \lambda * w^l \right] \quad (9)$$

$$b^l = b^l - \alpha \left(\frac{1}{m} * \Delta b^l \right) \quad (10)$$

Where α is the learning rate.

3 Simulated Annealing Algorithm

The simulated annealing algorithm (ANN) is a type of stochastic optimization algorithms based on the Monte-Carlo iterative solution strategy [6]. By simulating the annealing process in metallurgy, it performs well in finding out the optimal solution in a larger solution space [7]. In SAA, we need to choose appropriate variables as the energy signal. In each step of cooling, there is a probability to randomly generate a new network state. When the energy in the new state is higher than that in the old state, the new network is accepted at a certain probability. Note that it is necessary to restrict the way the network changes. Thus, we limit the way the network parameters changes when generating a new network state, and the “backtracking” method is also added to SAA, where “backtracking” means that the network will record the optimal solution within a certain number of training times and choose the better one for follow-up training after comparing the optimal solution to the current solution.

The process of simulated annealing algorithm can be divided into four steps:

1. Produce a new solution in the solution space.
2. Compute the network performance of new solutions.
3. Judge whether the new network is accepted or not, the judgment is based on an acceptance criterion. The most commonly accepted criterion is the Metropolis criterion: if the new solution is better, accept it, otherwise the probability function as followed will determine whether accept the new network [8].

$$P = e^{\frac{-\Delta f'}{T}} \quad (11)$$

4. When the new network is accepted, replace the network with the new one.

4 Optimization and Simulation

4.1 Network Optimization

During the training process of BPNN, the learning rate α affects the speed and precision of the training. If the learning rate is large, training speed will be accelerated, but more

curve information will be lost and final training accuracy will be hurt. In contrast, when the learning rate is too small, although the accuracy can be improved, the learning rate will be significantly reduced [9]. To ensure the training effect under the premise of improving the training speed and achieving the globally optimal solution, we will build a BPNN due to its stability, and optimize the network by using SAA, referred to as SABP.

The flow chart of SABP is shown in Fig. 1:

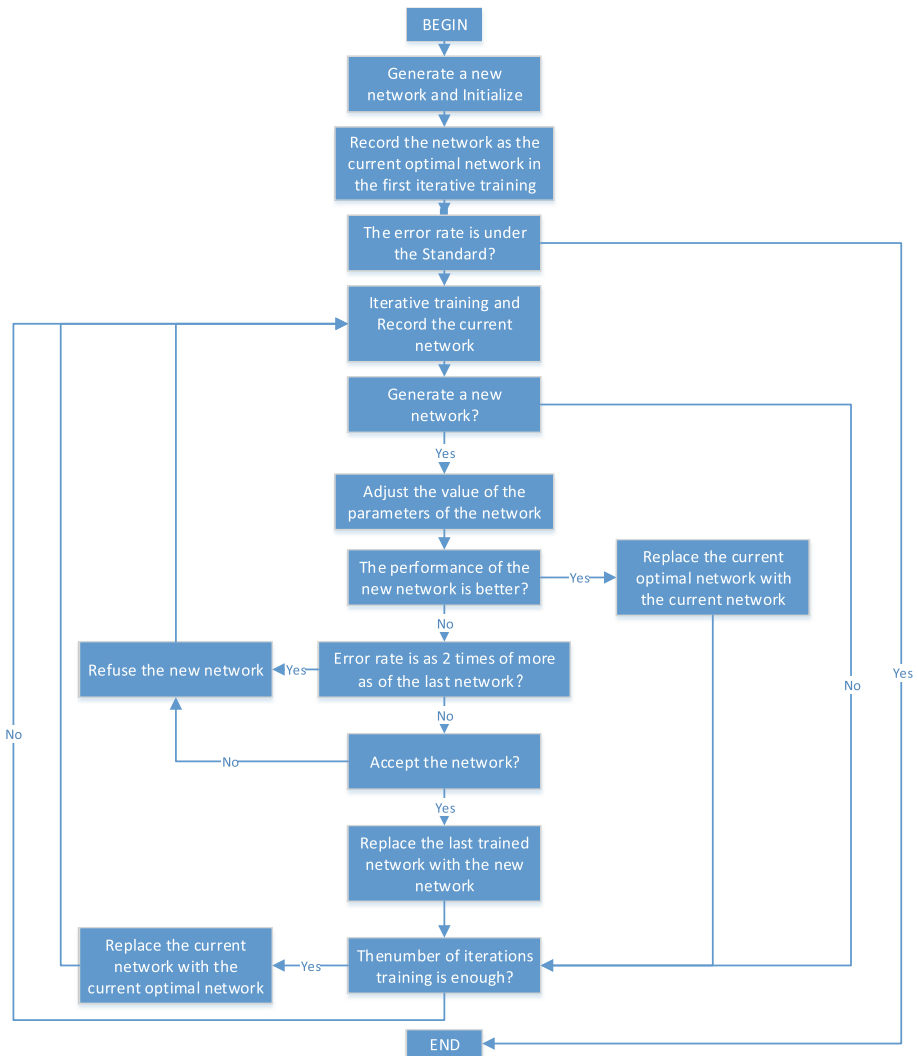


Fig. 1. Flow chart of the SABP

The optimized network logic flow is as followed:

1. Generate a new BPNN and set the parameters random values, and set the learning rate α and the expected training error rate fixed values. Determine the probability functions about generating the new network and about accepting the worse network.
2. Record the network after the first iterative training as the current optimal network.
3. Train the network, record and save the current network parameters before the start of each iteration training. In the process of iteration training, it is determined by the probability function whether to generate a new network and complete the iterative training or just to complete the iterative training.
4. If a new network is determined to be generated, the bias and weight parameters of the neurons in the network are adjusted.
5. After a single iteration training, the trained network performance is compared to the old one. If the trained network performance is improved, the new network is accepted. On the other hand, it is determined whether the new network is accepted or not.
6. Compare the performance of the network with the expected value, if the network performance is better than the expected value, end the training, if not, and carry on the next iterative training.
7. After completing an iterative training, compare the current network with the current optimal network, if the current network performance is better, then replace the current optimal network with the current network, if not, carry on the next iterative training. After comparing the current network performance with the current optimal network, if the current network performance is poor, the iterative training is carried out on the basis of the current optimal network, if not, replace the current optimal network with by the current network.

4.2 Generation of New Network and Network Backtracking

In the SAA, the error rate of each training is the energy state of the network. With reference to the Metropolis criterion, the probability function that produces the new solution shown in [10] is repeated as follows:

$$P = e^{-\frac{\sqrt{E_1 - E_2}}{0.95 * E_1}} \quad (12)$$

When $E_2 < 1 - P$, the network parameters are adjusted to produce a new network on the basis of the original network. The receptive probability function is as follow:

$$P = e^{-\frac{\sqrt{E_1 - E_2}}{0.85 * E_1}} \quad (13)$$

When $E_2 > 2 * P$, the new network is refused. Note that E_1 is the last iteration training error rate, E_2 is a random number between 0 and 1. The numbers 0.95 and 0.85 are constant determined by repeatedly training tests. Besides the functions, we also take

another restriction that if the error rate is twice as much as the error of the last training network, then the worse network is refused.

In our algorithm, the SAA produces a new network where the parameters of the neurons in the hidden layers are updated. The parameters include error, bias value, bias accumulation, weight and weight accumulation. Note that the updated parameters mainly include the bias value and weight, and the other parameters remain the same as in the original network. At first, we randomly classify the neurons in the hidden layers at a ratio of 5:3:2, and change the 50% weights linearly, and then multiple the weight of the 30% neurons by 0.1 to the remaining 20% neurons weight, in the course of the experiment we gradually adjust the proportion of neurons classification, the proportion of linear changes, and the proportion of superimposed. Finally when the network is to be updated, the weights of randomly selected 70% of the neurons in each hidden layer become 1.45 times of their original values, and for the remaining 30% of the neuron, 0.4 of their original values are added to their weights. In addition, the bias of each neuron in hidden layers is set to a random value.

Note that backtracking affects the effect of SAA and the training speed. After repeatedly training and testing, we decide to do backtracking every 250 iterations.

4.3 Simulation and Testing

Regarding the training samples and the testing samples, we resort to the *kddcup.data.corrected* dataset, each data with 41 eigenvalues as the input and one value as the output. Hence, we have 41 input neurons, 1 output neuron and we choose 2 hidden layers, each hidden layer with 21 hidden neurons after several screening tests. The expected error rate is set to 0.03, and the learning rate is chosen to be 0.005, after several screening tests.

Training samples and test samples are taken from the tag data set *kddcup.data.corrected*, we randomly select 10500 samples from the data set, 500 samples from the extracted data for training, and the remaining 10,000 for testing.

5 Experimental Results

In this section, we consider both BPNN and SABP algorithms for performance analysis. In order to reduce the effect of system error in the test, 500 training and forecasting are carried out for each one, and we keep the time of each training and prediction error.

In Table 1, we compare the error rate and training time of the two algorithms. As Figs. 2 and 3 illustrate, the error rate of SABP is 5.58%, and the training time average is 14.31 s. In contrast, the error rate of BPNN is 10.33%, and the training time average is 60.24 s. It can be seen that SABP outperforms BPNN in terms of the training speed.

Table 1. The result of the experiment

	Training time		Error rate	
	Average	Variance	Average	Variance
SABP	14.31 s	1143.49	5.58%	11.89
BPNN	60.25 s	3568.44	10.33%	8.22

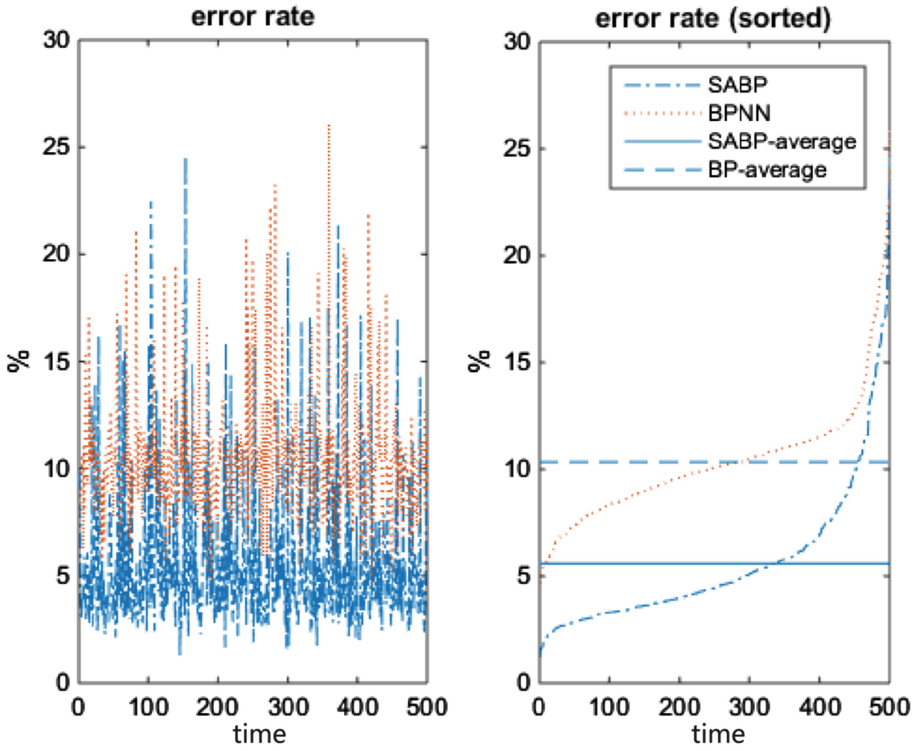


Fig. 2. Error rate of two groups

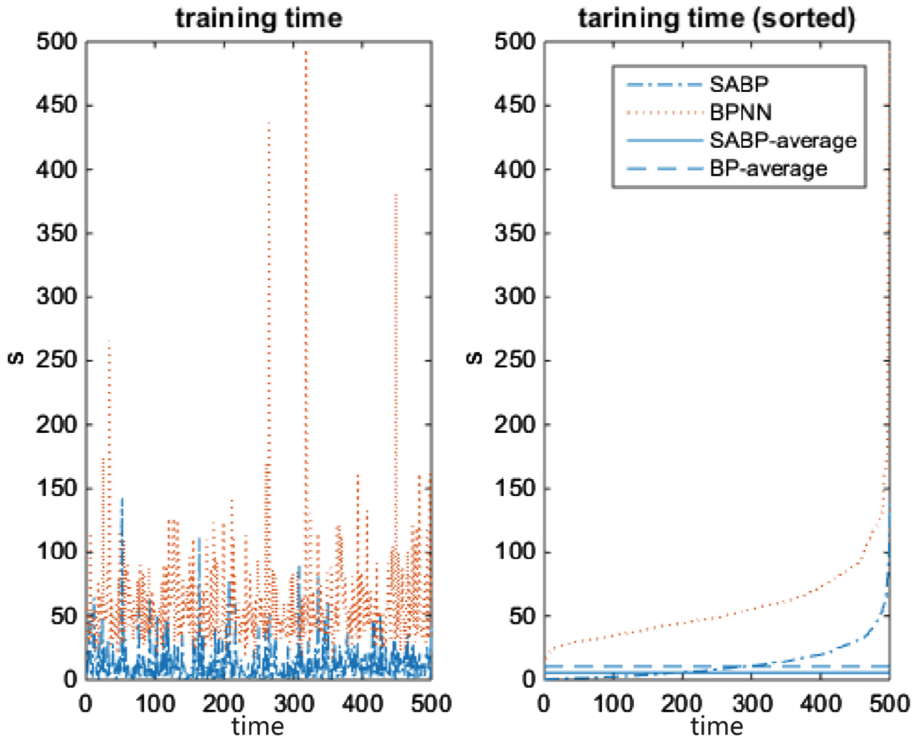


Fig. 3. Training time of two groups

6 Conclusion

In order to defend the rapidly changing network attack, NIDS needs to produce mistake or fail reports as few as possible, and learn and identify new attack type as soon as possible, which requires higher accuracy and faster training speed. Our result shows that our proposed SABP, incorporating BPNN with SAA, is an effective method for the neuron network to jump out of the local minimum and reduce the error rate, while the training speed of the network could also be significantly improved, and it outperforms BPNN. In the future, we would like to apply this new method in many other applications and evaluate its performance.

References

1. Huang, X., Zhang, M., Wang, X.: Protection system for network boundary security. In: 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 6. IEEE (2010)
2. Zhao, B., et al.: Convolutional neural networks for time series classification. *J. Syst. Eng. Electron.* **28**(1), 162–169 (2017)

3. Yao, X.H.: A network intrusion detection approach combined with genetic algorithm and back propagation neural network. In: 2010 International Conference on E-Health Networking, Digital Ecosystems and Technologies (EDT), vol. 1. IEEE (2010)
4. Qian, F., Ding, R.: Simulated annealing for the 0/1 multidimensional knapsack problem. *Num. Math.-Engl. Ser.* **16**(4), 320 (2007)
5. Wang, Y., Lu, C., Zuo, C.: Coal mine safety production forewarning based on improved BP neural network. *Int. J. Min. Sci. Technol.* **25**(2), 319–324 (2015)
6. Qi, J.-Y.: Application of improved simulated annealing algorithm in facility layout design. In: 2010 29th Chinese Control Conference (CCC). IEEE (2010)
7. Hongkai, W., Yanyong, G., Peijun, X.: Simulated annealing algorithm for the optimal translation sequence of the jth agent in rough communication. *J. Syst. Eng. Electron.* **19**(3), 507–512 (2008)
8. Lu, X.-Z., et al.: Structural optimization and segregation behavior of quaternary alloy nanoparticles based on simulated annealing algorithm. *Chin. Phys. B* **25**(5), 053601 (2016)
9. Sun, Y.-J., et al.: Improved BP neural network for transformer fault diagnosis. *J. Chin. Univ. Min. Technol.* **17**(1), 138–142 (2007)
10. Liu, L., Feng, G.: Simulated annealing based polynomial time QoS routing algorithm for MANETs. *J. Electron. (Chin.)* **23**(5), 691–697 (2006)

An Improved Binary Bee Colony Algorithm for Satellite Resource Scheduling Method

Pan Zhao^(✉), Xuebin Sun, and Ping Chen

Beijing University of Posts and Telecommunications (BUPT),
No. 10 Xitucheng Road, Haidian District, Beijing, China
hbupan@outlook.com

Abstract. Aiming at the problem of satellite resource scheduling for multi-space targets, drawn on the experience of encoding in the Particle Swarm Optimization (PSO) algorithm, we designed an encoding style to represent the constraint and the solutions to the problem and introduced binary artificial bee colony (BABC) algorithm based on Pareto multi-objective optimization. Compared with the artificial bee colony (ABC) algorithm, the only difference is that BABC used Logistics function mapping the values to the binary. In this paper we made some improvements including population initialization which use the constraint conditions to randomly generate then modify to a feasible solution and candidate solutions generation in a way of crossover used in the Genetic algorithm. In the optimal solution search process, the Pareto optimal solution of the population is recorded, which means a set of differentiated solutions with different advantages on different indexes is obtained. It is convenient to select the corresponding optimal solution according to the user's preference and the actual situation. The experimental results show that the improved binary artificial bee colony algorithm could solve the satellite resource scheduling problem, which provides a new idea for multi-space target satellite resource scheduling problem.

Keywords: Multi-objective optimization · BABC algorithm
Satellite resource scheduling

1 Introduction

In recent years, observational satellite technology has been developing continuously and has shown its important role in related fields such as geological exploration, environmental monitoring and spatial target monitoring [1]. The space target has the advantages of low running track, fast running speed and unpredictable flight path. At the same time, the observation of the space target needs the cooperation of multiple observation satellites, so the efficient dispatching of the observation satellite resources becomes an important research problem [2]. The problem of satellite resource scheduling can be attributed to the optimization problem of multi-optimization and multi-objective. The time complexity of this problem increases rapidly with the increase of spatial target, which makes it difficult to solve the problem by using traditional enumeration method. Artificial Bee Colony Algorithm [3] was proposed by D. Karaboga in 2005, which is a novel group based intelligent heuristic algorithm

designed to solve the problem of multivariable function optimization. In 2006, Chin Soon Chong introduced the colony algorithm into the field of combinatorial optimization problem, solved the problem of shop scheduling [4], and then successfully solved the traveling salesman problem [5]. Because the artificial bee colony algorithm has the characteristics of few parameters, simple and convenient, robust and strong, it has obvious advantages in the complex combinatorial optimization, and has rapidly become the research hotspot of the combinatorial optimization problem, widely used in the aspects of path planning, image processing and power Task scheduling [6–8]. In this paper, we introduce the bee colony algorithm into the above problem, and use the colony algorithm to solve the multi-constrained satellite resource scheduling problem.

2 Problem Description

The problem based on the following assumption that each observing satellite carries a sensor to monitor the space target, the monitoring of each space target requires w satellites, regardless of the time overhead of the satellite switch, the status of the satellite to the space object is visible and invisible [9]. Taking into account the above assumptions, the satellite resource scheduling problem can be described as:

- (1) The whole system has m satellites represented by the set $S = \{s_1, s_2, \dots, s_m\}$
The position of satellites in a constellation is abstracted into one dimension, which makes neighboring satellites adjacent to each other;
- (2) There are n space object represented by the set $O = \{o_1, o_2, \dots, o_n\}$
- (3) The scheduling process is divided into k time periods, and the set representation is $T = \{t_1, t_2, \dots, t_k\}$
- (4) The scheduling policy is represented by the following set, $\Phi = \{\phi(i, j, k)\}$, $i = 1, \dots, m, j = 1, \dots, n, k = 1, \dots, K$, $\phi(i, j, k)$ means that the space target j in the time period k is observed by the satellite i .

In view of the above problems, there are a number of indicators can be optimized from a different point of view [10].

- (1) Target observable rate I_s , the target observable rate represents the ratio of the number of targets that can be observed in each time period to the total number, and then averages all time periods. The index is calculated as follows:

$$I_s = \frac{\sum_1^K I_k}{K}, I_k = \frac{q}{n} \quad (1)$$

In this Eq. (1), q is the number of target observed in time period k , while n is the total number of space target. For the target observation rate index, the higher the index, the more monitoring the target, the greater the index.

- (2) Switching frequency I_η , is the sum of the number of times a target satellite observation changes calculated by the equation:

$$I_\eta = \sum_1^K I_k \quad (2)$$

In Eq. (2), K is the total number of periods, and I_k is the handoff times of the target satellite pairs for each time period handover. As each target - satellite switch brings energy and time consumption, the smaller the target, the better.

- (3) Slackness degree I_μ , the index represented by the variance of the satellite operating time calculated by the equation:

$$I_\mu = [\sum_{i=1}^M (T_m - T_a)]^{1/2} \tag{3}$$

T_m is the working time of each satellite, T_a is the average time for all satellite. Because of the huge cost of replacement and maintenance of on-orbit satellites, the smaller the degree of slackness, the more stable the working time of the satellite constellation.

In the optimization of the above indexes, we also need to consider the actual work of the satellite state. The constraints of the problem:

- (1) Each target requires multiple satellites to be monitored at any time;
- (2) Each satellite can only monitor one target at the same time;
- (3) The satellite has different visibility into the target at different times;

In this paper, we introduce the matrix group coding method, encode each stage, and then synthesize the coding matrix of different stages Matrix group. Because the satellite’s working state on the ground includes the following two types: available and unavailable. Encoding for both states, using 0/1 can be fully expressed, using the following methods for coding: the constraint matrix is $n * km$ dimensions, where the quantum of space target represented by n , the quantum of satellite represented by m , the quantum of time period represented by k . The elements of the matrix represent the availability of satellites at some stage, as shown in following example:

$$mat = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & \vdots & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & \vdots & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & \vdots & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \tag{4}$$

In Eq. 4, the coding matrix represents the constraints of 3 detection targets, 5 satellites, and 2 time periods.

The solution matrix is encoded in a matrix of the same dimension as shown in Eq. 3, taking into constraints. Using matrix group encoding method can not only accurately represent constraints of practical problems, but also has good adaptability, that can flexibly modify the resource allocation problem with different scenarios.

3 Binary Artificial Bee Colony Algorithm

Artificial Bee Colony Optimization Algorithm is inspired by the behavior of bees foraging. A complete bee group consists mainly of three different bees, which is so called employed, onlookers and scouts. Employed is the bees who carries a food

resource and will share the food information including distance, amount and quality to the onlooker bees by different dances when back to hive. After getting all information from different employed bees, the onlooker bees will choose the best food resource as foraging target. Scouts are the bees who randomly seek the new food resource and will become employed once find food resource and then acts as an employed bee. The algorithm corresponds above process. The solutions about the problem can be treated as the position of food resource mentioned above, and the fitness of the solution can be measured by the amount of nectar.

ABC algorithm implementation steps are as follows: First, the artificial bee colony algorithm randomly produces SN food sources by the equation:

$$x_i = x_l + rand * (x_u - x_l) \quad (5)$$

In the Eq. (5), x_l is the lower boundary of the feasible solution, x_u is the upper bound of the feasible solution, $rand$ is the uniform distributed random number on (0, 1). Then the colony searches MCN times for all food sources. In each cycle, the bee first generates a candidate food source in the neighborhood of the food source with the equation:

$$v_{id} = x_{id} + rand * (x_{id} - x_{jd}) \quad (6)$$

In the Eq. (6), the subscript i represents the source of food to be evolved, $j \in \{1, 2, \dots, SN\}$, $d \in \{1, 2, \dots, D\}$, j, d is chosen randomly, $rand$ is the uniform distributed random number on (-1, 1). And then compared the new generation solutions to the original, if the fitness value of the new solution better than the old one, the bee abandon the old solution and memorizes the new one. Otherwise it just do nothing just keep the old in memory.

According to the information that employed brought back, the onlooker bees select their favorite food source in a way of roulette by probability proportional calculated from the fitness of the food resource. Once a solution is chosen, the onlooker bee use Eq. (5) to produce a new solution around the original solution based on the old one and the selected one. If the fitness of the new solution is better than the old one, the bee also abandon the old one and select the new. The probability value, p_i by which an onlooker bee chooses a food source is calculated as follows:

$$p_i = \frac{fit_i}{\sum_{n=1}^{sn} fit_n} \quad (7)$$

In Eq. (7), fit_i fitness value corresponding to the solution, the relationship between the fitness value and the objective function has a corresponding rule as follows:

$$fit_i = \begin{cases} \frac{1}{1+|f_i|}, & f_i \geq 0 \\ 1 + |f_i|, & f_i < 0 \end{cases} \quad (8)$$

When a solution is passed through the *limits* without updating, a new solution is randomly generated to replace the original solution by the Eq. (4).

Base on ABC, Marinakis et al. Modeled the Binary Artificial Bee Colony (BABC) algorithm [11] based on the idea of binary particle swarm optimization [12] and binary differential evolution algorithm [13]. The algorithm use logistic function mapping the continuous values obtained by the initialization process and the neighborhood search to binary. The mapping relationship is shown below.

$$bin(x_i) = \frac{1}{1 + e^{-x_i}} \tag{9}$$

$$x'_i = \begin{cases} 1, & bin(x_i) > rand \\ 0, & bin(x_i) \leq rand \end{cases} \tag{10}$$

where Eq. (9) is the logistic function. In Eq. (10), *rand* are uniformly distributed between (0, 1) random numbers. The rest of the algorithm is the same as the traditional binary algorithm.

4 The BABC Algorithm for Satellite Resource Scheduling

The BABC algorithm inherits the neighborhood search method of the traditional ABC algorithm, and uses the Logistic function to map the obtained value to the binary. The main problem [14] is that the initial source of food and the neighborhood search of the food source are Continuous mapping to discrete, which can't calculate the complexity of the higher complexity of the optimization problem; The second is that when employed bee and onlooker bee doing its neighborhood searching, solutions updates within single-dimensional, result in the new solutions are very likely the same with the old ones. This makes the convergence of the algorithm slow, easy to fall into the local optimal solution, affecting the performance of the algorithm.

In order to obtain a solution that has good performance on multiple indexes, the optimal solution is saved using the Pareto optimal solution set. To solve this problem, the BABC algorithm makes the following adjustments.

- (1) The initialization of solutions. We randomly generate a 0–1 matrix that meet the constraints.
- (2) Neighbor search method. For the solution *I*, select *d* randomly, $d \in \{1, 2, \dots, D\}$ and then randomly select a solution *k*, $k \in \{1, 2, \dots, K\}$, $k \neq i$, use the function below to generate a new solution.

$$x'_i = overlap(x_i, x_k, d) \tag{11}$$

Function *overlap()* achieve these operations including using the column *d* of solution x_k to replace the same column in solution x_i , and then fix the solution according to the constraint matrix to a feasible solution.

- (3) Onlooker phase. Different with ABC algorithm of one objective function, we have three indexes need to be optimized, so the probability calculation formula define as follow:

$$p_i = \frac{Domin(x_i)}{N} \quad (12)$$

Where the $Domin(x_i)$ is the number of Dominant solution by x_i , and N is total number of solutions.

5 Simulation Results and Analysis

In this section, the MATLAB programming language is used to implement the algorithm, and the effect of the bee colony algorithm on the final results under different control parameters is discussed. The constellation data part uses the Walker Constellation to observe the space target, the constellation parameter is 24/3/1, and each space target observation needs more than 3 satellites. The parameters of bee colony algorithm are population size 200 and maximum iteration number 2000.

This experiment mainly analyzes the observational index of the number of space objects in different constellation data (Fig. 1).

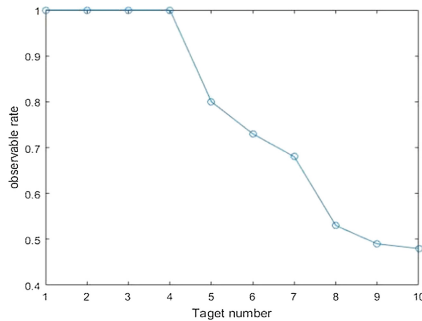


Fig. 1. The observation rate with different number of space target

As can be seen from the diagram, when the number of targets is not greater than 4, the observation rate is 1, which means that all the space targets can be effectively observed at each scheduling stage, and the observation rate drops obviously when the value is greater than 4. Because there are 24 satellites in the constellation, each target needs 3 to satellite observations, so the number of observations for the 8 largest, but because each satellite can only services one target, and the target of different satellite visible is different so it is difficult to reach the maximum number of observation in theory.

When the number of different targets in the region is different, the parameters of the switching times and the relaxation parameters are observed (Figs. 2, 3 and 4).

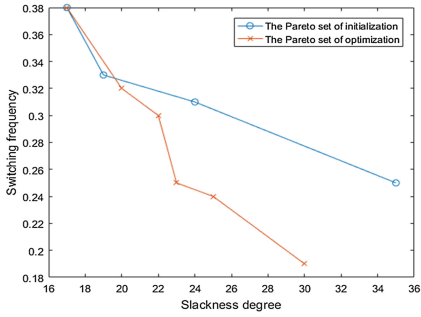


Fig. 2. The Pareto set of 3 space targets

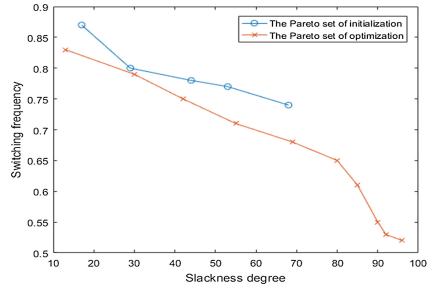


Fig. 3. The Pareto set of 5 space targets

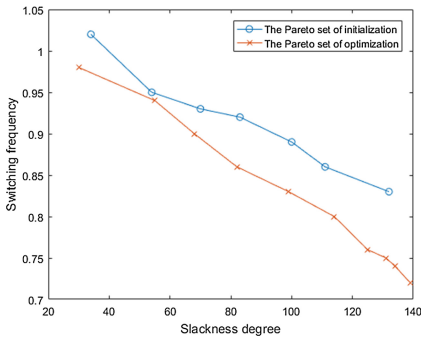


Fig. 4. The Pareto set of 10 space targets

It can be seen from the figure with the increase in the number of targets, switching times and relaxation index rose at the same time, switching frequency and relaxation index is a pair of contradictory indicators, a low degree of relaxation shows that all the satellite working time is balanced, but the equilibrium shows satellite observation on the change in different periods of time, even before the stage of the satellite target is still satisfy the constraints.

6 Conclusion

In this paper, binary bee colony algorithm is introduced to satellite resource allocation problem with multiple optimization objectives and multiple constraints. Meanwhile, the binary artificial bee colony algorithm is improved from continuous to discrete. One improvement is that when the initial solution is generated, because the upper and lower bounds of the 0–1 coding matrix are determined, it is not reasonable to use the traditional initial solution to generate the formula. For the 0–1 combinatorial programming problem, this paper directly uses the method of random generation combination and then the constraint correction to avoid the process from continuous

mapping to discrete process. Experimental results show that the proposed algorithm can solve the satellite resource allocation problem with multiple constraints and multiple objectives.

References

1. Ran, G.: Foreign civilian and commercial earth observation satellites. *Space Int.* **2016**, 41–48 (2016). 2016 year in review
2. Cheng, H., Wang, B., An, W.: A sensor scheduling method of LEO constellation based on information decision tree. *Acta Electronica Sinica* **38**(11), 2630–2634 (2010)
3. Karaboga, D., Akay, B.: A comparative study of Artificial Bee Colony algorithm. *Appl. Math. Comput.* **214**(1), 108–132 (2009)
4. Chong, C.S., Low, M.Y.H., Sivakumar, A.I., et al.: Using a Bee Colony algorithm for neighborhood search in job shop scheduling problems. In: *European Conference on Modelling & Simulation: Simulations in United Europe* (2007)
5. Wong, L.P., Low, M.Y.H., Chong, C.S.: Bee Colony Optimization with local search for traveling salesman problem. *Int. J. Artif. Intell. Tools* **19**(03), 305–334 (2010)
6. Xiong, W., Xu, B., Xu, M.: Differential Bee Colony algorithm for non-convex economic load dispatch. *Control Decis.* **26**(12), 1813–1817 (2011)
7. Xiao, Y., Yu, W.: Bee Colony algorithm for image edge detection. *Appl. Res. Comput.* **27**(7), 2748–2750 (2010)
8. Hu, Z., Zhao, M.: Research on robot path planning based on ABC algorithm. *Electr. Weld. Mach.* **39**(4), 93–96 (2009)
9. Zheng, Y., Yin, Y., Yong, D., et al.: Satellite resource scheduling algorithm based on Pareto front and particle swarm optimization. *Comput. Eng.* **42**(1), 193–198 (2016)
10. Xie, K., Han, Y., Xue, M., et al.: Algorithm for sensor management in the space-based infrared LEO constellation. *J. Astronaut.* **28**(5), 1331–1336 (2007)
11. Marinakis, Y., Marinaki, M., Matsatsinis, N.: A hybrid discrete Artificial Bee Colony - GRASP algorithm for clustering. In: *International Conference on Computers & Industrial Engineering*, pp. 548–553. IEEE (2009)
12. Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. In: *1997 IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation*, vol. 5, pp. 4104–4108. IEEE (2002)
13. Pampara, G., Engelbrecht, A.P., Franken, N.: Binary differential evolution. In: *IEEE Congress on Evolutionary Computation, CEC 2006*, pp. 1873–1879. IEEE (2006)
14. Liu, T., Zhang, L., Zou, K., et al.: Multiuser detection based on differential evolution binary Artificial Bee Colony algorithm. *Adv. New Renew. Energy* **18**(1), 5–10 (2013)

Automatic Liver Segmentation on CT Images

Torecan Celik¹, Hong Song^{1(✉)}, Lei Chen¹, and Jian Yang²

¹ School of Software, Beijing Institute of Technology, Beijing, China
anniesun@bit.edu.cn

² School of Optics and Electronics, Beijing Institute of Technology, Beijing, China
jyang@bit.edu.cn

Abstract. In this paper, a new coarse-to-fine framework is proposed for automatic liver segmentation on abdominal computed tomography (CT) images. The framework consists of two steps including rough segmentation and refined segmentation. The rough segmentation is implemented based on histogram thresholding and the largest connected component algorithm. Firstly, gray value range of the liver is obtained from image histogram, then the liver area is extracted from the rest of an image according to the largest connected component algorithm. The refined segmentation is performed based on the improved GrowCut (IGC) algorithm, which generates the label seeds automatically. The experimental results show that the proposed framework can efficiently segment the liver on CT images.

Keywords: Liver segmentation · CT images · Improved Grow-Cut
Histogram thresholding

1 Introduction

Automatic segmentation for liver computed tomography (CT) images is an important part in computer-aided liver diagnosis. CT has been widely used for clinical diagnosis of hepatic disease because of its high resolution. Accurate liver segmentation from abdominal CT scans is critical for computer-assisted diagnosis and therapy. Generally, radiologists or physicians have to manually delineate the liver region slice by slice, which is tedious and time-consuming due to the large amount of data. There are mainly two reasons to explain this phenomenon. The first one is the intensity of the liver is similar to other organs or muscles. The other one is different patients or even on the same patient may exist variation in both shape and scale [1–8].

Recently, a large variety of methods have been developed to improve the liver segmentation procedure. These methods are commonly based on region growing, clustering, classification algorithms, deformable models or level sets, statistical shape models, probabilistic atlases, and graph cuts.

[3] proposed a 3D statistical shape model to segment the liver from CT images. They firstly used a training set of shapes to build up a statistical model based on iterative technique. Each shape used in building model is defined by some anatomically specific points sampled on the liver surface. Then they computed the mean shape and positioning

it on the image. After that, they apply adjustment to every single shape. However, because the shape of the liver is highly variable, it is difficult to select the landmarks and the authors did not evaluate their model in clinical.

[2] proposed a region-growing approach. It can provide acceptable results on contrast enhanced CT images. The approach starts from a small region (environment of input curve, or point) and if the intensity of the neighboring voxel is corresponds to a pre-defined range, it will be added to the actual region. The region-growing method can close round the vessels and tumors efficiently (in contrast with active contour), but it's sensitive to select original input and if the intensity is similar, it is easily flow into the neighboring organs.

And [8] combined the methods of distance maps, thresholding and mathematical morphology to delineate the skin, bones, lungs, kidneys and spleen, then extracted the liver. They build a 3D reference model from manually segmented livers and adjusted into the image with rigid and affine registration. Then the model deformed to get the final result.

In this paper, an automatic liver segmentation framework without any interaction is proposed. Because different CT machines may generate different color ranges, a calibration step is needed for each different CT machine to perform automatic segmentation. The framework consists of two steps, rough segmentation based on automatic liver detection algorithm and refined segmentation based on IGC [9–11]. The rough segmentation is implemented based on histogram thresholding and the largest connected component algorithm. Firstly, gray value range of the liver is obtained from image histogram, then the liver area is extracted from the rest of an image according to the largest connected component algorithm. The refined segmentation is performed based on the improved GrowCut (IGC) algorithm, which generates the label seeds automatically. An automatic checking algorithm is proposed to choose the rough segmentation results which need to be done refined segmentation. The experimental results show that the proposed framework can segment the liver on the CT image series efficiently and automatically.

2 Methods

The proposed segmentation framework is shown in Fig. 1. It consists of the following main steps. Firstly, the input CT image sequences are roughly segmented using automatic liver detection algorithm, then IGC is applied to do the refined segmentation. Various post-processing steps, such as filtering, filling holes, erosion and masking are implemented on the refined segmentation liver region.

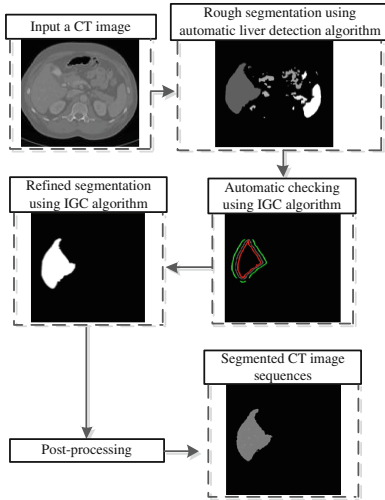


Fig. 1. The proposed automatic liver segmentation framework

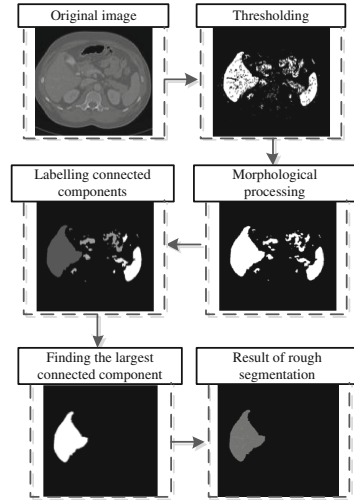


Fig. 2. The flowchart of automatic liver detection

2.1 Rough Segmentation Using Automatic Liver Detection Algorithm

Since the intensity of the liver is similar with the adjacent other organs, without preprocessing, a direct liver extraction process may unavoidable extract undesirable components from its adjacent organs as fault positive/negative errors [1]. To solve this problem, we use the intrinsic characteristics of the liver. Because the liver is between the stomach, intestines and thorax in the abdomen. It is the largest organ in human body. Using these characteristics, our goal is to find the largest connected component in CT images.

The automatic liver detection algorithm consists of four main steps, thresholding, morphological processing, labeling and finding the largest connected component. The flowchart of the automatic liver detection algorithm is shown in Fig. 2.

The original CT image, which is shown in Fig. 3(a) was first processed by the thresholding method, which can be described as follows,

$$g(x, y) = \begin{cases} 1 & \text{if } (f(x, y) > T_1 \text{ and } f(x, y) < T_2) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $g(x, y)$ is the image after thresholding operation, $f(x, y)$ is the original grayscale image, T_1 and T_2 are the intervals of the liver. The obtained threshold image is shown in binary image. Because the gray value range of the surrounding organs is similar to the liver, it may be connected with some other components which do not belong to the liver. In order to remove these surrounding components, which is shown in Fig. 3(b), we perform morphological processing method on them, here we use erosion to remove the noise pixels in the image. Because there are some pixels look like black holes exist in the component after erosion, another morphological processing approach is applied to

fill these holes using flood filling and median filtering technology to make the image smoother. The median filter can be expressed as follows,

$$g(x, y) = \text{Median}(f_s(x, y)) \quad (2)$$

where S is a temporary cluster in the whole image. It assumes a center pixel and create a 3×3 or 5×5 window with surrounding neighbor pixels. $f_s(x, y)$ is the value of pixel in S , (x, y) is the coordinate of the pixel. $\text{Median}(f_s(x, y))$ is the middle value of all the values in S , which is used to instead the center pixel value. We can see from Fig. 3(c) that the morphological processing step can provide a smoother result with less irrelevant components and noise.

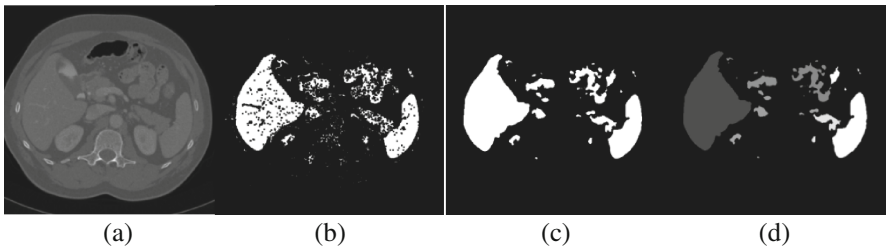


Fig. 3. (a) Original CT image, (b) Thresholded and eroded image, (c) Smooth image after morphologic processes, (d) The labeled connected components

The next step is to label the connected component which has the same label. The labelling connected component algorithm is based on an iterative method. It searched the binary pixels in the image and labelled them with their adjacent labels according to the pixel's neighborhood. Usually there are two widely used neighborhood methods. One is the Moore neighborhood with 8-connectivity and another is von Neumann neighborhood which is using 4-connected pixels. The labelling algorithm first chose the binary pixels, marked them with current label and assumed these pixels as the center, then checked their neighbor pixels. If the value of neighbor pixel is binary, then the pixel is marked by the same label with the center pixel. The result of labeling the connected component is shown in Fig. 3(d).

In the last step, we know that liver is the biggest organ in the abdominal CT image, our task is to find the largest connected component in the image. Here we use the histogram of labeled images to automatic find the largest component because the liver accounted for most of the marked areas. Using this method we can extract the largest labelled component from all the labelled components and then get the liver area. The labeled result is shown in Fig. 4(a). Using the labeled result multiplied by the original image, the final segmented result is shown in Fig. 4(b).

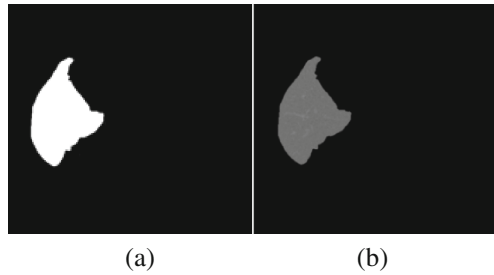


Fig. 4. (a) The labeled result of the Liver, (b) The final segmentation result

However, some results of the above automatic liver segmentation algorithm are not smooth and accurate enough, such as under-segmented or over-segmented cases as shown in Fig. 5. So refined segmentation is been used in the proposed segmentation framework after rough segmentation.

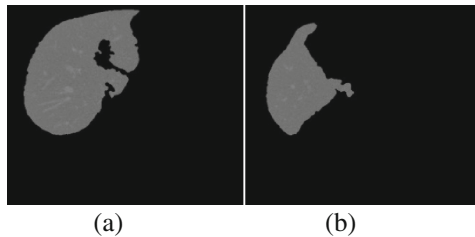


Fig. 5. (a) Under-segmented image, (b) Over-segmented image

2.2 Refined Segmentation Using IGC Algorithm with Automatic Checking

After rough segmentation, some images need to be refined. In order to find these images, We checked the total pixel number of an image and compared it with the previous one. By using thresholding methods, we got the images that need to be further segmented automatically.

In this comparison, if the difference between total pixel number of the current segmented image is more than the thresholding value, that means current segmented image has some problems, otherwise image does not need refined segmentation. When the algorithm detects any shrinkage between current segmented images and previous segmented images, connection method, which connects refined and rough segmentation, uses other predetermined threshold values to figure out whether the image need refination or not.

As mentioned above, the proposed segmentation method is totally automatic that decides which images need refination by using this algorithm. Rough segmentation with automatic liver detection is connected to refined segmentation with IGC automatically by this algorithm. In this paper, we applied the IGC algorithm to refine the rough segmentation results. The IGC algorithm was proposed by our research group, which

makes good use of the continuity of CT series in space and can automatically generate the seed labels to improve the efficiency of segmentation. The details about the IGC algorithm is given in the reference [12].

2.3 Post-processing

After all these processes, some morphologic processes such as erosion and dilation are used to smooth edge of the image. Hereby the user can obtain the final result after post-processing.

3 Evaluation

Three groups of CT images were used to evaluate the segmentation performance. The slice thickness was 1.0 mm and the spacing between slices was 0.5 mm. Each slice had 512×512 size. All segmentation framework were implemented in C++ language with OpenCV library on a 64-bit Windows 8.1 OS. All experiments carried on the computer with Intel(R) Core i7 (2.4 GHz), 8 GB memory.

According to the experimental results, a dataset consists of around 200 CT images and this segmentation method handles all these CT images in 40–45 s.

The results of this automatic liver segmentation method (M1) is compared with another segmentation method (M2) which includes rough segmentation with SKFCM and IGC. For evaluation of these methods, there are three evaluation criterions, accuracy, overlap, and total process time of the method (PTM). Accuracy is a common criterion that is used to evaluate performance of segmentation methods widely. Overlap shows the degree of overlap between segmentation results by computer algorithm and manual segmentation results. The closer overlap is to 1, the better segmentation result will be. The accuracy and overlap are defined as follows,

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$Overlap = \frac{TP}{TP + FP + FN} \quad (4)$$

where TP denotes the number of true positive pixels, TN denotes the number of true negative pixels, FP denotes the number of false positive pixels and FN denotes the number of false negative pixels (Table 1).

Table 1. The evaluation of methods.

	Methods	Accuracy (%)	Overlap (%)	PTM (s)
Data 1	M1	99.69	94.61	42.12
	M2	99.58	94.48	102.45*
Data 2	M1	99.17	93.62	44.40
	M2	99.08	93.59	105.04*
Data 3	M1	99.67	93.73	41.28
	M2	99.69	94.16	100.67*

* Interactions are not included in this process time.

The results show that the proposed automatic liver segmentation method is quite fast and reliable.

4 Conclusion

Liver diseases are a widely spread problem among all demographics of people. That is why we implemented a system related to the liver. Also in this project, we tried to remove all interactive methods and manual intervention. Thus, we proposed a full-automatic liver segmentation method.

The implementation consists of two main steps: rough segmentation with automatic liver detection and refined segmentation.

Firstly, in the rough segmentation, we used thresholding methods based on histogram to label the image and found the labeled largest components which finally identifies the liver. Therefore, we obtained the liver regions from all abdominal CT images.

Secondly, we used Improved Grow-Cut method to do refine segmentation for the under-segmentation or over-segmentation because of the blood vessels or similar surrounding tissues. We assessed a threshold to solve these problems which checked the change of the liver region pixels. If the pixel number of detected liver region increases or decreases over than the threshold value, we determined to refine concerned image slices with Improved Grow-Cut. Also, the method contains the automatic seed generation for Grow-Cut algorithm. That means we handled refined segmentation using Grow-Cut method without any interaction.

This fully automatic system does not contain any manual interaction so we hope these systems will be used widely in the clinical. In the future, these systems also can be combined with machine learning algorithms, thus the systems will recognize diseases without any doctors. Even some robotic systems can directly begin the treatment phase or surgery on the patients.

Acknowledgement. This work was supported by the National Hi-Tech Research and Development Program (2015AA043203), and the National Science Foundation Program of China (81430039, 61672099, 81430039, 61501030). This study received financial support from Frontier and interdisciplinary innovation program of Beijing Institute of Technology.

References

1. Yang, X., Yu, H., et al.: A hybrid semi-automatic method for liver segmentation based on level-set methods using multiple seed points. *Comput. Methods Programs Biomed.* **113**(1), 69–79 (2014)
2. Foruzan, A., Chen, Y., et al.: Segmentation of liver in low-contrast images using K-Means clustering and geodesic active contour algorithms. *IEICE Trans. Inf. Syst.* **96**, 798–807 (2013)
3. Zhang, X., Tian, J., Deng, K., et al.: Automatic liver segmentation using a statistical shape model with optimal surface detection. *IEEE Trans. Biomed. Eng.* **57**(10), 2622–2626 (2010)
4. Li, X., Luo, S., Li, J.: Liver segmentation from CT image using fuzzy clustering and level set. *J. Sig. Inf. Process.* **4**(3), 36–42 (2013)
5. Li, B.N., Chui, C., Chang, S., et al.: Integrating spatial fuzzy Clustering with level set methods for automated medical image segmentation. *Comput. Biol. Med.* **41**(1), 1–10 (2011)
6. Chen, S., Zhang, D.: Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **34**(4), 1907–1916 (2004)
7. Kannan, S.R., Ramathilagam, S., Sathya, A.: Effective fuzzy C-means based kernel function in segmenting medical images. *Comput. Biol. Med.* **40**(6), 572–579 (2010)
8. Zayane, O., Jouini, B., Mahjoub, M.A.: Automatic liver segmentation method in CT images. *Can. J. Image Process. Comput. Vis.* **5**(1), 85–92 (2011)
9. Vezhnevets, V., Konouchine, V.: GrowCut: interactive multi-label ND image segmentation by cellular automata. In: *Graphicon Novosibirsk Akademgorodok*, pp. 150–156 (2005)
10. Hamamci, A., Unal, G., Kucuk, N., et al.: Cellular automata segmentation of brain tumors on post contrast MR Images. *Med. Image Comput. Comput. Assist. Interv. MICCAI* **13**(3), 137–146 (2010)
11. Yamasaki, T., Chen, T., Yagi, M., et al.: GrowCut-based Fast Tumor Segmentation for 3D Magnetic Resonance Images. *SPIE Medical Imaging, International Society for Optics and Photonics* **8314**, 208–220 (2012)
12. Song, H., Zhang, Q., Wang, S.: Liver segmentation based on SKFCM and Improved GrowCut for CT images. In: *IEEE International Conference of Bioinformatics and Biomedicine (BIBM)*, pp. 331–334 (2014)

An Improved Blind Spectrum Sensing Algorithm Based on QR Decomposition and SVM

Yaqin Chen^{1,2}(✉), Xiaojun Jing^{1,2}, Wenting Liu^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
cyqbuptstu@163.com

² Key Laboratory of Trustworthy Distribution Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ School of Engineering and Computer Science,
Oakland University, Rochester, USA

Abstract. Spectrum sensing, a basic functionality in cognitive radio, aims at detecting the presence or absence of primary user (PU). As one of the most popular spectrum sensing methods, Covariance-based sensing works based on the correlation between signal samples. However, its performance sharply declines in low Signal Noise Ratio (SNR) environment. To improve detection performance of covariance-based sensing as far as possible, an improved blind spectrum sensing scheme is proposed in this paper on the basis of QR matrix decomposition and support vector machine (SVM). In the proposed scheme, QR matrix decomposition is applied to the co-variance matrix of received signal firstly, and then the main features are constituted by extracting and arranging orderly the upper triangular elements of R matrix. After that, SVM is used to conduct the obtained features and determine whether PU exists. The proposed algorithm does not need the prior information of PU and noise. Simulation results demonstrate that the proposed method has a better performance than conventional covariance-based methods, especially in low SNR scenarios.

Keywords: Spectrum sensing · Covariance-based sensing · QR decomposition
SVM

1 Introduction

In recent years, spectrum scarcity has caused widespread concern. However, it is reported by FCC that the usage rate of fixed spectrum varies from 15% to 85% [1], which means that the traditional spectrum allocation method leads to the incomplete utilization of spectrum. To solve the problem, cognitive radio (CR) was proposed.

Spectrum sensing (SS), a key technique in CR, which aims at detecting the existence of PU. Conventional sensing methods include energy detection, cyclostationary feature detection, likelihood ratio test and so on [2].

All of the above algorithms need the prior information about signal or noise, which may not be realistic in practice. Therefore, some blind spectrum sensing algorithms emerged. Among these blind sensing algorithms, eigenvalue-based [3, 4], and covariance-based sensing methods [5] are widely accepted. Because they do not need any prior information and take the correlation between signal samples into account. It was voted that [6] proposed a blind scheme, features from the Cholesky decomposition of covariance matrix as the criterion to determine whether the radio frequency band is vacant, which was proved to have good detection performance. In addition, to overcome channel fading and hidden terminal, [7] put forward a cooperative sensing algorithm.

As we all know, SS is actually a binary classification problem. Therefore, more and more machine learning classification algorithms are applied in SS in recent years. In [8], Awe OP extracted eigenvalues as the features of signal then used SVM for classification, it is proved by experiment that the algorithm has better performance. Then, a paper proves the superiority of SVM [9]. Considering the mobility of secondary users, [10] used the random forest to achieve a better network throughput.

In this paper, motivated by the work in [5–10], we proposed a blind spectrum sensing algorithm based on QR decomposition and SVM. Firstly, the covariance matrix of cooperative secondary users (SUs) is estimated. Secondly, we employ QR decomposition of the covariance matrix to extract the features of signals when PU exit or not. Finally, SVM is applied to classify whether PU presents.

The rest of this paper is organized as follows. In Sect. 2, the system model is given. In Sect. 3, the process of feature extraction and the application of SVM is described. In Sect. 4, simulation results are discussed. Finally, we draw our conclusion in Sect. 5.

2 System Model

2.1 Spectrum Sensing

The main task of spectrum sensing is to detect the presence of PU. Thus, spectrum sensing can be represented as a binary hypothesis testing problem, which may be written as

$$\begin{cases} H_1 : & y(k) = s(k) + n(k) \\ H_0 : & y(k) = n(k) \end{cases} \quad k = 1, 2, \dots, K \quad (1)$$

where K denotes the number of samples and $y(k)$ indicates the signal of SU received from the PU transmitter. Additionally, $s(k)$ and $n(k)$ respectively denote the signal of PU and noise. When PU exists, SU receives the signal of PU and noise, represented as H_1 . Otherwise, SU only get noise signal, represented as H_0 .

2.2 Cooperative Covariance-Based Detection

In covariance-based detection [5], $s(k)$ and $n(k)$ meet the basic assumptions: (1) $n(k)$ is an independent, identically distributed Gaussian signal, satisfying $E(n(k)) = 0$, $E(n^2(k)) = \sigma_n^2$. (2) The samples of $s(k)$ are correlated.

Supposing that the number of cooperative SUs is L . Then we get the following vectors:

$$Y = [y_1(k) \quad y_2(k) \quad \dots \quad y_L(k)]^T \quad (2)$$

$$S = [s_1(k) \quad s_2(k) \quad \dots \quad s_L(k)]^T \quad (3)$$

$$N = [n_1(k) \quad n_2(k) \quad \dots \quad n_L(k)]^T \quad (4)$$

where

$$Y = \begin{bmatrix} y_1(1) & y_1(2) & \dots & y_1(K) \\ y_2(1) & y_2(2) & \dots & y_2(K) \\ \vdots & \vdots & & \vdots \\ y_L(1) & y_L(2) & \dots & y_L(K) \end{bmatrix}$$

Then the $L \times K$ ($0 < L/K < 1$) statistical covariance matrix of Y , S , N can be written as

$$R_Y = E[Y \cdot Y^H] \quad (5)$$

$$R_s = E[S \cdot S^H] \quad (6)$$

$$R_N = E[N \cdot N^H] = \sigma_n^2 I_L \quad (7)$$

$$\begin{cases} H_1 : & R_Y = R_s + \sigma_n^2 I_L \\ H_0 : & R_Y = \sigma_n^2 I_L \end{cases} \quad (8)$$

From the above formulae, we can notice that when PU is absent, $R_s = 0$. Because the samples of noise are independent, R_Y is a diagonal matrix. When PU is present, due to the correlation of PU signal, the off-diagonal entries of R_s are non-zero. In this case, R_Y is not a diagonal matrix.

3 The Improved Algorithm

3.1 The QR Decomposition of Signal

In pattern recognition, QR decomposition has been widely used. Because Q matrix in QR decomposition is a group of orthogonal eigenvectors, R matrix covers all information of matrix and reflects the main features of signal. And the main advantage of QR decomposition is the numerical stability, which is suitable for spectrum sensing with Gaussian signal and noise. In [11], QR decomposition is successfully applied in channel identification.

Based on the above theories, we decompose R_Y by using QR decomposition.

$$R_Y = QR \tag{9}$$

where

$$Q = [q_1 \quad q_2 \quad \dots \quad q_L]$$

$$q_i \cdot q_j = 0 \quad 1 \leq i \leq L$$

In the hypothesis of H_0 , $R_Y = R_N = \sigma_n^2 I_L$. According to (9), we get

$$R = \sigma_n^2 I_L = \begin{bmatrix} \sigma_n^2 & & & \\ & \sigma_n^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$$

In the hypothesis of H_1 , because of the correlation between SUs, R matrix of R_Y is not a diagonal matrix.

From the above analysis, we get

$$\left\{ \begin{array}{l} H_1 : R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1L} \\ & r_{22} & \dots & r_{2L} \\ & & \ddots & \vdots \\ & & & r_{LL} \end{bmatrix} \\ H_0 : R = \begin{bmatrix} \sigma_n^2 & & & \\ & \sigma_n^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} \end{array} \right. \tag{10}$$

Therefore, the upper triangular matrix R can be used to differentiate PU from noise. After arranging the elements of R in rows, the features when PU present or not are extracted and represented as a vector.

3.2 SVM Based Sensing Algorithm

In the classification algorithms, SVM is regarded as one of the best classifiers. In this paper, we use nonlinear SVM to achieve the detection of PU.

Combing the features and corresponding labels, the training set and the testing set can be obtained. Assume the testing set as $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)\}$, M is the number of training samples. x_i , a $1 \times (L \times (L + 1)/2)$ vector, is the feature, consisting of the upper triangular elements of matrix. y_i , the corresponding label, equals 1 when PU present, otherwise -1.

The nonlinear SVM achieves classification based on the optimal separation hyper-plane. The optimal hyperplane and classification decision function can be represented respectively as [10].

$$\begin{aligned} w^* \cdot \varphi(x) + b^* &= 0 \\ f(x) &= \text{sign}(w^* \cdot \varphi(x) + b^*) \end{aligned} \quad (11)$$

where $\varphi(x)$ is a mapping function, which maps x_i into a high dimensional space.

According to the conditions of the optimal hyperplane, Lagrange function can be constructed. Then the question of the optimal hyperplane can be converted to

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^M \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^M \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, M \end{aligned} \quad (12)$$

where α represent Lagrange multiplier. $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$. C is a positive constant penalty parameter.

Based on (12), we can get the optimal solution of α as $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_M^*)^T$.

Then the optimal solution of w is calculated by

$$w^* = \sum_{i=1}^M \alpha_i^* y_i x_i \quad (13)$$

Selecting α_i^* , which satisfies $0 \leq \alpha_i^* \leq C$, the optimal solution of b can be obtained

$$b^* = y_j - \sum_{i=1}^M \alpha_i^* y_i K(x_i, x_j) \quad (14)$$

Finally, the classification decision function can be written as

$$f(x) = \text{sign}\left(\sum_{i=1}^M \alpha_i^* y_i K(x \cdot x_i) + b^*\right) \quad (15)$$

4 The Simulation Result and Discussion

In simulation, the modulation of PU signal and the noise are respectively assumed to be OFDM and AWGN. The carrier frequency and the sampling frequency are respectively 100 MHz and 400 MHz. The value of L is 5, and the value of K is 1000. The simulation mainly includes three steps as follows.

1. We generate the training set and testing set according to the third section of this paper. The size of them is respectively 8000 and 2000.
2. The training set then is sent into SVM to get the classification decision function.
3. Classification decision function obtained from step 2 is applied to feature vectors of the testing set. Then the output of SVM are compared with corresponding labels in testing set to compute the probability of detection P_d and false alarm P_{fa} .

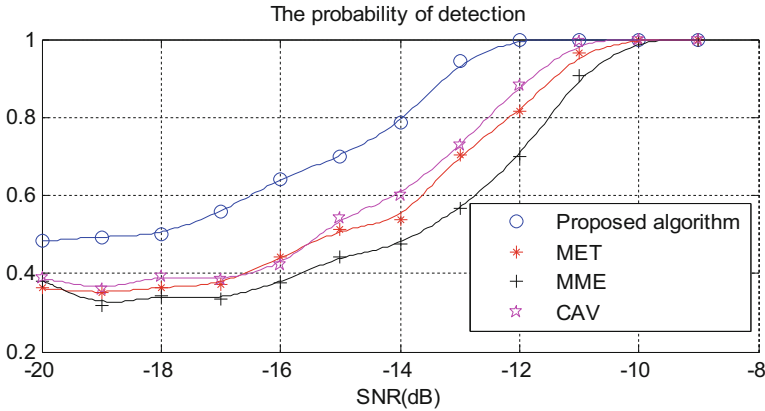


Fig. 1. The comparisons of the detection probability

Figure 1 compares the proposed algorithm with maximum-minimum eigenvalue (MME) [3], maximum-trace ratio (MET) [4], and covariance absolute value (CAV) [5]. In this figure, their P_{fa} at the certain SNR are same. It is obvious that the P_d of proposed method reaches 49% at the SNR of -20 dB, while the P_d of MME, MET, and CAV are below 40%. When SNR between -20 dB and -12 dB, the P_d of proposed method outperforms other algorithms about 15%.

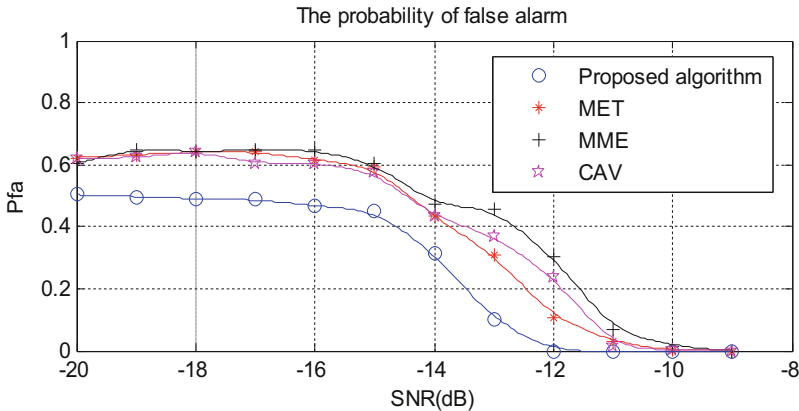


Fig. 2. The comparisons of the false alarm probability

Figure 2 compares their P_{fa} for given P_d at various SNRs. It is worth noting that the P_{fa} of proposed method is lower than other methods. At the SNR of -12 dB, the P_{fa} of the proposed method converges to 0, while others are above 0.1.

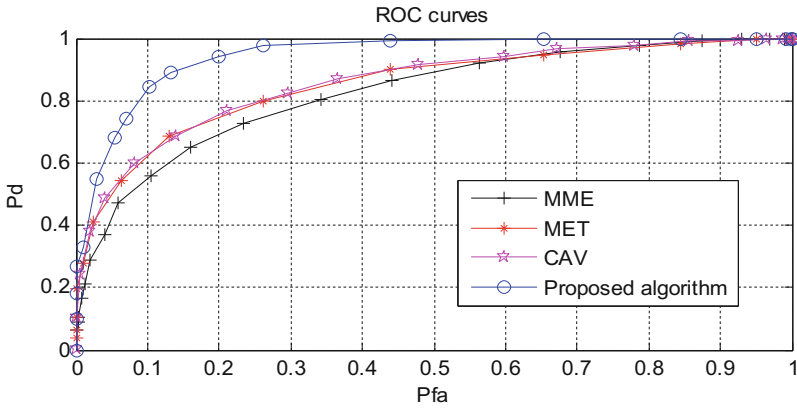


Fig. 3. The ROC curves at the SNR of -15 dB

Figure 3 shows their ROC curves at the SNR of -15 dB. When P_{fa} equals 0.1, the P_d of proposed method reaches 0.84, which is greater than other algorithms. It is clear that the proposed method has better detection performance than MME, MET, and CAV.

In summary, the detection performance of the proposed method is better than conventional covariance-based methods. The first two figures show that the proposed method is more stable than MME, MET, and CAV when SNR is low. The reason is that the traditional methods cannot distinguish between signal and noise at low SNRs. While features extracted from QR decomposition cover the all key information, which can distinguish signal and noise effectively. The last figure shows the superiority of the combination of QR decomposition and SVM.

5 Conclusions

In this paper, a blind cooperative spectrum sensing algorithm is put forward combing QR decomposition and SVM. This method does not demand the information of signal or noise and has good performance than conventional covariance-based methods. Simulation results verified the superiority of the proposed scheme, and our future work will focus on the influence of different modulation types.

Acknowledgement. Project 61471066 supported by NSFC.

References

1. Kolodzy, P.: Spectrum policy task force report. In: Federal Communications Commission Technical report et Docket 40(4), 147–158 (2002)
2. Li, B., Zhao, L., Sun, M., Zhou, Z., Nallanathan, A.: Spectrum sensing for cognitive radios in time-variant flat fading channels: a joint estimation approach. *IEEE Trans. Commun.* **62**(8), 2665–2680 (2014)
3. Kortun, A., Ratnarajah, T., Sellathurai, M., et al.: On the Eigenvalue-based spectrum sensing and secondary user throughput. *IEEE Trans. Veh. Technol.* **63**(3), 1480–1486 (2014)
4. Huang, L., Fang, J., Liu, K., et al.: An Eigenvalue-moment-ratio approach to blind spectrum sensing for cognitive radio under sample-starving environment. *IEEE Trans. Veh. Technol.* **64**(8), 3465–3480 (2015)
5. Zeng, Y., Liang, C.: Spectrum sensing algorithms for cognitive radio based on statistical covariances. *IEEE Trans. Veh. Technol.* **58**(4), 1804–1818 (2008)
6. Yang, X., Lei, K., Peng, S., et al.: Blind detection for primary user based on the sample co-variance matrix in cognitive radio. *IEEE Commun. Lett.* **15**(1), 40–42 (2011)
7. Qian, L., Hu, Q., Qian, Y., Wu, G.: Cooperative communications for wireless networks: techniques and applications in LTE-advanced systems. *IEEE Wirel. Commun.* **19**(2), 22–29 (2012)
8. Awe, O.P., Zhu, Z.: Eigenvalue and support vector machine techniques for spectrum sensing in cognitive radio networks. In: *Technologies and Applications of Artificial Intelligence*, pp. 223–227. IEEE (2014)
9. Thilina, M., Choi, W., Saquib, N.: Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE J. Sel. Areas Commun.* **31**(11), 2209–2221 (2013)
10. Wang, D., Yang, Z.: A novel spectrum sensing scheme combined with machine learning. In: *International Congress Image and Signal Processing, BioMedical Engineering and Informatics*, pp. 1293–1297. IEEE (2016)
11. Shigyo, D., Tanabe, S., Ishibashi, K.: QR-decomposed generalized belief propagation for MIMO detection. In: *IEEE International Symposium on Information Theory and ITS Applications* (2017)

Sentiment Analysis Using Modified LDA

Jingyi Ye^{1,2(✉)}, Xiaojun Jing^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
yejingyi@naver.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China

³ School of Engineering and Computer Science, Oakland University, Rochester, USA

Abstract. The technology of the Internet develops rapidly recent years, the public tends to share their reviews, opinions and ideas on the Internet. The forms of these subjective texts are free and concise, and they contain a wealth of sentiment information. In this paper, a modified latent Dirichlet allocation (LDA) model and support vector machine (SVM) are used for sentiment analysis of subjective texts. Analysis of sentiment could help producer to enhance the products and guide user make better choices as well. We apply a modified LDA model using term frequency-inverse document frequency (TF-IDF) algorithm to mine potential topics, find the most relevant words of the topic and represent the document. Then we use SVM to categorize the texts into two classes: positive and negative. Experiment results show that the performance of the modified LDA approach is better than the traditional LDA model.

Keywords: Opinion mining · Sentiment analysis · Latent Dirichlet allocation
TF-IDF algorithm

1 Introduction

The technology of the Internet and the integration of dynamic web developed rapidly in the past decades. People begin to enjoy the fun of network applications and express their mood, thoughts and ideas on the network. Thus, a large number of potentially valuable sentiment texts appear on the Internet.

Sentiment analysis has been widely used in many areas, and can be beneficial to many aspects. From the perspective of individual users, sentiment analysis could affect the individual's understanding and attitude towards the specific object. Form an enterprise point of view, sentiment analysis of user views and service experience can enable enterprises to know their products for the merits of an accurate grasp and develop

Project 61471066 supported by NSFC.

effective strategies in business competition and product development. From the government point of view, sentiment analysis can help government promptly understand people's views and attitudes [1].

Recently, researchers show a great interest in exploring new text representation models for improving the accuracy and efficiency of text processing. The theoretical idea of the topic model is that document is a mixture of several topics, each of which contains multiple terms of word distribution. Topic model obtains the semantic related topic collection hidden in the document through the common information of words in the document. Topic model transforms the document from word space to topic space, and expresses the document in a lower dimension space. The origination of topic model was from latent semantic indexing (LSI), and then topic model evolved to a variety of forms, especially latent Dirichlet allocation (LDA) and LDA based models. The topic model has a sentiment feature that can extract the implicit sentiment topic from the data set. Due to the good mathematical foundation and flexible extensible of topic model, it has been widely used in text mining and information processing tasks [2].

It is time-consuming to extract words manually in a document, so we need to extract words automatically. People use algorithms to divide the document into words, give the corresponding weight of each word according to some methods. Many approaches can be utilized for the weight calculation, such as frequency function, entropy function and term frequency-inverse document frequency (TFIDF). TF-IDF is a statistical method with a high accuracy and recall, and it can be applied to a document in a corpus for evaluating the importance of a word. The principle of this method is that if a word appears in the document at a high frequency and is rarely present in other documents, it is considered that the word has a good class distinction and is suitable for the classification [3].

In this paper, we analyze the sentiment of subjective documents with a modified LDA model and SVM. We use TF-IDF algorithm to modify the traditional LDA model. According to the potential topic mining method, the documents are represented by the modified LDA model. Then the topics are divided into two classes via support vector machine (SVM), namely positive and negative. Experiment results prove the effectiveness and practicability of the modified model.

The outline of this work is shown as follows. In Sect. 2, a brief introduction of sentiment analysis, LDA model and TF-IDF algorithm is presented. A specific description of the proposed approach is followed in next section. After that, the experiment results are shown in Sect. 4, and conclusions are made in the final section.

2 Related Work

2.1 TF-IDF Algorithm

Jones first proposed the concept of Opposed to Document Frequency (IDF) in [4] in 1972. He pointed out that the weight of words can be given according to the frequency of words appeared in the corpus. If a word has a higher number of occurrences in the corpus, it has a lower information entropy and the corresponding weight, and vice versa.

Back to 1973, the TF-IDF algorithm was first proposed by Salton, and it was effectively applied to the field of information retrieval [5]. In 1988, a detailed description of the use of multiple words weighting method was proposed [6]. TF-IDF mainly reflects the following idea: the higher frequency of a word, the stronger its ability to distinguish the content of the document (TF) and the wider the scope of a word in the corpus, the lower attributes of the document content (IDF) [6].

2.2 Latent Dirichlet Allocation

LDA is a generation probability model, which is the method of modeling the topic information of text data [2], with good mathematical basis and flexible expansibility. LDA could mine words with topic, through the three-tier Bayesian model. Titov [6] proposed a multi-granularity LDA model and applied it to sentiment summary generation with a multi-topic sentiment model. Zhao [7] proposed the ME-LDA model, which combines the maximum entropy with the topic model.

2.3 Sentiment Analysis

With the arrival of large data age, the analysis of massive data could access valuable products or services. However, most of the information is unstructured and difficult to manually analyze. Sentiment analysis is a new research direction that rises to deal with the analysis of implied emotional information [8]. Sentiment analysis can automatically analyze the subjective text, effectively identify and excavate the emotional information. The main tasks of sentiment analysis include the extraction, classification and retrieval. Sentiment extraction refers to the extraction of relevant sentiment words and evaluation of objects. Sentiment classification is to distinguish subjective and objective texts and determine the polarity of subjective text. Sentiment retrieval is used to retrieve documents containing relevant sentiment information to meet users' query needs. In this paper, we mainly concerns on the extraction and classification tasks.

3 The Proposed Approach

Basic steps of the proposed approach are as follows: we organize and preprocess the documents in the corpus at first, then we use the modified model to learn text representation of training set and test set, finally train the SVM classifier to complete the sentiment classification of test set.

3.1 Pretreatment

The documents in the corpus are chapters composed of sentences. In order to fit the structured text needed by the classification algorithm, it is necessary to preprocess the corpus.

The main pretreatment in this paper is to remove the stop words. The purpose of this process is to reduce the spatial complexity and time complexity of the method and to

improve the accuracy of the follow-up feature selection. Stop words usually do not make sense or have contribution to the classification results, including punctuation, modal particles, prepositions and conjunctions, such as “are”, “above”, “and”, etc. The implementation of this process requires a vocabulary of stop words, and these words can be filtered out by applying an approach of string matching. For instance, the target is “a little girl held an apple from the basket and cried”, “a”, “an”, “from”, “the”, “and” are stop words, and the remaining words are “little girl held apple basket cried”.

3.2 Model Description

After the preprocessing, the modified LDA model is used to deal with the corpus. We use TF-IDF algorithm to modify LDA model.

TF-IDF Algorithm

TF-IDF is a widely used algorithm in information retrieval field. In recent years, researchers have used TF-IDF algorithm to calculate the weight of features and achieved good results [5].

$$W = TF \times IDF = TF \times \frac{1}{DF}. \quad (1)$$

The frequency of word T in document D is TF, which is used to calculate the capability of the word to describe the document. IDF represents the inverse of the frequency of the document D containing the word T in the corpus, which is used to calculate the capability of the word to distinguish the document. If the frequency of a word is high in its own document but low in other documents, this word has a strong ability to distinguish it from other documents and is assigned to a high weight.

Latent Dirichlet Allocation

The generation probability model LDA imitates the process of human writing. If we want to write a document, it is often necessary to determine what topics to write. After determining the topics, we will use some words highly related to the topics to describe them. Thus, a document usually consists of multiple topics, each topic is described with the high frequency words associated with the topic.

Blei proposed an unsupervised full probability generation model, LDA, in 2003 [2]. LDA has a clear internal structure and mathematical basis, which could be calculated by efficient probability estimation algorithm. LDA model has been widely used in text classification, text modeling, image processing and information retrieval and other fields.

LDA model is a method of modeling documents, topics and words, which transforms the traditional word vector expression into the topic vector expression [9]. The benefit of doing this is obvious that the expression based on topics reduces the dimension of feature space. LDA model has a clear logical structure, containing the document layer, the topic layer and the word layer. Each layer is adjusted by variables and parameters. Figure 1 shows the graphical representation of LDA model.

The box indicates that the content is repeated, the number of repetitions is in the lower right corner, the shadow represents the observed value, the empty node represents the implied random variable or parameter, and the arrow indicates the dependency.

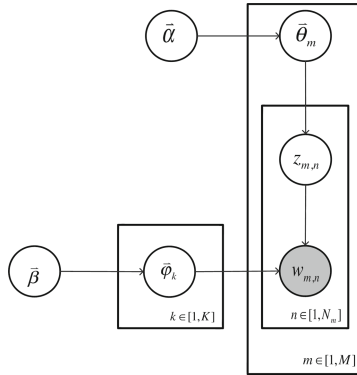


Fig. 1. Graphical model representation of LDA

The symbols in LDA model are explained as follows:

- (1) K : the number of topics,
- (2) M : the number of documents,
- (3) N_m : the total number of words in document m ,
- (4) α : Dirichlet prior parameters of multinomial distribution θ_m under each document,
- (5) β : Dirichlet prior parameters of multinomial distribution φ_k under each topic,
- (6) $z_{m,n}$: the topic of the word n of the document m ,
- (7) $w_{m,n}$: the word n of the document m ,
- (8) θ_m : the topic distribution under the document m ,
- (9) φ_k : the word distribution under the topic k .

Modified LDA

TF-IDF algorithm is used to improve the performance of LDA. The main idea of the modified LDA model is to replace the untreated words in the traditional LDA model using the words treated by TF-IDF algorithm. The words treated by TF-IDF algorithm have a weight and have the ability to express the importance of the word to the current document and to distinguish it from other documents. The structure of modified LDA is shown in Fig. 2, and t_k is used to describe the word distribution which is treated by TF-IDF algorithm in the slash part.

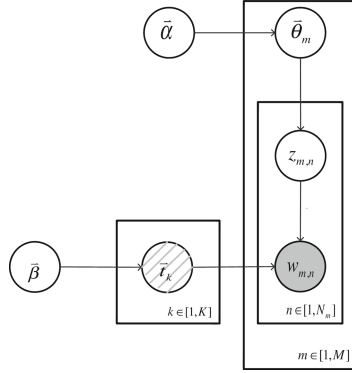


Fig. 2. Graphical model representation of modified LDA

The corpus D contains all documents, which served as the input part of the modified LDA model. K is the number of topics. The output of modified LDA is the most relevant words of each topic. The generative process of the modified model is as follows:

For each document, we generate the multinomial distribution of topic θ_m , which is derived from the Dirichlet distribution of the parameter α .

For each topic, we generate the multinomial distribution of word t_k , which is derived from the Dirichlet distribution of the parameter β . For each topic, the sum of $z_{m,n}$ is calculated by using Gibbs sampling and an approximate posterior on θ_m and t_k is obtained. The formula is as follows:

$$p(\Theta, \Phi | D^{train}, \alpha, \beta) = \sum_z p(\Theta, \Phi | z, D^{train}, \alpha, \beta) \times p(z | D^{train}, \alpha, \beta). \tag{2}$$

We repeat the above procedure to get the output of modified LDA. Through the above process, we get a list of topics for each document, and for each topic we get some of the most relevant words with the topic.

4 Experiment Results

We demonstrate the performance of modified LDA model by experiment in this section. The dataset we used is the sentiment dataset from Cornell University, which contains 2000 movie reviews with both positive and negative data of 1000. In the experiment, we use 1500 reviews as training data and 500 reviews as test data. The classifier in the experiment we used is SVM. Also we use five-fold cross validation method and measure the performance with precision, recall and F1-value.

The experiment compares the influence of different topic numbers on LDA and modified LDA and the result of F1-value is shown in Fig. 3 and the result of precision and recall is shown in Table 1. The results show that the modified LDA has a better performance than traditional LDA model. When topic number is 40, we get the best performance with precision at 0.9, recall at 0.87 and F1-value at 0.87.

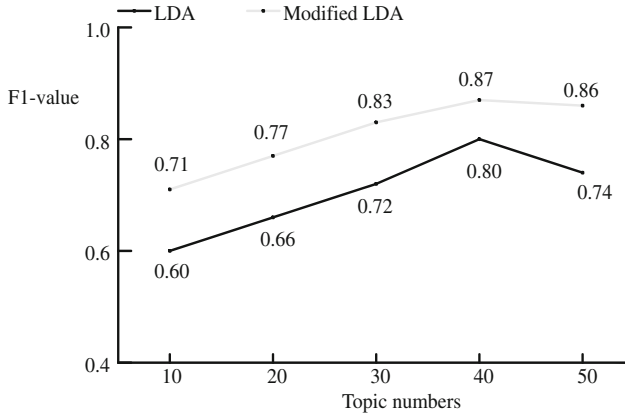


Fig. 3. The classification performance of F1-value using LDA and modified LDA to represent the document, with different topic numbers.

Table 1. The classification performance of precision and recall using LDA and modified LDA to represent the document, with different topic numbers.

Topic numbers	LDA		Modified LDA	
	Precision	Recall	Precision	Recall
10	0.63	0.61	0.82	0.73
20	0.67	0.67	0.85	0.78
30	0.72	0.72	0.88	0.84
40	0.8	0.8	0.9	0.87
50	0.74	0.73	0.86	0.86

This section mainly discusses the corpus selection, evaluation criteria, LDA model and modified LDA model experimental results. The text sentiment classification of modified LDA model has improved the classification quality than traditional LDA model. The comparative experiment using movie reviews data show the effectiveness of the modified LDA model.

5 Conclusions

Sentiment analysis is an important research area in natural language processing field. It is of great significance for the research of Internet public opinion supervision, commodity sales and information screening. We have studied the text representation model in sentiment analysis problem, and modified the LDA model with TF-IDF algorithm. Experiments show the effectiveness of modified LDA. In this paper, we use modified LDA model in binary classification problem, in the future work we can use the modified model in multi-classification problem.

References

1. Cambria, E.: Affective computing and sentiment analysis. *IEEE Intell. Syst.* **31**(2), 102–107 (2016)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
3. Guo, A., Yang, T.: Research and improvement of feature words weight based on TFIDF algorithm. In: 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference, pp. 415–419 (2016)
4. Jones, K.S.: A statistical interpretation of term specificity and its application in retrieval. *J. Documentation* **28**(1), 11–21 (1972)
5. Salton, G., Yu, C.T.: On the construction of effective vocabularies for information retrieval. In: Proceedings of the 1973 Meeting on Programming Languages and Information Retrieval, pp. 11–21 (1972)
6. Titov, I., McDonald, R.: Modeling online reviews with multi-grain topic models. In: The 17th International World Wide Web Conference, pp. 111–120 (2008)
7. Zhao, W.X., Jiang, J., Yan, H., Li, X.: Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 56–65 (2010)
8. Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., Bandyopadhyay, S.: Enhanced SenticNet with affective labels for concept-based opinion mining. *IEEE Intell. Syst.* **28**(2), 31–38 (2013)
9. Li, Y., Zhou, X., Sun, Y., Zhang, H.: Design and implementation of Weibo sentiment analysis based on LDA and dependency parsing. *China Commun.* **13**(11), 91–105 (2016)

Co-training Based on Multi-type Text Features

Wenting Liu^{1,2(✉)}, Xiaojun Jing^{1,2}, Yaqin Chen^{1,2}, and Jia Li³

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
liuwentinghb@126.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications,
Beijing, China

³ School of Engineering and Computer Science, Oakland University,
Rochester, USA

Abstract. Sentiment classification is intended to classify the sentiment color categories expressed by the text. This paper illustrates the sentiment classification method based on the semi-supervised algorithm that aims to improve performance by using unlabeled data. This paper proposes a novel co-training style semi-supervised learning algorithm in order to improve semi-supervised learning ability. In our algorithm, there are three classifiers trained on the original labeled data, where the text representation for each classifier is unigram, bigram, and word2vec, respectively. And then these classifiers can use unlabeled data to update themselves. In detail, any of two classifiers have the same label, then add the new labeled data to a training set of the third classifier. By combining different types of features, our algorithm can extract text information from multiple views which contribute to sentiment classification. In addition, this algorithm doesn't require redundant and sufficient perspectives. Experiments show that our algorithm is superior to traditional co-training algorithm and partial semi-supervised learning algorithm.

Keywords: Sentiment classification · Semi-supervised learning
Co-training · Multi-type text features

1 Introduction

In recent decades, user-generated subjective texts quickly emerged, which contains a large quantity of useful information. In order to analyze and mine valuable opinions of the texts, sentiment analysis came into being. Sentiment classification divides the target text into positive or negative through analyzing the subjective texts [5].

At present, many researchers interest in supervised learning of sentiment classification. However, supervised learning depends on massive labeled data. To solve this problem, semi-supervised learning is applied to sentiment analysis, it takes advantage of both labeled and unlabeled data.

Blum and Mitchell (1998) proposed a high-performance semi-supervised learning algorithm called co-training [1]. The algorithm uses labeled data to train classifiers on two different views, and then adds the new labeled data which was predicted by the

other classifier to training set of each classifier so that the classifier can use the new labeled data to update itself. The algorithm requires two sufficient and conditional independent views, “sufficient” means that each view contains enough information to generate a strong classifier, “conditional independent” means that the two views should be independent. However, it is difficult to meet both sufficient and independent conditions in most practical applications. Goldman and Zhou (2000) put forward an improved co-training algorithm that does not need to satisfy multi-views condition. However, it needs two different supervised learning algorithms which divide the example space into multiple equivalence classes, and then labeling data through cross validation technique. The negative impact of the extensive apply cross validation to the algorithm include a high time complexity.

In this paper, we proposed a novel co-training algorithm for sentiment classification which does need to meet sufficient and independent conditions, nor does it need to use two totally different supervised learning algorithms which divide the example space into multiple equivalence classes. Thus it is more adaptable. Compared with the above-mentioned algorithms, our algorithm made several changes to attain a better performance. The innovation of our approach includes following two aspects: (1) our approach adopts three classifiers compared with co-training. (2) the text features of each classifier are different.

The remaining part of this paper is organized as follows. Section 2 makes a brief introduction of semi-supervised sentiment analysis. Section 3 illustrates our novel co-training approach for sentiment classification, Sect. 4 evaluates our approach based on the experimental results, Sect. 5 draws the conclusion.

2 Related Work

Supervised sentiment classification is the current mainstream method, it was first introduced into the sentiment classification task by Pang et al., and has achieved good classification performances. A substantial number of follow-up studies have focused on enhancing the performance of supervised learning.

An increasing amount of researchers focus on semi-supervised sentiment classification for the past few years. Wan (2009) proposed an algorithm based on co-training which employs English and Chinese as different views for sentiment classification, English and Chinese have significant logical expression difference [3]. Li et al. (2010a) proposed a co-training approach which exploits personal view and non-personal view for sentiment classification [4]. Dasgupta and Ng (2009) combine several technologies including active learning, spectral clustering, transductive learning and ensemble learning to sentiment classification [10]. However, the accuracy rate of the experiment was low in [2], an improved co-training algorithm which exploits two different supervised learning algorithm for co-training was proposed by Goldman and Zhou. In [7], the two authors adopt an algorithm which uses three classifiers for co-training. Like the above algorithms, our algorithm is also based on co-training, it uses original labeled data set to train three classifiers on different text representation models [11].

3 Improved Co-training Algorithm

3.1 Algorithm Principle

The traditional co-training algorithm needs to satisfy the sufficient and independent conditions. In the actual scenarios, two sufficient and conditional independent views are difficult to find. In order to tackle the problem, our approach use three classifiers. Zhou and Li (2005) have proved that using three classifiers neither needs to meet the sufficient and independent conditions nor needs to use different supervised learning algorithms. In addition, these three classifiers in our algorithm should have greater differences. If the three classifiers are all the same, the results of labeled data obtained by any two classifiers are consistent with the result of the third classifier. Under these circumstances, our algorithm degenerate into self-training algorithm. In the co-training, the diversity of the classifiers can be ensured by satisfying independent and sufficient conditions. In [6], Goldman and Zhou (2000) put forward an algorithm which does not need to meet redundant and sufficient conditions, using two different supervised learning algorithm enable the two classifiers to be diverse. Zhou (2007) proposed a tri-training algorithm uses bootstrap sampling technique to gain three diverse classifiers [7]. Our approach adopts different feature representation models to achieve the diversity of three classifiers. In detail, extracting three different types of feature from original labeled data, and using them to train three classifiers, these classifiers then update through exploiting unlabeled data. each text representation model has its own unique advantages for sentiment classification, considering some of them are complementary and interrelated to some extent. Our algorithm fully takes advantage of different types of features which contribute to a better performance of sentiment classification, thus our algorithm is more advantageous than above-mentioned algorithms.

The main procedure of our algorithm is as follows:

- a. Generating three classifiers from original labeled training set that uses different types of features.
- b. During the training process, put new labeled data to the third classifier's training set if the other two classifiers have the same prediction.
- c. Using updated labeled data set to train classifiers.
- d. Continuing to iterate until a certain stop condition is reached.

The pseudocode of the algorithm is as follows.

3.2 Text Feature

This paper chooses three different types of features, which are bigram, unigram and word2vec.

N-gram

Unigram is individual word tokens separated by a punctuation mark or a whitespace, bigram is pairs of adjacent word tokens. For instance, consider the following sentence:

“I love this new phone, and its music experience is great”.

Firstly, remove the stop words, then the sentence becomes the following: ‘love phone, music experience great’.

The features (Unigram, Bigram) are shown in Table 1. Individual features are separated by square brackets “[]”.

The pseudocode of our algorithm

```

Input: Original labeled data  $L$ 
      Unlabeled data  $U$ 
      Learning algorithm
Output: New classifier  $C$ 
Procedure:
  for  $i \in \{1, 2, 3\}$  do
    extract text feature  $F_i$  from  $L$ 
    use  $L$  to train a classifier  $h_i$  that text feature is  $F_i$ 
  end for
  set  $L_i$  equals to  $\phi$  ( $i \in \{1, 2, 3\}$ )
  repeat until none of  $L_i$  ( $i \in \{1, 2, 3\}$ ) changes
    set  $S_i$  equals to  $\phi$  ( $i \in \{1, 2, 3\}$ )
    for every  $x$  in  $U$  do
      use classifier  $h_i$  to predict label of  $x$ 
      if  $h_j = h_k$  ( $j \neq i$  and  $k \neq i$ )
        then  $S_i = S_i \cup \{(x, h_j(x))\}$ 
      end for
    for  $i \in \{1, 2, 3\}$  do
       $L_i = L_i \cup S_i$ 
      remove duplicate sample for  $L_i$ 
      use  $L_i \cup L$  to learn a classifier  $h_i$ 
    end repeat
  end repeat

```

Table 1. Text feature from the sample data

Feature set	Text features
Unigram	[love][phone][music][experience][great]
Bigram	[love-phone][phone-music][music-experience][experience-great]

word2vec

Word2vec is an open source and efficient tool published by Google in 2013, and each word is characterized as a numerical vector. The distributed characterizations of each word are obtained by training neural network with one hidden layer [8].

There are two types of word2vec models, the continuous bag-of-words model and Skip-gram model. CBOW model predicts the center word $w(t)$ when its context is known, while the skip-gram model, on the contrary, predicts its context under the condition that the center word $w(t)$ is already known. Because the training procedure of CBOW model is similar to the training procedure of Skip-gram model, the following part only introduces the training of CBOW model (Fig. 1).

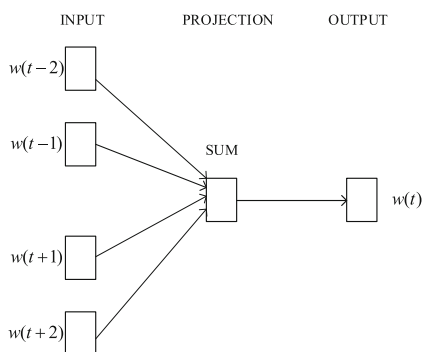


Fig. 1. The training process of the CBOW model

CBOW's network structure includes three layers, according to the data process, which are input layer, hidden layer and output layer. The objective function formula is as follows:

$$L = \sum_c \log P(w|context(w)), \quad (1)$$

Where c represents corpus, w represents center word, and $context(w)$ represents the context of w .

Input layer: contain $2c$ words vector in $context(w)$.

Projection layer: the $2c$ vectors of the input layer are summed and accumulated, which is

$$x_w = \sum_{i=1}^{2c} v(context(w_i)). \quad (2)$$

Output layer: output layer corresponds to a binary tree, word appear in the corpus as a leaf node, and weight is the occurrence number of the word in the corpus.

The output of the result needs to be a softmax normalized, and it's as follows:

$$p(w|\text{context}(w)) = \frac{e^{y_w \cdot i_w}}{\sum_{i=1}^N e^{e^{y_w \cdot i}}}. \quad (3)$$

When the neural network training is completed, you can find word vector of all words. Interestingly, when using a word vector to express a word, it can be found a similar law: king – “man” + “woman” = “queen”. It can be seen that the word vector is very conducive to the expression of the semantic features of the word.

In our approach, the feature vector of each document is the centroid of the word embeddings of the document [12]. The formula for centroid of a document M is as follows:

$$\vec{M} = \frac{1}{|M|} \sum_{i=1}^M \vec{w}_i, \quad (4)$$

where $|M|$ is the number of tokens in N and \vec{w}_i is the word vector of word w_i .

4 Experimentation

4.1 Experimental Settings

Data Set: We extracted the raw texts from IMDB movie reviews. The 50,000 reviews dataset was split evenly into 25,000 training sets and 25,000 test sets. The sample data are generally evenly distributed (25,000 positives, 25,000 negatives). It also includes an additional 50,000 unlabeled documents. We randomly selected 5% or 10% of the sample as the initial labeled sample, and remaining data as unlabeled data set. Word2vec model is built by the additional 50,000 unlabeled documents. Ten-fold cross validation as the final experimental result.

Features: We remove stop words if unigram is used, and don't remove them if bigram is used. For word2vec, we learn the vector representation of words through training 50000 unlabeled data, and then average all vectors of the words as the feature vector of each review.

Classification Algorithm: SVM algorithm has good performance in emotion classification, this paper uses SVM in sklearn package.

4.2 Experimental Results

In order to reflect the classification performance, the following algorithm and our algorithm for comparison:

Baseline: Supervised learning algorithm with the original labeled data only, in this paper, we use SVM classifier.

Self-training: Firstly, use the original labeled set to train the classifier, then use the classifier iteratively add the highest confidence sample to the labeled set.

Co-training: A co-training algorithm both using feature partition and language translation strategies.

Figure 2 shows the classification performance comparison of the various semi-supervised learning algorithm when the initial labeled set is 5% or 10% of the total training set. From the results we can see that our algorithm obtains the best classification effect, the classification accuracy rate is far better than baseline, compared with Self-training and Co-training, when the initial labeled set is 5% of the total training set, the accuracy rate of our algorithm has been increased by 5.9% and 2.8%, respectively. When the initial labeled set is 10% of the total training set, our algorithm attains the accuracy rate improvement by 6.7% and 2.2%.

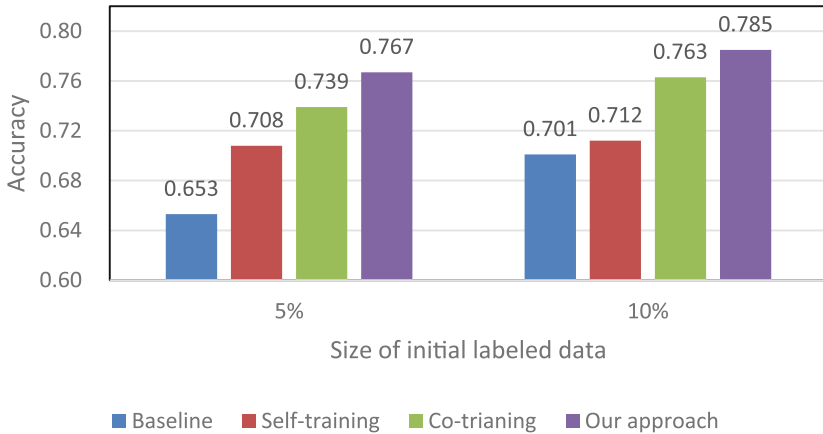


Fig. 2. Performance comparison of different algorithm

5 Conclusions

In this paper, we illustrate a novel co-training style algorithm in semi-supervised sentiment classification. The algorithm uses original training set to train three classifiers, where the text representation model for each classifier is unigram, bigram, and word2vec, respectively and then these classifiers are refined with unlabeled data. Our approach is clearly superior to self-training algorithms and innovative co-training algorithms for semi-supervised sentiment classification, according to the analysis of the experimental data.

Acknowledgments. Project 61471066 supported by NSFC.

References

1. Zhou, X., Wan, X., Xiao, J.: Cross-lingual sentiment classification with bilingual document representation learning. In: Meeting of the Association for Computational Linguistics, pp. 1403–1412 (2016)
2. Xiang, B., Zhou, L.: Improving Twitter sentiment analysis with topic-based mixture modeling and semi-supervised training. In: Meeting of the Association for Computational Linguistics, pp. 434–439 (2014)
3. Xie, S., Wang, T.: Dividing for combination: a bootstrapping sentiment classification framework for micro-blogs. In: International Conference on Information Science and Cloud Computing, pp. 78–84. IEEE (2014)
4. Kim, Y., Zhang, O.: Credibility adjusted term frequency: a supervised term weighting scheme for sentiment analysis and text classification. In: Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 79–83 (2014)
5. Xu, S., Liang, H.Z., Baldwin, T.: UNIMELB at SemEval-2016 tasks 4A and 4B: an ensemble of neural networks and a Word2Vec based model for sentiment classification. In: International Workshop on Semantic Evaluation, pp. 183–189 (2016)
6. Jeevankumar, M., Jain, P., Chetan, M.: Opinion analysis of text on the basis of three domain classification. In: International Conference on Automatic Control and Dynamic Optimization Techniques, pp. 173–177 (2016)
7. Li, S., Huang, L., Wang, J.: Semi-stacking for semi-supervised sentiment classification. In: Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing, pp. 27–31 (2015)
8. Le, T., Mikolov, T.: Distributed representations of words and phrases. In: Proceedings of the 31th International Conference on Machine Learning, Beijing, pp. 1188–1196 (2014)
9. Silva, N.F.D., Hruschka, E.R., Hruschka Jr., E.R.: Biocom Usp: tweet sentiment analysis with adaptive boosting ensemble. In: International Workshop on Semantic Evaluation, pp. 123–128 (2014)
10. Giorgis, S., Rousas, A.: A weighted ensemble of SVMs for Twitter sentiment analysis. In: Proceedings of SemEval-2016, pp. 96–99. Association for Computational Linguistics (2016)
11. Gao, W., Li, S., Lee, S.Y.M.: Joint learning on sentiment and emotion classification. In: ACM International Conference on Information & Knowledge Management, pp. 1505–1508 (2013)
12. Kusner, M.J., Sun, Y., Kolkin, N.I., Weinberger, K.Q.: From word embeddings to document distances. In: Proceedings of the 32nd International Conference on Machine Learning, vol. 37, Lille, France (2015)

GNSS Spoofing Jamming Recognition Based on Machine Learning

Pan Gao^{1,2,3}✉, Songlin Sun^{1,2,3}, Zhen Zeng^{1,2,3}, and Chenwei Wang⁴

¹ School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing, China
18957137840@163.com

² Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China

³ National Engineering Laboratory for Mobile Network Security,
Beijing University of Posts and Telecommunications, Beijing, China

⁴ DOCOMO Innovations, Inc., Palo Alto, USA

Abstract. While a vast amount of applications and services are based on the Global Navigation Satellite System (GNSS), GNSS needs to deal with jamming, and how to carry out spoofing jamming recognition is the key of achieving high accuracy performance. In this paper, we apply machine learning approaches to GNSS spoofing jamming recognition. In particular, first, we investigate the scheme by employing the classical isometric mapping (ISOMAP) and Laplacian Eigen mapping (LE) algorithm to extract intrinsic feature vector for classification recognition from the original high-dimensional data. Next, we compare this scheme to another two feature vector extraction algorithms developed from principal component analysis (PCA), wavelet transform and singular value decomposition (WT-SVD). Finally, we consider the recognition rates of the four algorithms based on the support vector machine (SVM) classifier, and the effectiveness and the robustness of our scheme are verified via simulations.

Keywords: GNSS · Jamming recognition · Machine learning
Feature extraction

1 Introduction

As the key function of a vast amount of information technologies, GNSS plays a significant role in military and commercial applications. On the other hand, it might be vulnerable to a variety of jamming attacks, especially in military services. In particular, as a major mode of jamming, spoofing jamming can deceive the GNSS receiver to produce the wrong navigation positioning and timing results, which could significantly affect GNSS [1].

In order to detect the spoofing signals and mitigate the impact of the spoofing attacks, a variety of methods have been proposed. For example, software-defined radio was introduced for GNSS jamming detection, classification and localization in [2]. In [3], a spoofing countermeasure based on the power measurements of a single rotating antenna

was proposed, which can be implemented in a static receiver. In addition, several methods were proposed for implementation in a conventional receiver with a single antenna, such as moving receiver methods, signal quality monitoring techniques, and multi-modal detection method [4, 5]. Moreover, to identify GNSS spoofing jamming by recognizing GNSS spatial signal characteristics, an improved method based on variance analysis(ANOVA) was proposed [6]. Also, a GNSS anti-spoofing method based on particle filter was proposed in [7], the feature of which is to effectively detect the spoofing attack by capturing the maximum weight of the abnormal particles.

While the prior works mainly focused on designing models and signal processing algorithms to resolve GNSS spoofing jamming recognition problems, in this paper, we take a completely different approach based on machine learning. Specifically, we employ the Hammerstein model in [8] to model a transmitter/spoofers with their wireless channels, and propose a new scheme. The feature extracted by our proposed the new scheme can effectively obtain the global or local structural features of the data space, which is helpful to improve the accuracy of recognition made by the SVM classifier.

The rest of this paper is organized as follow. The system model and the spoofing recognition method are introduced in Sect. 2. Section 3 introduces machine learning dimensionality reduction algorithm and the SVM classifier. The experimental simulation results are provided in Sect. 4. Finally, we conclude this paper in Sect. 5.

2 System Model

The structure of a typical transmitter/spoofers we consider in this paper is simplified and shown in Fig. 1.

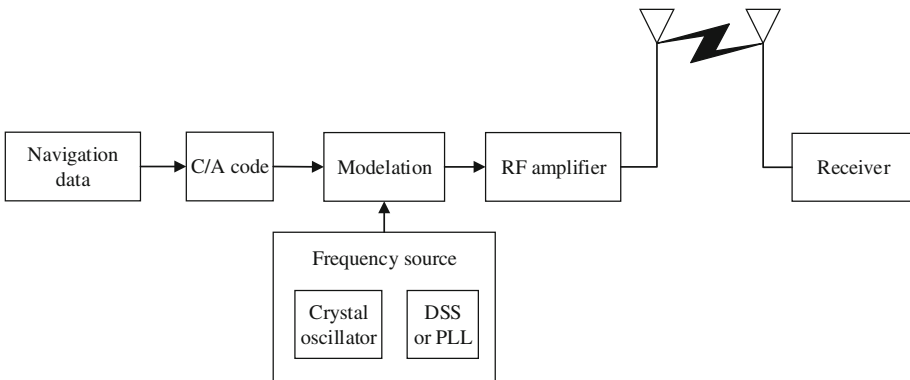


Fig. 1. Structure of a typical transmitter/spoofers

In this paper, the Hammerstein model developed in [8] is employed to simulate the spurious jamming signal, which is extremely similar to the genuine signal parameter. The transmitter/spoofers is reasonably equivalent to a static nonlinear subsystem, which can be described by a memoryless polynomial model. In addition, the wireless communication channel is modeled as a dynamic linear subsystem, which is represented by an

FIR filter. Thus, the relationship between the model input and output can be readily written as follows:

$$y(n) = \sum_{k=0}^{N-1} h_k \sum_{i=1}^M b_{2i-1} |d(n-k)|^{2i-2} d(n-k) + w(n), \quad (1)$$

where $y(n)$ is the received signal, and $d(n)$ denotes the input signal; M is the number of polynomial coefficients of the nonlinear subsystem, b_{2i-1} is the i -th weight of the polynomial coefficients, and N is the order of the FIR filter; h_k represents the channel response coefficient, and $w(n)$ is the additive white Gaussian noise and $w(n) \sim N(0, \sigma^2)$.

Figure 2 shows the main spoofing jamming recognition components of a GNSS receiver, and we will illustrate it in details later. In this paper, machine learning is used to seek a low-dimensional feature embedded in the high-dimensional data and to construct the feature sets. After that, all the feature sets are fed into a classifier for training and classification. The accuracy of classification is defined as the spoofing jamming recognition rate.

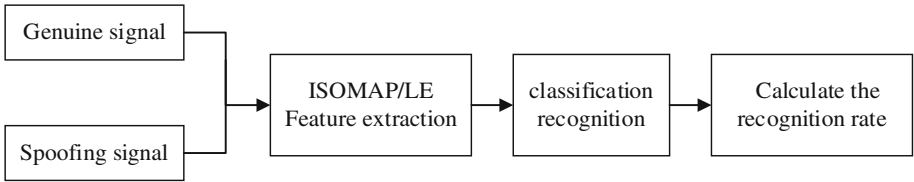


Fig. 2. Receiver spoofing jamming recognition

3 The Proposed Scheme Based on Machine Learning

3.1 Isometric Feature Mapping (ISOMAP)

Based on the multi-dimensional scale transformation (MDS) method, ISOMAP strives to maintain the intrinsic geometry of the data sets, i.e., to maintain the distance between the two points [9]. While the distance matrix constructed by MDS reflects the Euclidean distance between sample points, the distance matrix constructed by ISOMAP captures the geodesic distance between them. Therefore, the key of ISOMAP is how to calculate the geodesic distance between sample points.

The steps of ISOMAP are summarized as follows:

1. Select the neighborhood to construct the neighborhood graph G . Calculate the Euclidean distance between each sample point x_i and the remaining sample points.
2. Calculate the shortest path. When the graph G has an edge $x_i x_j$, the shortest path is $d_G(x_i, x_j) = d(x_i, x_j)$, otherwise $d_G(x_i, x_j) = \infty$.

3. Calculate d-dimensional embedding. Apply MDS method to distance matrix D_G .

$$H = -(I - 1_N 1_N^T) D_G (I - 1_N 1_N^T) / 2. \tag{2}$$

Let d be the number of top eigenvalues $\lambda_1, \dots, \lambda_d$ of H and the corresponding eigenvectors u_1, \dots, u_d consist of $U = [u_1, \dots, u_d]$, and $T = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_d^{1/2}) U^T$ is d-dimensional embedded results.

From the above steps, it can be seen that ISOMAP is a globally optimal method. It can capture the manifold distance between the high-dimensional sample points, and its computational complexity is $O(N^3)$ [9].

3.2 Laplacian Eigen (LE)

The LE mapping algorithm has a very intuitive dimensionality reduction goal. Specifically, if any two sample points are very close to one another in the high dimensional space, then the corresponding two points after dimension reduction are also very close in the low dimensional space [9]. That is, when the sample point x_i and x_j are very close, LE uses a positive weight to contact the two sample points. Usually, the values of these weights can be set to $W_{ji} = 1$, but they can also be set by the attenuation function:

$$w_{ij} = \exp(-\|x_i - x_j\|^2 / \sigma^2), \tag{3}$$

where σ^2 is a proportional parameter. Suppose the diagonal elements $D_{ii} = \sum_j w_{ij}$ of diagonal matrix is represented by D , and then the low dimensional coordinate τ_i of LE is used to minimize the sum of functions:

$$E(T) = \sum_{ij} \frac{w_{ij} \|\tau_i - \tau_j\|^2}{\sqrt{D_{ii} D_{jj}}}. \tag{4}$$

In addition, the unique solution can be obtained by adding centralization and standardization restrictions to T . Since the sum function can be written

$$E(T) = \text{Tr}(T(I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}) T), \tag{5}$$

we can solve the smallest d -eigenvalues of the matrix $\Phi = I - D^{(-1/2)} W D^{(-1/2)}$ and $T = [u_2, \dots, u_{d+1}]$ is the calculated embedded result.

As mentioned in [9], the LE requires only a small amount of computation, and the computational complexity is $O(dN^2)$.

3.3 Support Vector Machine (SVM)

In machine learning, SVM is a supervised learning model, which is usually used for pattern recognition, classification and regression analysis [10]. By introducing a dummy variable into each attribute of the data, SVM can be applied to data classification. SVM is a type of second-class classification models and its basic model is defined as the linear classifier with the largest interval in the feature space. That is, the learning strategy of SVM is the interval maximization, which can be transformed into a convex quadratic programming problem. Moreover, the Gauss kernel is introduced by SVM to learn nonlinear SVM, which translates the nonlinear problem into a linear separable one. In this paper, the Gaussian kernel we consider is given by:

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right). \tag{6}$$

The corresponding SVM is a Gaussian radial basis function classifier. In this case, the SVM classification decision function is:

$$f(x) = \text{sign} \left\{ \sum_{i=1}^{N_i} a_i^* y_i \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) + b^* \right\}. \tag{7}$$

3.4 Recognition Process

After completing the feature extraction, we need to select and design the classifier to classify the feature vector. Taking into account the actual anti-jamming process, we need to ensure that testing and classification of the classifier are efficient, adaptive, and not easy to fall into the local minimum regimes, which results in recognition error in the experimental simulation. Based on the results made by numerous papers in classification, we finally choose SVM for classification.

Collecting all the four steps above, we establish a spoofing jamming recognition model flow chart, as shown in Fig. 3, where the feature extraction module is to first seek a low-dimensional feature embedded in the high-dimensional data, and then the feature

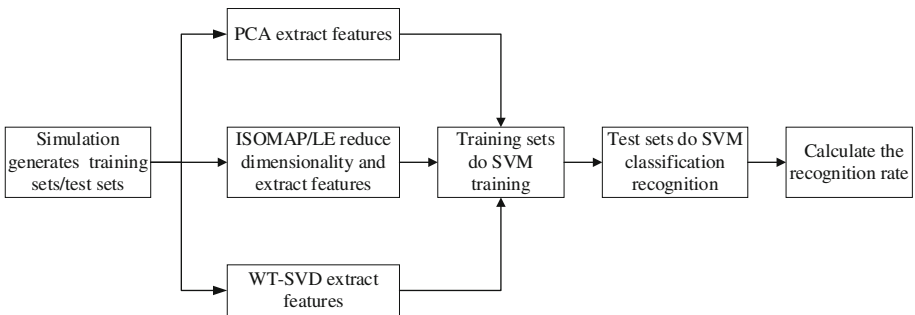


Fig. 3. GNSS spoofing jamming recognition block diagram

sets are constructed. Finally, all the feature sets are fed into the SVM classifier for training and classification.

4 Simulation Results

In order to verify the effect of machine learning on the recognition of spoofing jamming, we assume that jamming is spoofing jamming and the GNSS C/A code signal is generated by simulation based on the Hammerstein model, and the signal modulation is QPSK. In addition, we assure that the SNR is in the range of [1 dB, 10 dB]. Also, we assume that 1000 training sets (500 sets of genuine signals and 500 sets of spoofing signals, respectively) and 1000 testing sets are independently generated by simulation, each data with 500 dimensions.

The parameters of ISOMAP and LE algorithm are chosen to be $k = 4$ and $d = 2$. Meanwhile, in order to comprehensively investigate the recognition effect of machine learning algorithm, we in this section also consider the traditional PCA algorithm and the WT-SVD algorithm by simulation. We state our results in the following.

First, the time complexity of each algorithm is illustrated in Table 1, which shows the average runtime for each SNR value with 1000 training sets.

Table 1. Average runtime of four algorithms

Algorithms	LE	ISOMAP	WT-SVD	PCA
Average runtime/s	3.6977	86.1822	31.7243	0.9879

From Table 1, it can be seen that the runtime of LE is much faster than ISOMAP, because ISOMAP needs to use the Dijkstra algorithm to calculate the shortest path. Moreover, PCA is the fastest; WT-SVD is faster than ISOMAP, but slower than LE.

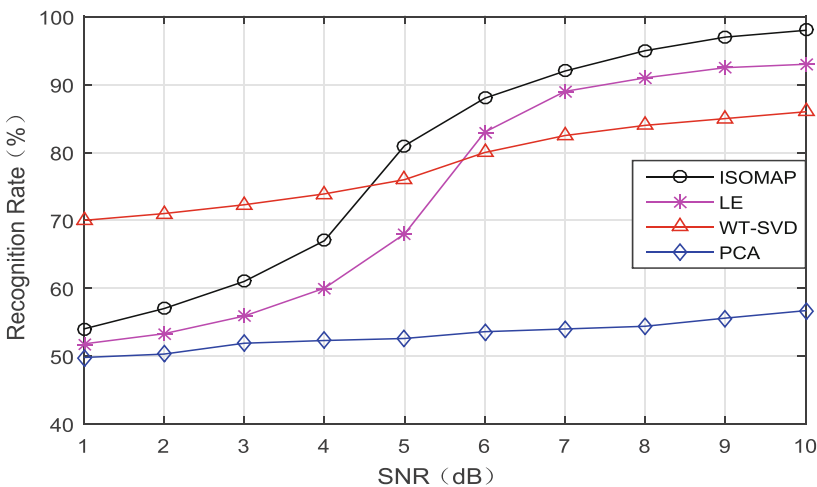


Fig. 4. The recognition rate varies with the SNR

The classification accuracy of the four algorithms introduced in Sect. 3 are shown in Fig. 4. The recognition rate of each algorithm is approximately about 50% around $\text{SNR} = 1$ dB except for WT-SVD. As expected, the average recognition rates of ISOMAP, LE and WT-SVD algorithms have been greatly improved when the SNR increases. Moreover, the average recognition rates of ISOMAP and LE exceed WT-SVD when $\text{SNR} = 6$ dB. Additionally, the average recognition performance of ISOMAP has always been higher than LE because ISOMAP is a globally optimal method and LE is local optimal. In the classification recognition, feature extraction is extremely essential. However, the PCA method poorly extracts the characteristics of non-linear data, which leads to its low recognition rate and no more than 60%.

Overall, the experimental simulation shows that two nonlinear dimensionality reduction algorithms in machine learning are superior to the traditional WT-SVD and PCA algorithms in feature extraction and recognition performance based on the SVM classifier. Moreover, our simulation results illustrate that the quality of feature selection determines the recognition accuracy.

5 Conclusions

In order to improve the recognition rate of GNSS spoofing, we apply machine learning to GNSS spoofing jamming recognition. In particular, we analyze several nonlinear dimensionality reduction algorithms in feature extraction and recognition performance based on the SVM classifier. In future, we would like to explore algorithm design and data analysis for more practical scenario.

Acknowledgment. This work is supported by National Natural Science Foundation of China (No. 61471066).

References

1. Jeong, S., Kim, T., Kim, J.: Spoofing detection test of GPS signal interference mitigation equipment. In: International Conference on ICT Convergence 2014, pp. 651–625 (2014)
2. Bartl, S., Berglez, P., Hofmann-Wellenhof, B.: GNSS interference detection, classification and localization using software-defined radio. In: IEEE European Navigation Conference (ENC), pp. 159–169 (2017)
3. Fei, W., Li, H., Lu, M.: GNSS spoofing countermeasure with a single rotating antenna. *IEEE Access* **5**, 8039–8047 (2017)
4. Li, J., Zhang, J., Chang, S., Zhou, M.: Performance evaluation of multimodal detection method for GNSS intermediate spoofing. *IEEE Access* **4**, 9459–9468 (2016)
5. Kerns, A.J., Wesson, K.D., Humphreys, T.E.: A blueprint for civil GPS navigation message authentication. In: IEEE Position Location and Navigation Symposium, PLANS, pp. 262–269 (2014)
6. Yuan, D., Li, H., Lu, M.: A method for GNSS spoofing detection based on sequential probability ratio test. In: Proceedings IEEE/ION PLANS 2014, Monterey, CA, pp. 351–358 (2014)

7. Borio, D.: PANOVA tests and their application to GNSS spoofing detection. *IEEE Trans. Aerosp. Electron. Syst.* **49**(1), 381–394 (2013)
8. Hong, X., Chen, S., Gong, Y., Harris, C.J.: Nonlinear equalization of Hammerstein OFDM systems. *IEEE Trans. Signal Process.* **62**(21), 5629–5639 (2014)
9. Ye, D., Desjardins, B., Hamm, J., et al.: Regional manifold learning for disease classification. *IEEE Trans. Med. Imaging* **33**(6), 1236–1247 (2014)
10. Shrutkirti, D.K., Puranik, V.G.: Remote detection of photoplethysmographic signal and SVM based classification. In: *IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)*, pp. 128–132 (2016)

TinyPEP: Tiny Pairwise-Key Establishment Protocol for Wireless Sensor Networks

Wei Liu^{1(✉)} and Rong Luo²

¹ Chongqing University of Technology, Chongqing 400054, China
liu-wei@cqut.edu.cn

² Tsinghua University, Beijing 100084, China

Abstract. Setup time is one of the most critical factors for transitory initial key based pairwise key establishment protocols in wireless sensor networks. In this paper, we propose TinyPEP, which greatly reduces the setup time of key establishment by removing unnecessary information exchanges and introducing backoff mechanism directly. TinyPEP also provides a compensation scheme for unconnected nodes. Setup time and totally connected probability are theoretically analyzed and experimentally simulated. The results show that, by choosing parameters carefully, the proposed protocol is scalable for different network densities. When the size of backoff window is 8192 slots, the setup time is less than 5.2 s and the totally connected probability is larger than 97% for typical network densities.

Keywords: Security · Pairwise key establishment · Wireless sensor networks

1 Introduction

Wireless sensor networks (WSNs) always need to be deployed in open environments. Moreover, the sensed information is generally sensitive and private. As a result, security becomes an important factor for WSNs. Confidentiality and authenticity are basic requirements for most applications. Due to resource constrains, existing security schemes typically use symmetric cryptosystems to implement these two functionalities. Symmetric cryptosystems use identical pairwise keys between the communication pair, so efficient mechanisms are needed to establish these keys.

There have been many protocols proposed for pairwise key establishment in WSNs. Transitory initial key based protocols are important members of them. However, none of them has been widely accepted as a standard, either because the complexity is unaffordable by resource constrained sensor nodes or because the security is defective which can be utilized easily by adversaries. In this paper, a transitory initial key based pairwise key establishment protocol, TinyPEP, is proposed. The contributions of this work include: (1) Setup time is significantly reduced by removing unnecessary information exchanges and introducing backoff mechanism directly; (2) A compensation scheme with acceptable overheads is proposed for unconnected nodes to reestablish pairwise keys with original neighbors even after time limit has expired.

The rest of this paper is organized as follows. In Sect. 2, related works are discussed. After giving the scenario assumptions and threat model in Sect. 3, we describe the proposed protocol briefly in Sect. 4. The compensation scheme for unconnected nodes is also presented in this section. In Sect. 5, setup time and totally connected probability are theoretically analyzed and simulation results are presented in Sect. 6. We conclude this paper in Sect. 7.

2 Related Works

Several key establishment protocols have been proposed over the last years, which can be classified into four classes: Trustable Node Assistance (TNA) protocols, Random Key Pre-distribution (RKP) protocols, Plaintext Key Exchange (PKE) protocols and Transitory Initial Key (TIK) based protocols.

TNA protocols use auxiliary nodes to establish pairwise keys. [1] proposed to use base stations for this purpose. This protocol is very efficient for small scale networks. However, it isn't scalable well with network size. [2] improved this scheme with many auxiliary nodes scattered in the network randomly. Nodes establish pairwise keys with the assistance of auxiliary nodes within several hops. There is an obvious vulnerability for these protocols. If the adversary has compromised an auxiliary node, it is able to establish pairwise keys with any node at any location in the network.

[3] presented the idea of RKP. In this scheme, each node is preloaded with a number of keys randomly chosen from a large pool. In order to achieve adequate probability for each node to establish secure links with its neighbors, the number of keys preloaded has to be very large. Moreover, in order to find out shared keys between neighbors, large number of messages have to be exchanged. Several works extended this scheme or gave more detailed analysis, and deployment knowledge was introduced to improve it. However, essential weaknesses still present. As compromising a node reveals many legitimate keys, the adversary can replicate arbitrary number of malicious nodes to join into the network.

[4] presented the idea of PKE. This scheme is based on the observation that an adversary can only appear at certain location at a given time. Nodes establish pairwise keys by generating keys randomly and exchanging them in plaintext. This scheme needs no keys preloaded and the process is absolutely distributed. However, malicious nodes are able to join into the network arbitrarily, even without compromising any nodes.

LEAP [5] presented the idea of TIK firstly, and increased its security level in LEAP+ [6]. In these schemes, each node is preloaded with identical initial key which will be cleaned up after a time limit. Nodes establish pairwise keys using the initial key after deployment. OTMK [7] introduced randomness into TIK based schemes to restrict the impact of initial key compromise in local area. However, setup time of key establishment is too long in LEAP+ and OTMK, which increases the probability of initial key compromise. If the time limit is decreased for security, connected probability will be inevitably affected. Unfortunately, there are no compensation schemes for unconnected nodes in these schemes.

3 Scenario Assumptions and Threat Model

In most sensor network applications, arbitrary point to point communication is unnecessary. General patterns include communications among direct neighbors and between sink (or base station) and nodes. The keys shared between sink and nodes are always preloaded before deployment. Consequently, establishing pairwise keys for direct neighbors is enough. When needed, other keys can be constructed after pairwise keys have been established [6]. Therefore, our protocol only focuses on establishing pairwise keys for direct neighbors. In this paper, we only consider single hop network. Performance of TinyPEP in multihop network is left for future work.

In this paper, we assume that adversaries possess powerful capabilities to eavesdrop, inject or tamper with packets. However, these behaviors are localized because an adversary can only appear at certain place at a given time. The number of adversaries is much smaller compared with the network size, so adversaries are unable to eavesdrop on the entire network. The adversary is able to compromise a node and obtain all secret information definitely. However, it has to take certain amount of time to execute these operations [7]. The adversary is able to compromise a small portion of sensor nodes in large scale sensor networks within a short period of time. However, it cannot compromise all nodes within such a short time.

4 TinyPEP: Tiny Pairwise-Key Establishment Protocol

4.1 The Key Establishment Process

In TinyPEP, the key establishment process can be divided into four steps, which are Keying Materials Preload, Neighbor Discovery, Pairwise Key Generation, and Initial Key Cleanup respectively [8].

Keying Materials Preload. Before network deployment, a trustable node generates an initial key K_0 randomly. It assigns unique address i for each sensor node randomly. Then, the trustable node generates $K_{i,B}$ for sensor node i randomly. With K_0 and i , the trustable node computes identification key for node i as follows:

$$K_i = H(K_0 || i || K_0) \quad (1)$$

The trustable node preloads $(K_0, K_{i,B}, K_i)$ into sensor node i and $K_{i,B}$ into base station. K_0 is used for key establishment, $K_{i,B}$ is used for secure communication between node i and base station, and K_i is used for authenticating i 's identity.

Neighbor Discovery. After deployment, the key establishment process starts. Each node generates a random number $nonce_i$ to introduce discrepancy spatially and chooses its backoff time randomly for identity message broadcast. Then, each node senses the channel for potential broadcasts from its neighbors. If a node finds that the channel is clear at all times during its backoff, it will broadcast its own identity message $illnonce_i$. If a node finds that certain neighbor begins to broadcast during its backoff, it will receive the broadcast. After receiving, it computes pairwise key with this neighbor and records corresponding address-key pair into its neighbor table. Then it chooses backoff time

again. Assume that there are no collisions during this phase, all neighbors will receive the identity message broadcasted by the node that ends its backoff firstly. Repeat this process, until reach the time limit T_{limit} .

Pairwise Key Generation. Assume that node j has received i 's identity message, it computes pairwise key with i as follows:

$$K_{j,i} = H(K_0, f(j, i), K_0) \tag{2}$$

where f represents a function which can generate the same pairwise key while only receiving the other's identity message

Initial Key Cleanup. If time limit T_{limit} expires, no matter whether there are nodes without broadcasting their identity messages or not, the protocol has to be finished to avoid potential initial key compromise. Each node cleans up its initial key and transmits its random number to the base station which is encrypted and authenticated with $K_{i,B}$, then enter into normal application phase.

4.2 Compensation Scheme for Unconnected Nodes

In the process of pairwise key establishment, if more than one node ends backoff firstly at the same time, their identity messages will collide. Then, no neighbors can receive their messages successfully. As a result, neighbors can't generate pairwise keys with them. They are totally unconnected from the network, even though they can receive some identity messages from other nodes.

There are two solutions for this totally unconnected problem. First, if a node which has broadcasted its identity message finds that there have been no broadcasts in several backoff phases and time limit T_{limit} hasn't expired, it will attempt to broadcast its identity message again in the next backoff phase. In other words, if the setup time is much shorter than T_{limit} , key establishment process can repeat several times to improve totally connected probability.

Second, if time limit T_{limit} has expired and node i finds that it is unable to communicate with any neighbor, then it can conclude that its identity message has collided with others. In this case, as node i has pairwise keys with some neighbors, which can be used as proof of its legality in this area, it will transmit special message to base station for assistance. The format of this message is as shown in Fig. 1, where m, n and p represents neighbor addresses from which node i receives identity messages.

i	$Addr(m, n, p \dots)$	$K_{i,m}(K_{i,n}(K_{i,p}(\dots(nonce_i))))$
-----	-----------------------	---

Fig. 1. Format of compensation message

The message is encrypted and authenticated with $K_{i,B}$. How many neighbors and pairwise keys are included in the message depends on the number of identity messages received by i and the desired connectivity degree. In order to reduce communication overheads, the random number are repeatedly encrypted using pairwise keys. After

receiving this message, relay nodes route it to base station without checking. When the base station receives this message, it will compute corresponding pairwise keys using K_0 and random numbers of declared nodes in the message. Then, the base station checks the validity of the repeatedly encrypted random number. If pass, the base station will inform the declared nodes with corresponding pairwise keys. These messages are encrypted and authenticated with keys between base station and the declared nodes. After receiving this information, the declared nodes add node i to their neighbor tables and inform i that connection is available.

5 Mathematical Analysis

5.1 Supported Number of Nodes

Upper bound and lower bound of supported number of nodes can be analyzed theoretically according to the key establishment process described above. We only give final results here and detailed derivation process can be find in [8].

$$\frac{T_{limit}}{(m-1)T_{uint} + T_T + T_C} < n < \frac{T_{limit}}{T_T + T_C} \text{ or } \frac{T_{min}}{(m-1)T_{uint} + T_T} < n < \frac{T_{limit}}{T_T} \quad (3)$$

where T_T denotes transmit time of identity message, T_C denotes time to compute pairwise key, m denotes the number of slots in backoff window.

5.2 Totally Connected Probability

In this paper, totally connected means that every node establishes bidirectional pairwise keys with all its neighbors. In order to guarantee total connection, there must be no collisions in each backoff phase. So, the totally connected probability can be expressed as follows [8]:

$$P = P(k)P(k-1) \cdots P(2)P(1) \quad (4)$$

where $P(k)$ represents the probability with which there are no collisions when k nodes left haven't broadcasted their identity messages. $P(k)$ can be expressed as follows:

$$P(k) = k \sum_{i=1}^{m-1} (m-i)^{k-1} / m^k \quad (5)$$

6 Experimental Results

6.1 Explanations for Protocol Parameters

In order to be comparable with analogous schemes, such as LEAP+ and OTMK, we choose Mica2 [9] as our target platform. Mica2 node is equipped with an ATmega128 microcontroller, which has a timer working at 32 kHz. In TinyPEP, node address and random number are 2 bytes and 4 bytes respectively, so length of the identity message is 16 bytes, which also includes 8 bytes MAC header and 2 bytes CRC. Because the transmission rate of Mica2 is 19.2 Kbps, the time to transmit/receive identity message is 6.667 ms. MD5 is used as the one-way secure hash function, and the key size is chosen as 128bit. It needs less than 2 ms to compute pairwise keys in Mica2. Consequently, pairwise key computation can be delayed to the next backoff phase when using Mica2 nodes.

6.2 Connected Probability

In TinyPEP, connected probability only depends on whether there are collisions and how many collisions occur. We simulated the totally connected probability and the probability with which no more than two nodes collided. The results are illustrated in Fig. 2. It is obvious that the connected probability is tightly coupled with the size of backoff window. If the backoff window is large enough, totally connected probability can exceed 97% for typical network densities. The size of backoff window should be determined by maximal network density possibly appeared in practical applications. It is very clear that the probability with which more than two nodes have their identity messages collided is considerably low. This value can be lower than 0.05% with backoff window of 8192 slots for typical network densities. Then we can conclude reasonably that the compensation scheme for unconnected nodes in TinyPEP doesn't consume much energy as only little nodes need to invoke the compensation scheme.

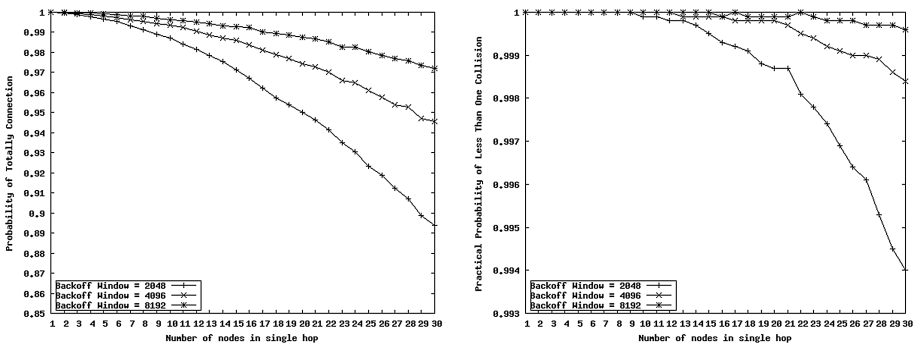


Fig. 2. Probability of totally connected (left) and no more than two unconnected nodes (right)

6.3 Setup Time Estimation

In order to guarantee adequate connected probability, backoff window has to be large enough. This leads to longer setup time inevitably. Thus, it is necessary to determine the supported number of nodes under reasonable sizes of backoff window and different time limits. Figure 3 gives the maximal and minimal supported number of nodes theoretically. However, these values can only reflect possible range of supported number of nodes. Practically supported number of nodes locates in this range. In order to find the practical values, we simulated the practical probability of finishing key establishment in given time limit with backoff window of 8192 slots. The simulation results are also given in Fig. 3. It is obvious that the practical supported number of nodes is larger than the theoretical minimum.

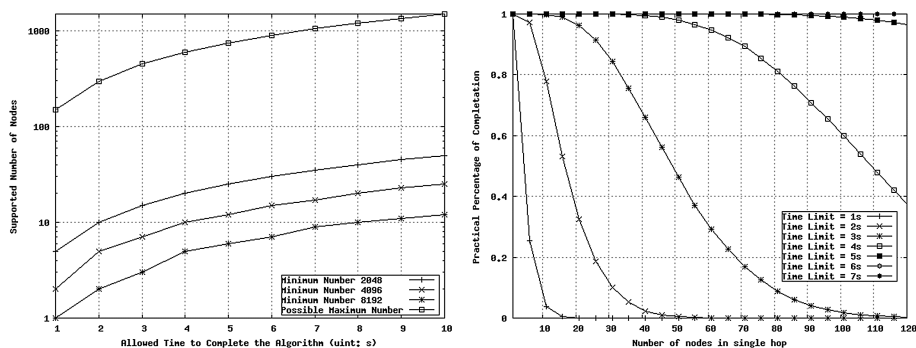


Fig. 3. Theoretically (left) and practically (right) supported number of nodes

To compare the setup time with other TIK based schemes, practical setup time for typical network densities is simulated. The maximal values for different network densities are plotted in Fig. 4. For typical network densities, the maximal setup time is less than 5.2 s, which has been improved significantly compared with LEAP+ and OTMK. LEAP+ needs 170 s to finish for 20 nodes per single hop, while OTMK needs 100 s to finish for 10 nodes per single hop. Even the simplified OTMK also needs 20 s for 10 nodes per single hop.

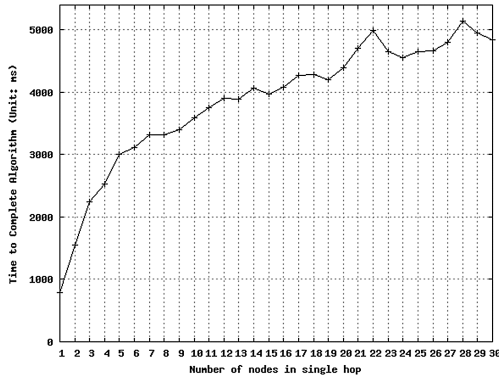


Fig. 4. Practical setup time for typical network densities

7 Conclusions

Pairwise key establishment is a critical security primitive for wireless sensor networks. There have been many protocols proposed in this domain. However, performance and security of existing protocols are not satisfactory. In this paper, we propose TinyPEP, a tiny pairwise key establishment protocol, to address these problems. Using TinyPEP, setup time is significantly decreased by removing unnecessary information exchanges and introducing backoff mechanism directly into key establishment. A compensation scheme with acceptable overheads is proposed for unconnected nodes to reestablish pairwise keys with original neighbors even after the time limit has expired. Simulation results show that, by choosing parameters carefully, the proposed protocol is scalable for different network density. The setup time is less than 5.2 s and the totally connected probability is larger than 97% for typical network density, which outperforms other transitory initial key based protocols, such as LEAP, LEAP+ and OTMK.

Acknowledgements. This research was supported in part by China NSFC Grant 61601069 and Scientific and Technological Research Program of Chongqing Municipal Education Commission KJ1600935.

References

1. Perrig, A., Szewczyk, R., Wen, V., et al.: SPINS: security protocols for sensor networks. In: Proceedings of ACM MobiCom, pp. 189–199 (2001)
2. Dong, Q., Liu, D.: Using auxiliary sensors for pairwise key establishment in WSN. *ACM Trans. Embed. Comput. Syst.* **11**(3), 59:1–59:31 (2012)
3. Eschenauer, L., Gligor, V.: A key-management scheme for distributed sensor networks. In: Proceedings of ACM Conference on Computer and Communications Security, pp. 41–47 (2002)
4. Anderson, R., Chan, H., Perrig, A.: Key infection: smart trust for smart dust. In: Proceedings of IEEE Conference on Network Protocols, pp. 206–215 (2004)

5. Zhu, S., Setia, S., Jajodia, S.: Leap: efficient security mechanisms for large-scale distributed sensor networks. In: Proceedings of ACM Conference on Computer and Communications Security, pp. 62–72 (2003)
6. Zhu, S., Setia, S., Jajodia, S.: LEAP+: efficient security mechanisms for large-scale distributed sensor networks. *ACM Trans. Sens. Netw.* **2**(4), 500–528 (2006)
7. Deng, J., Hartung, C., Han, R., et al.: A practical study of transitory master key establishment for wireless sensor networks. In: Proceeding of IEEE Conference on Security and Privacy for Emerging Areas in Communication Networks, pp. 289–299 (2005)
8. Liu, W., Luo, R., Yang, H.: A lightweight key establishment protocol for wireless sensor networks. *J. Electron. Inf. Technol.* **32**(4), 869–874 (2010)
9. Mica2 Datasheet: <http://www.memsic.com>. Accessed 12 July 2017
10. Newell, A., Yao, H., Ryker, A., et al.: Node-capture resilient key establishment in sensor networks: design space and new protocols. *ACM Comput. Surv.* **47**(2), 24:1–24:34 (2014)
11. Porambage, P., Braeken, A., Schmitt, C., et al.: Group key establishment for secure multicasting in IoT-enabled wireless sensor networks. In: Proceedings of IEEE LCN, pp. 482–485 (2015)
12. Harn, L., Hsu, C.F.: Predistribution scheme for establishing group keys in wireless sensor networks. *IEEE Sens. J.* **15**(9), 5103–5108 (2015)

An Innovative Indoor Location Algorithm Based on Supervised Learning and WIFI Fingerprint Classification

Cong Chao^(✉) and Men Xiaoran

School of Electrical and Electronic Engineering,
Chongqing University of Technology, Chongqing 400054, China
congchao@cqut.edu.cn

Abstract. By studying the characteristics of WIFI fingerprint signals and combining supervised learning methods in machine learning, an innovative indoor location algorithm based on Naïve Bayes and WIFI fingerprinting is presented. In the experiment, the router is selected as the generator of WIFI signal, and the RSSI fingerprint of the signal is collected to form the fingerprint library. The Naive Bayes models are used to train the data, and the server is used to calculate the position in order to realize the fast positioning of the intelligent terminal. Experiment is designed with an indoor environment including 6 positioning points, scanning interval is set to 5 s, and the learning time is set to 10 min. The experiment result shows that the system and algorithm perform well and the accuracy of positioning is higher than 80%.

Keywords: Indoor positioning · WIFI fingerprint · Supervised learning Classification

1 Introduction

Indoor positioning mainly refers to the indoor environment monitoring via wireless communication, inertial, base station, etc. In order to meet the needs of various indoor environments and indoor structures, there are some technical fields aiming at indoor positioning technology. Among all of the indoor positioning techniques, indoor positioning with WIFI has become an important branch of indoor positioning in various solutions of indoor positioning.

At present, there are three kinds of positioning technology based on WIFI in the market. First is the trilateration localization algorithm [1] with received signal strength measurement. This method estimates the distance between the measured point and the known points by calculating the signal intensity and the known signal attenuation model. Second is fingerprint location method [2, 3] based on received signal strength. The technique compares the signal strength of the real-time detection with the signal intensity characteristics of each position point in the fingerprint library, and matches the best reference point as the point to be detected. The third is signal time measuring method based on time of signal flight between two nodes in a round-trip [4]. The time consuming is used to calculate the distance, and circles are drawn for the representing

distance values between location points, and overlapping part of the circle is the measuring position.

This paper focuses on the second method, which is fingerprint localization algorithm based on signal strength. The algorithm discusses the acquisition and analysis of WIFI fingerprint feature, uses supervised learning algorithm to classify WIFI fingerprint information with position tag through the application of machine learning, improves the accuracy and timeliness of positioning.

2 Background Technology

2.1 WIFI Technology

WIFI is an important part of WLAN. It connects electronic devices such as mobile phones and computers to a wireless LAN. The emergence of WIFI is to improve the interconnectivity between wireless network products based on the IEEE 802.11 standard [5].

Through the introduction of the analysis of the current indoor positioning technology, WIFI technology, which is regarded as a branch of indoor positioning technology, has unique advantages [5]. Firstly, the WIFI technology itself has a great advantage of wide transmission range with hundreds of meters, which means in a certain area, lesser routers can be arranged for enough number of fingerprint acquisition; secondly, the provisions of WIFI IEEE 802.11 by the standard unified, can realize the interconnection between different types of WIFI generators, which means a good generalization can be achieved between intelligent device, wireless routers and mobile phones; thirdly, the transmission rate of WIFI is very fast, sometimes can reach 600 Mbps, which means the requirement of fast indoor positioning; finally, WIFI has advantages of high penetration rate, low cost, and very good popularity rate, which leaving room for fast growth.

2.2 WIFI Fingerprint Algorithm

WIFI fingerprint algorithm uses the Received Signal Strength Indication (RSSI) of WIFI signal omit from Access Points (APs) to detect indoor position. When AP positions are settled, the RSSI from AP will change by physical location, which composes the unique identified location information in this position, similar to the human fingerprint information [6]. Based on this feature, WIFI fingerprint algorithm builds a prior fingerprint library for different locations, and then use it for real-time localization.

The algorithm can be divided into two stages: offline training stage and online positioning stage (Fig. 1).

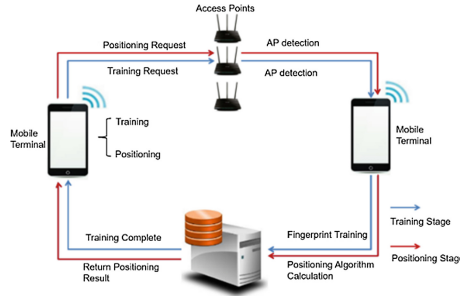


Fig. 1. Indoor positioning process

Offline Training Stage

A number of sample points are selected, where the WIFI RSSI information are recorded with position labeling. Thus, a collection of RSSI and a corresponding position label can be established as a fingerprint F_i :

$$F_i = \begin{bmatrix} MAC_1, R_1 \\ MAC_2, R_2 \\ MAC_3, R_3 \\ \vdots \\ MAC_n, R_n \end{bmatrix} \tag{1}$$

Where n means the number of WIFI RSSI can be measured in sample position, MAC_n means the MAC address, R_n represents RSSI value. In general, the more WIFI RSSIs can be measured, the better the accuracy of positioning is.

Fingerprints (F_i) from all sample positions must be collected in offline stage to establish a database of fingerprint.

Online Positioning Stage

The fingerprint E is formed similarly by obtaining the WIFI fingerprint of the current position:

$$E = \begin{bmatrix} mac_1, r_1 \\ mac_2, r_2 \\ mac_3, r_3 \\ \vdots \\ mac_n, r_n \end{bmatrix} \tag{2}$$

The online detected fingerprint E is referred to offline database of all F fingerprint information for comparison by classification or matching algorithms, to find the best matching fingerprint in database, and returns the corresponding position (or coordinates) of the fingerprint.

In the process of training and matching, the traditional method is to use distance calculation method to calculate fingerprint similarity [7, 8]. In this paper, an innovative

algorithm based on supervised classification method in machine learning, Naïve Bayes algorithm is applied to improve the precision and accuracy of WIFI fingerprint matching and positioning.

2.3 Naïve Bayes Algorithm

The Naïve Bayes classifier is a Bayes algorithm based probability classifier, which is applied to a series of strong (naïve) independent hypothetical features in machine learning. Bayes algorithm is the use of the Bayes formula for the calculation of posterior probability under the condition of the object with prior probability, the calculation results in the selection of the corresponding category with maximum probability as the category of the object. The general formula of Bayes theorem is as follows:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad (3)$$

Where, $A_1, A_2, A_3, \dots, A_n$ represent all A events, and $P(A_i)$ is the prior probability of events. $P(B|A_i)$ is the conditional probability of the event A_i after the occurrence of the event B, and $P(A_i|B)$ is the conditional probability of the A_i after the event B is known.

Naïve Bayes is an effective method for constructing classifiers for WIFI fingerprints. The model analyzes actual situations, and then assign the class labels with the eigenvalues to the situation. This is not a single classifier training algorithm, instead of using a series of algorithms based on the same principle, Naïve Bayesian classifier assumes that all features between samples is irrelevant, although some characteristics are interdependent or determined by other features, but the classifier will be considered in determining when these features are independent in probability distribution.

3 Supervised WIFI Fingerprint Classification

3.1 Training Data Definition

As described in Sect. 2.2, WIFI fingerprint includes MAC address and RSSI signal intensity values, both of which are measured during positioning. Algorithm calculates the matching level or relation between RSSI & MAC pair and location of points in training stage, and encapsulates the content in collection as fingerprints. Therefore, modeling the training data is very important by constructing a classifier to train the classifier fingerprint.

In the experiment, the method of WIFI positioning is used. The more WIFI detection, the more the fingerprint acquired, and the higher accurate the location will be detected after machine learning. This experiment will not only use the WIFI information from fixed AP, but also use the WIFI information which can be detected around the indoor environment (such as AP from different floor, AP from telecommunication businesses, etc.). The machine learning method will be used to train the database, and the positioning results will be more accurate.

3.2 Bayesian Model for WIFI Fingerprint Classification

Based on the theory of Naïve Bayes and the actual situation of the experiment, a classification model is established. In theory, it is a conditional probability model. According to Bayes theorem, there are:

$$P(C = c_i|F_1, \dots, F_n) = \frac{P(C)P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)} \tag{4}$$

Where C stands for positional class variables (c_1, c_2, \dots, c_n) , the probability of occurrence of c_i at each location is independent of each other; $F_1, F_2, F_3 \dots F_n$ represents the fingerprints that are measured at each reference location when building a fingerprint library.

Because the denominator of Formula (4) only depends on the characteristics of the variable F_n , the general characteristic variables is given, denominator can be directly regarded as a constant; and the numerator can be regarded as a joint distribution model of $P(C, F_1, F_2, F_3 \dots F_n)$. Then Formula (4) can be written in the form of conditional probability. According to Bayesian theory, if the assumption of conditional independence is true, variables F_i are independent from each other, if i is not equal to j . At this point, the conditional distribution expression of the positional class variable C is represented as:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= \frac{1}{Z} P(C)P(F_1|C)P(F_2|C)P(F_3|C)\dots \\ &= \frac{1}{Z} P(C) \prod_{i=1}^n P(F_i|C) \end{aligned} \tag{5}$$

Where Z is a scaling factor which is dependent only with $F_1, F_2, F_3 \dots F_n$. Thus, the characteristic variable is a known constant, and decomposition into prior probability P(C) and conditional probability $P(F_i|C)$ can greatly improve the reliability of the model.

The classifier can be constructed from the probability model, which contains the model and the corresponding decision rules, given that the fingerprint acquired from online stage is E, there is:

$$P(E|C = c_i) = \prod_{j=1}^n P(F_j = E|C = c_i) \tag{6}$$

With a posteriori probability criterion, the classifier formula is defined as follows:

$$\text{classify}(E) = \text{argmax}\{P(C = c_i) \prod_{j=1}^n P(F_j = E|C = c_i)\} \tag{7}$$

The posterior probability calculated by Bayes formula is the probability that the object belongs to this category, and the greater the probability value of the class, the more likely the object belongs to this class. This process is fast, the more data use for training, the longer the study time is, and the higher the positioning accurate can reach.

3.3 Algorithm Flow

Based on above discussion, the workflow for classification algorithm is as follows:

- (1) deploy adequate WIFI APs in indoor environment, and make sure that the WIFI signal is evenly distributed in all locations;
- (2) in learning phase, define and go through all sample positions in indoor environment, then collect WIFI information to form fingerprint F_i and position labels of the sample position. Send the information to server for supervised classification;
- (3) calculate probabilities $P(C_1|F_c), P(C_2|F_c), \dots P(C_n|F_c)$ according to Formula (6);
- (4) find the maximum $P(F_j = E|C = c_i)$ of the computational probability in step (3), that is, the corresponding coordinate F_c of the fingerprint E , and the E is the coordinate of the undetermined position.

4 Experimental Design and Result Analysis

4.1 Experiment Scheme

The whole positioning system is composed of three parts, which are the field AP, the server and the mobile terminal.

Field AP can arbitrarily select devices with WiFi signal emission, such as routers, portable Wi-Fi, Raspberry, Pis and so on. Ideally, these modules need to be distributed more evenly in the indoor scene, so that enough WiFi fingerprints can be obtained at different locations.

The server is mainly used with supervised classification algorithm and TCP/IP communication server, Web pages and the database service, carrying out the fingerprint data acquisition and other sensor data can be released by the MQTT server service through mobile Internet is sent to the server, and saved to the database.

The mobile terminal is mainly used for the collection of fingerprint sample data in the learning phase, the collection of F_i and the online fingerprint E acquisition at the online positioning stage. In the course of the experiment, the mobile terminal is composed of a smartphone and a APP equipped with it. The experimental personnel can be moved to the handheld intelligent mobile phone positioning area fingerprint data acquisition through APP; acquisition and learning is completed, also need to test again online positioning handheld intelligent mobile phone, check the effect of positioning algorithm (Fig. 2).

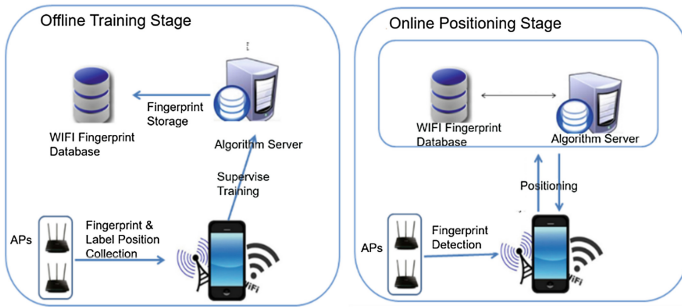


Fig. 2. Two stages for indoor positioning

4.2 Indoor Positioning Environment

There are 4 empty classrooms formed as “2 × 2” array as the experimental indoor environment. Each room has 1 AP on both sides of the north and south to ensure the WiFi fingerprint coverage, as shown in Fig. 3. In the experiment, the Android based smartphone was used as a fingerprint and display location tool.

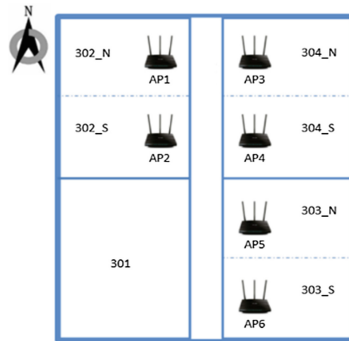


Fig. 3. Indoor environment

4.3 Experiment Result

Because of the characteristics of supervised learning algorithm, the main work of experiment is focused on fingerprint collection. In the experiment, the scan cycle is set to 5 s, and the learning time is 10 min (Table 1).

Table 1. Experimental data display

Exp No.	Actual Position	Algorithm Results (lists the top 3 highest probability locations)	
		Positioning Result	Probabilities
1	304_S	304_S	41
		302_S	18
		304_N	17
2	304_N	304_N	67
		303_S	10
		303_N	8
3	302_N	302_S	42
		302_N	37
		303_N	19
4	302_S	302_S	71
		303_S	17
		303_N	11
5	303_N	303_N	58
		303_S	20
		302_S	18
6	303_S	303_S	69
		302_S	15
		303_N	12

By comparing above results, it is found that the algorithm in Exp No. 1, 2, 4, 5, 6 can calculate a highest probability of location corresponding to the actual location, which means positioning result is correct. The accuracy rate of the algorithm is 83%.

In Exp No. 3, the algorithm mistakenly calculated highest probable location in 302_S, which is a neighbor location to the actual location 302_N, and 302_N had a very similar probability in this experiment. Which means Naïve Bayes method may occasionally classify a wrong category between very close locations.

The above is just a part of the data, but by many experimental results analysis, the algorithm is relatively stable, the positioning accuracy of 4 m, the latter can also be re optimization algorithm, to achieve higher positioning accuracy.

5 Summary

In this paper, a supervised classification based algorithm is established to study the indoor location problem. In this algorithm, WIFI fingerprint model, including the MAC address and the signal strength of RSSI, is used for data training and positioning. Moreover, Naïve Bayes classification method is applied to estimate the location of given WIFI fingerprint. The positioning results are accurate, the calculation method of position machine learning experiments, as long as a small amount of data acquisition

can get accurate positioning rules, save manpower cost, and positioning effect than the traditional method.

References

1. Su, H.-K., Liao, Z.-X., Lin, C.-H., Lin, T.-M.: A hybrid indoor-position mechanism based on bluetooth and WiFi communications for smart mobile devices. In: 2015 International Symposium on Bioelectronics and Bioinformatics (ISBB) (2015)
2. Chen, C., Han, Y., Chen, Y., Zhang, F., Ray Liu, K.J.: Time-reversal indoor positioning with centimeter accuracy using multi-antenna WiFi. In: 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (2016)
3. Ismail, A.H., Kitagawa, H., Tasaki, R., Terashima, K.: WiFi RSS fingerprint database construction for mobile robot indoor positioning system. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (2016)
4. Bisio, I., Cerruti, M., Lavagetto, F., Marchese, M., et al.: A trainingless WiFi fingerprint positioning approach over mobile devices. *IEEE Antennas Wirel. Propag. Lett.* **13**, 832–835 (2014)
5. Ohta, M., Sasaki, J., Takahashi, S., Yamashita, K.: WiFi positioning system without AP locations for indoor evacuation guidance. In: 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE) (2015)
6. Hin, B.J., Lee, K.W., Choi, S.H., et al.: Indoor WiFi positioning system for Android-based smartphone. In: International Conference on Information and Communication Technology Convergence (2010)
7. Krumm, J., Harris, S., Meyers, B., et al.: Multi-camera multi-person tracking for EasyLiving. In: Proceedings of the 3rd IEEE International Workshop on Visual Surveillance, Dublin, Ireland, pp. 3–10 (2000)
8. Orr, R.J., Abowd, G.D.: The smart floor: mechanism for natural user identification and tracking. In: Proceedings of the Conference on Human Factors in Computing Systems, Hague, pp. 275–276 (2000)

Satellites and Remote Sensing

Deep Learning and Machine Learning for Object Detection in Remote Sensing Images

Guowei Yang¹(✉), Qiang Luo¹, Yinding Yang², and Yin Zhuang²

¹ Beijing Institute of Spacecraft System Engineering, Beijing 100094, China
yangguowei2000@sina.com

² Beijing Key Laboratory of Embedded Real-Time Information Processing
Technology, Beijing Institute of Technology, Beijing 100081, China

Abstract. Object detection is one of the most effective ways to analyze the remote sensing (RS) images. In this paper, we focus on the prevalent object detection framework based on deep learning technology for RS images which contains three different stages, namely the region proposals generation, feature extraction, and classification. The review provides a clear picture of the challenges and possible development trends in this field. Typical methods under this framework are extensively reviewed and analyzed. Comparisons among traditional methods with deep learning methods are presented, in which supervised and unsupervised methods for RS scene target detection are deeply discussed.

Keywords: Deep learning · Object detection · Remote sensing

1 Introduction

The purpose of object detection in the field of remote sensing (RS) images is to find and locate the objects we are concerned such as ship [1] and airplane [2]. It is the foundation of RS field and plays an important role in many other applications such as changing detection, environment monitoring and target classification. Object detection in RS field often suffers from several special difficulties, including lacking training data and complex background interference.

The prevalent object detection framework generally consists of three stages, namely the generate region proposals, feature extraction, and classification. In this paper we aim to provide a review about the above framework, especially for the methods based on deep learning technology. We provide comparisons and extensive analysis among the learning-based methods in RS field. The contributions of this paper include reviewing the most prevalent methods in RS object detection field, providing comparisons between the deep learning based methods and the traditional methods, and providing research directions for application of deep learning in the field of RS scenes.

2 The Framework of Object Detection in RS Images

The most effective and prevalent framework in object detection of RS images contains three different stages. In the first stage, region proposals are extracted from the whole image. Each proposal is formed as a bounding box which indicates a probable object. After getting the entire region proposals, a robust feature is extracted from every proposals. Based on the feature, a classification is designed for the final decision. Typically, an image preprocessing operation and non-maximum suppression (NMS) can lead to a better performance. Figure 1 shows the overall framework of object detection system and the typical methods in each stage.

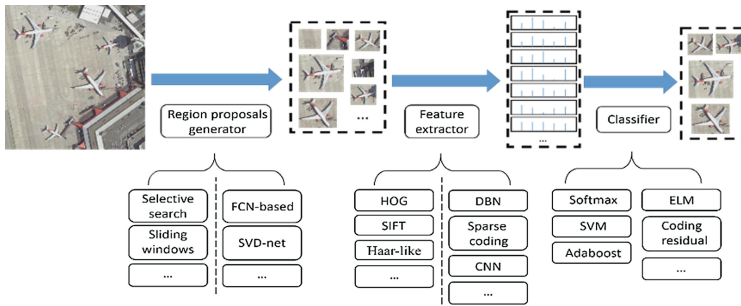


Fig. 1. Overall framework of object detection.

3 Region Proposals Generation

The mechanisms of feature descriptor and classifier have made a huge improvement of performance in the detection system. However, it is extremely time-consuming by using sliding windows to search the location of object in the traditional methods. Many outstanding object detection framework uses the region proposals generation algorithm [2, 4]. The goal of region proposal generation is to eliminate substantial irrelative areas which contain no objects in a short time. It contributes to designing more sophisticated classifiers without considering too many negative samples. A good region proposals generation algorithm should be able to suit for multi-scales object, have a high recall performance and converge in fast speed.

3.1 Non-network Methods

Despite the great success of deep learning, many object systems use non-network methods as their foundation to generate region proposals in RS field [1–5] and natural image field [2]. Hosang et al. [1] provide a holistic review about many different region proposal generating algorithms. All these algorithm are based on knowledge which means they don't have many parameters that need training. Many of these methods have been used in RS images. Zhao et al. [2] adopt multi-scale sliding windows to generate region proposals. Qi et al. [1] provide a saliency-based method to generate

region proposals. The saliency is computed in phase spectrum of Fourier transformation which can reduce the complexity significantly. Chen et al. [5] adopt sliding window technology on gradient images which can effectively reduce the number of region proposals. Cheng et al. [23] use the selective search method to generate region proposals and achieve good performance in variant objects. These methods have been proved to have good performance in some specific scenarios.

3.2 Network-Based Methods

Network-based methods attempt to generate region proposals through neural network. The most prevalent architecture of neural network to deal with image is convolutional neural network. It can generate a robust description of the whole image. However, in the region proposal stage, the description is ineffective. Many people engaged in finding strategies to generate region proposals by networks methods. Zhang et al. [8] adopt candidate region proposal network (CRPNet) along with the localization network to generate region proposals. CRP-net is a fully convolutional neural network, and is trained by images with $128 \text{ pixel} \times 128 \text{ pixel}$. In the forward period, it can be fed with images in any size and generate a candidate region map which is the object probability of each location. Zhou et al. [9] create the SVD-net which combines singular value decomposition (SVD) and convolutional neural network to generate region proposals. The proposed network contains three layers. The first two convolutional layer's parameters are determined by the SVD, and the third layer is to learn through the SGD algorithm. This network achieves efficient performance in practical offshore ships detection.

Network-based methods have many advantages. They can be easily deployed on the GPU and merged with the network-based stages afterwards. Besides, such methods are usually needed to be trained by data, which means we can include the object information in the first stage of the entire frame work so as to reduce the number of region proposals.

4 Feature Extraction

After we get the region proposals of the whole image, we need to extract the high level feature of each image patch. A good feature extractor should have the ability to handle variations such as intensity, scale, rotation and affine projection, and provide the same or similar representation of the same object. Features in this stage are crucial for the precision of the entire system.

There are many methods to extract features. The traditional hand-made extractor such as HOG [17], SIFT [5], Haar-like [18] are used for many years. However, after development for many years, they have reached the upper limit of performance especially for complex scenarios.

4.1 Sparse Coding Based Feature Descriptor

The general sparse coding cost function on a set of m input vectors can be expressed as Eq. (1):

$$\min_{\alpha, \varphi} \sum_{j=1}^m \left\| x^j - \sum_{i=1}^k \alpha_i^j \varphi_i \right\|_2^2 + \lambda \|\alpha\|_F, \tag{1}$$

where $\lambda \|\alpha\|_F$ is the penalty term which is regularized by L_0 or L_1 norm. The first term in Eq. (1) is the signal reconstruction process. α is the sparse representation of input x , and through the sparse coding, it can generate high level features of target representation without redundant information. The general sparse representation technique is adopted in RS scene target detection framework as shown in Fig. 2.

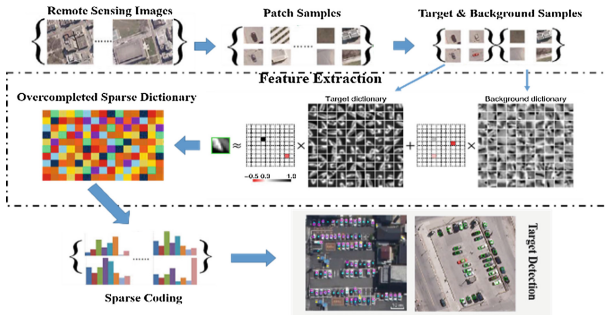


Fig. 2. Target detection framework of sparse representation

In recent years, a large number of sparse coding methods have been proposed and applied for RS scene classification. In the field of RS scene target detection, Yokoya [27] raise a general object detection framework based on sparse coding combine with the Hough voting. This method shows better performance of cars and ships detection in more complex RS scenes for optical imagery. Then, Chen [31] adopts high-resolution aerial images based sparse representation and super pixel technology which achieves vehicle detection in complex urban area. Zhang [32] proposes a method of sparse representation based binary hypothesis model for target detection in hyperspectral images. This binary hypothesis model utilizes the sparse coding reconstruction residual principle to achieve target detection. In general, the sparse representation has the abstraction ability of high level feature extraction, which achieves excellent performance in RS scene target detection.

4.2 Deep Belief Networks (DBNs) Based Feature Descriptor

Although supervised deep learning (DL) algorithms such as CNN can achieve high performance, they typically need large amount of labeled data. However, in RS image dataset, large amount of labeled images are not available. As a result, how to use the

original unlabeled images to learn the potential structure of the image is an attractive research direction. Deep belief network (DBN) (as shown in Fig. 3) is the first DL architecture that can be effectively trained with unlabeled data only [19]. Every unit in the same layer is connected with the neighboring layers. The top two layers are undirected while the connection of other layer pairs are directed.

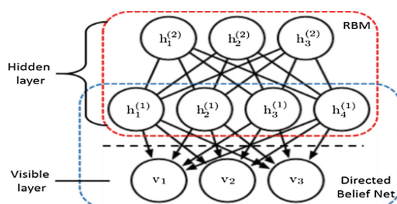


Fig. 3. The structure sample of DBN

Chen et al. [26] use four hidden layers DBNs as their feature extractor. For each two connected layers, it can be viewed as a restricted Boltzmann machines (RBMs), which can be trained easily through the CD algorithm [19]. Diao et al. [24] adopt a six layers DBN to extract the global features of the region proposals and the visualization of the features also shows the high distinguishable ability which is better than HOG-based features.

4.3 Convolutional Neural Network (CNN) Based Feature Descriptor

Convolutional neural network (CNN) is the most representative DL model when dealing with 2-D data such as image. The architecture of CNN enables it to generate a highly robust map from the raw pixels to an abstract representation which can greatly increase the performance of the following classifier. Unlike human-made feature extractor, CNN works in a supervised manner and needs to be trained by data. There are two ways to train a CNN model. One way is fine-tuning from a large network which is already trained by a large dataset (such as Image Net) [23, 25]; another way is to train a small network [20–22].

Cheng et al. [23] fine-tune their rotation-invariant CNN (RICNN) which is based on Alex Net in NWPU VHR-10 dataset. This model can extract rotation-invariant feature from the region proposals. Chen et al. [20] build a hybrid DNN (HDNN) which contains three convolution and max-pooling layers. This model is trained in vehicle dataset and achieves better performance than HOG and LBP.

Table 1 shows the features of different feature extractors.

Table 1. Comparison between different feature extractors

Feature extractor	Representation ability	Data dependency	Speed
Sparse coding	2	2	3
DBN	3	2	2
CNN	4	4	1
HOG, LBP	1	1	4

5 Decision and Regression

After we get a good representation for the image, we typically need a classifier as decision maker. It will finally predict the label of every region proposal. A good classifier should have the ability to reject the bad region proposals which do not contain an object and keep the right region proposals, which determines the precision. For each region proposal, regression is also a strong method to improve its location precision. As described in Eq. (2), the Smooth L1 loss type regression is prevalent in DL framework. The classifier maybe sometimes unclear because it is always combined with the feature extractor. Besides, Support vector machine (SVM) [9–14], extreme learning machine (ELM) [15], and adaBoost [16] are also used in RS image object detection system.

$$Smooth_{l_1} \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \quad (2)$$

Zou et al. [9] use SVD-net to extract the feature of the whole image which is followed by a SVM classifier to determine which region proposals of ship is the right one. In [15] Tang et al. apply the ELM to pool the feature generated from the CNN model and make classification decision. Also, using soft max as a classifier directly can lead to a good performance in some circumstances [24].

With the development of imaging technology, the resolution of image is getting higher and higher. This makes it possible to recognize the fine-category of objects such as the model of an airplane. This topic has been developed in the natural image which is called the fine-grained classification [3].

6 Conclusions and Future Works

In this paper, we make a holistic review about the prevalent object detection system in RS field. The DL framework which is rising in the field of machine learning have been introduced to the RS object detection field which out performs many traditional methods. Although some DL methods have been proposed, there are still many non-DL methods such as selective search and edge boxes etc. [1] that can achieve impressive results. The DL-based methods have a huge advantages to get high-level representation especially when large amount of training data is available. As for the classifier, once a good feature is obtained, which is typically through DL methods, it is not difficult to find a suitable classifier.

The following are some potential research directions. Generating region proposals with high recall and high accurate location is one of them, which is a great challenge and still remains open in RS scenes. Another is generating a robust feature for the region proposals without much labeled data, which is a crucial point since the unlabeled data are usually obtained, while few labeled data can be acquired in RS field.

References

1. Qi, S., Ma, J., Lin, J., Li, Y., Tian, J.: Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images. *IEEE Geosc. Remote Sens. Lett.* **12**(7), 1451–1455 (2015)
2. Zhao, A., Fu, K., Sun, H., Sun, X., Li, F., Zhang, D., Wang, H.: An effective method based on ACF for aircraft detection in remote sensing images. *IEEE Geosc. Remote Sens. Lett.* **14**(5), 744–748 (2017)
3. Zhang, N., Donahue, J., Girshick, R., Darrell, T.: Part-based R-CNNs for fine-grained category detection. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014, Part I*. LNCS, vol. 8689, pp. 834–849. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_54
4. Hosang, J., Benenson, R., Dollár, P., Schiele, B.: What Makes for Effective Detection Proposals? *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(4), 814–830 (2016)
5. Chen, X., Xiang, S., Liu, C., Pan, C.H.: Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **11**(10), 1797–1801 (2014)
6. Zhang, F., Du, B., Zhang, L., Xu, M.: Weakly supervised learning based on coupled convolutional neural networks for aircraft detection. *IEEE Trans. Geosci. Remote Sens.* **54**(9), 5553–5563 (2016)
7. Zou, Z., Shi, Z.: Ship detection in spaceborne optical image with SVD networks. *IEEE Trans. Geosci. Remote Sens.* **54**(10), 5832–5845 (2016)
8. Xia, Y., Wan, S., Yue, L.: A novel algorithm for ship detection based on dynamic fusion model of multi-feature and support vector machine. In: *2011 Sixth International Conference on Image and Graphics (ICIG)*, pp. 521–526, August 2011
9. Bi, F., Zhu, B., Gao, L., Bian, M.: A visual search inspired computational model for ship detection in optical satellite images. *IEEE Geosc. Remote Sens. Lett.* **9**(4), 749–753 (2012)
10. Zhu, C., Zhou, H., Wang, R., Guo, J.: A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geosci. Remote Sens.* **48**(9), 3446–3456 (2010)
11. Bi, F., Liu, F., Gao, L.: A hierarchical salient-region based algorithm for ship detection in remote sensing images. In: Zeng, Z., Wang, J. (eds.) *Advances in Neural Network Research and Applications*. LNEE, vol. 67. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12990-2_85
12. Xu, J., Sun, X., Zhang, D., Fu, K.: Automatic detection of inshore ships in high-resolution remote sensing images using robust invariant generalized hough transform. *IEEE Geosci. Remote Sens. Lett.* **11**(12), 2070–2074 (2014)
13. Tang, J., Deng, C., Huang, G., Zhao, B.: Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. *IEEE Trans. Geosci. Remote Sens.* **53**(3), 1174–1185 (2015)
14. Shi, Z., Yu, X., Jiang, Z., Li, B.: Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. *IEEE Trans. Geosci. Remote Sens.* **52**(8), 4511–4523 (2014)

15. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893, June (2005)
16. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, p. I (2001)
17. Lowe, D.: Distinctive Image Features from Scale-invariant Keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
18. Hinton, G., Osindero, S., Teh, Y.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2014)
19. Wang, J., Song, J., Chen, M., Yang, Z.: Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. *Int. J. Remote Sens.* **36**(12), 3144–3169 (2015)
20. Jin, X., Davis, C.: Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks. *Image Vis. Comput.* **25**(9), 1422–1431 (2007)
21. Cheng, G., Zhou, P., Han, J.: Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **54**(12), 7405–7415 (2016)
22. Diao, W., Sun, X., Zheng, X., Dou, F., Wang, H., Fu, K.: Efficient saliency-based object detection in remote sensing images using deep belief networks. *IEEE Geosci. Remote Sens. Lett.* **13**(2), 137–141 (2016)
23. Zhou, P., Cheng, G., Liu, Z., Bu, S., Hu, X.: Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. *Multidimension. Syst. Signal Process.* **27**(4), 925–944 (2016)
24. Chen, X., Xiang, S., Liu, C., Pan, C.H.: Aircraft detection by deep belief nets. In: 2013 2nd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 54–58, November (2013)
25. Yokoya, N., Iwasaki, A.: Object detection based on sparse representation and hough voting for optical remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **8**(5), 2053–2062 (2015)
26. Chen, Z., Wang, C., Wen, C., Teng, X., Chen, Y., Guan, H., Li, J.: Vehicle detection in high-resolution aerial images via sparse representation and superpixels. *IEEE Trans. Geosci. Remote Sens.* **54**(1), 103–116 (2016)
27. Zhang, Y., Du, B., Zhang, L.: A sparse representation-based binary hypothesis model for target detection in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **53**(3), 1346–1354 (2015)

Design and Implementation of Automatic Interpretation System for Remote Sensing Image

Linna Ni^(✉) and Yu Jiang^(✉)

Beijing Institute of Spacecraft System Engineering, Beijing, China
linnani01@126.com, tiancaijiangyu@126.com

Abstract. The high-precision remote sensors on satellite provide massive image data which brings new challenges to data processing and interpretation. The existing data processing systems are mostly semi-automatic which have the problems as low efficiency and low interpretation quality. This paper designs and implements a ground test system for processing and interpreting remote sensing images automatically and efficiently. The system introduces a two-step feature parameter correlation calculation algorithm to interpret the images. The authors realize the system based on the structure of universal server + FPGA which capabilities can achieve a processing rate of tens of gigabit per second.

Keywords: Remote sensing image · Image interpretation
Feature parameter correlation calculation

1 Introduction

With the development of space technologies in recent years, the number of satellites produced in each year increases rapidly. Meanwhile, the satellite become more complicated while one satellite usually carries several different high-precision sensors which provide image products including visible-light image, multispectral image, hyperspectral image, etc. [1]. Thus, the amounts of produced image data grow exponentially from 100 Mbit/s to Gbit/s level only in a decade which brings huge challenge to the ground test system [2–5].

In AIT (assembly, integration and test) phase, the ground test system receives the image data from satellites and implements calculation, analysis and interpretation task to verify the performance and functionality of the satellite sensors [6]. Most of the existing ground systems are semi-automatic which means the interpretation work is more or less based on the manual operation [7–12]. As the amounts of image data grow rapidly, the data processing delay becomes unacceptable which depends on the data volume. Also, the quality of testers' interpretation work may drop off when the testers face massive data.

In view of this situation, this paper introduces a highly automatic system for data processing and interpretation based on the innovative algorithm and hardware structure. In Sect. 2, the two-step feature parameter correlation calculation algorithm is described which is adopted in the system to dig out different kinds of errors in the images.

In Sect. 3, the system based on the structure of universal server + FPGA is implemented and the data flow of analysis and interpretation procedure is introduced. Section 4 is the experiment results section in which the system performance is presented together with the screenshot of the system working interface. In the last section, the authors make the conclusion of the system followed by some considerations of future.

2 Design of Image Interpretation Algorithm

In AIT phase, the ground system processes and interprets the received the image data from satellites to detect different kinds of image errors. The image data is normally compressed before it is transmitted from the satellite to the ground. The ground test system decompresses the data to restore the image [13, 14]. Based on the compression algorithm, the fixed line number of image data is called “one picture” as the smallest compression unit. The interpretation algorithm proposed in this paper is based on the comparison between consecutive pictures. The interpretation algorithm can be divided into 2 steps. In Step1, the system detects the dramatic change by comparing neighboring images which is considered as the error occurs. In Step2, the error is located. The 2-steps interpretation algorithm is described as below:

Step1: the histogram values of all the pictures are calculated firstly. The histogram equation is shown as below:

$$p(r) = \frac{n(r)}{NUM} \quad (1)$$

Where $p(r)$ is the probability of pixels with gray value of r in the picture; $n(r)$ is the number of pixels with gray value of r ; NUM is the number of pixels in the whole picture. The histogram of one picture is the set of $p(r)$ with all possible gray values of r .

Next, the histogram correlation values between every two neighboring pictures are calculated which can be given by the equation:

$$Cov = \frac{\sum_{r=1}^{2^N} p_{cur}(r) \cdot p_{fro}(r)}{\sqrt{\sum_{r=1}^{2^N} p_{cur}^2(r)} \sqrt{\sum_{r=1}^{2^N} p_{fro}^2(r)}} \quad (2)$$

Where Cov is the histogram correlation value; N is the bit number of pixels; $p_{cur}(r)$ is the probability of pixels with gray value of r in current picture; $p_{fro}(r)$ is the probability of pixels with gray value of r in the former picture.

The flow diagram of Step1 is shown as Fig. 1. The bottom line of the figure is the calculation procedure of histogram values while the middle line shows how Cov is figured out.

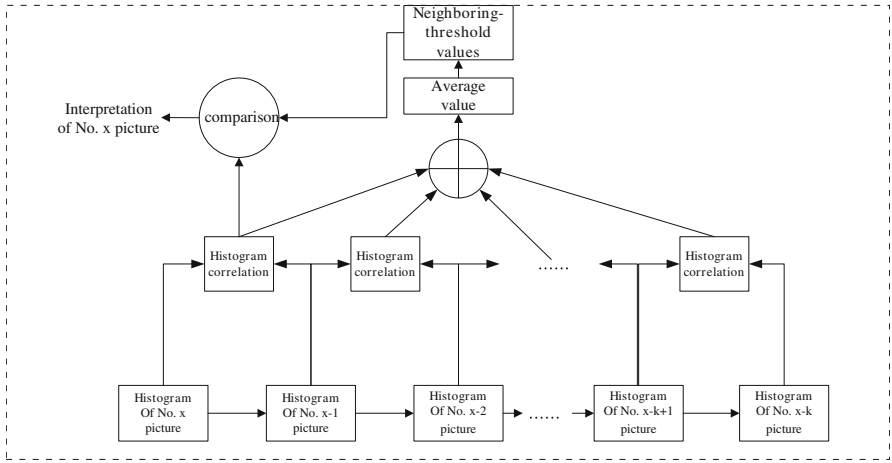


Fig. 1. The flow diagram of histogram correlation interpretation algorithm

Next, we compare each histogram correlation value with its neighboring-threshold values which are based on the average value of last several histogram correlation values. We consider the picture as abnormal when its histogram correlation value is out of the range of the neighboring-threshold values. It is noticed that the number of histogram correlation values for average calculation is configurable. The range of neighboring-threshold values is also configurable, for example as 90%–110% of the average value. Thus, the algorithm is self-adaption for different application scenes.

Step2: while the abnormal picture is detected in Step1, the next step is to locate the error pixels in the picture. In AIT phase, the images are normally simple and formalized which means the gray value of a pixel has high correlation with its neighbor pixels. Thus we use the method called neighboring pixel comparison to compare the central pixel and its 8 neighboring pixels for error detection. Firstly, we calculate the average gray value of the neighboring pixels of pixel c using the equation as below:

$$L_p(c) = \frac{\sum_{k=1}^8 x_k(c)}{8} \tag{3}$$

Next we calculate the difference between $L_p(c)$ and the gray value of the central pixel which is $x_o(c)$.

$$err = x_o(c) - L_p(c) \tag{4}$$

Similar to Step1, we introduce the neighboring-threshold values which are based on the average value of last several pictures and compare the neighboring-threshold values with the value of current picture to detect the error pixels. The flow diagram of Step2 as described above is shown in Fig. 2.

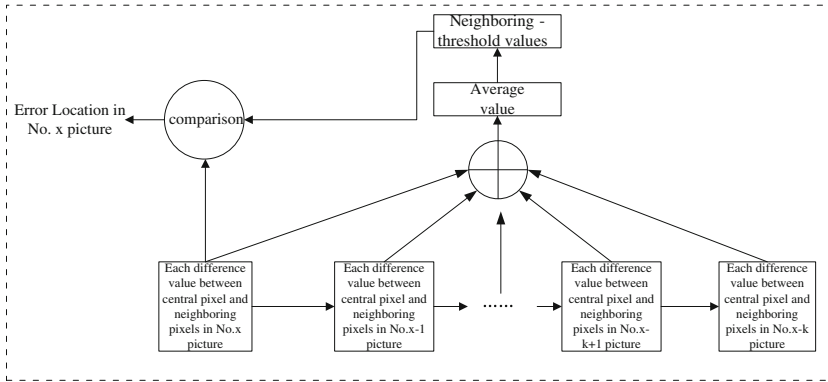


Fig. 2. The flow diagram of histogram correlation interpretation algorithm

3 Implementation of Remote Sensing Image Interpretation System

The traditional system based on the universal servers and the data processing software is unable to meet the requirement of high-speed data processing and interpretation. On the other hand, by the parallel pipeline functions of FPGA, the bottle-neck of speed can be resolved on a large scale. This paper proposes a structure of universal server+ FPGA to implement the remote sensing image interpretation of massive data. The structure of the system is shown in Fig. 3.

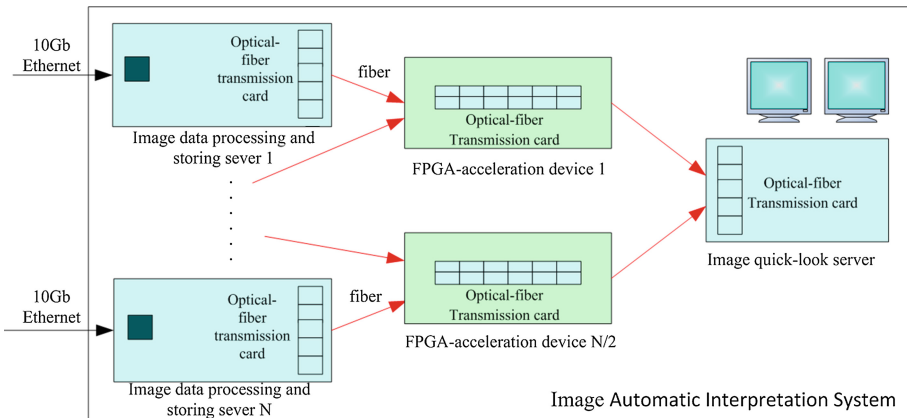


Fig. 3. The structure of the remote sensing image interpretation system

The system is composed of 3 parts:

- (1) The image data processing and storing sever receive the image data from the preceding devices, check the data format and report the format error, store the raw data and transmit the data to the FPGA-acceleration devices.

- (2) The FPGA-acceleration devices are used to implement the two-step data interpretation algorithms. As FPGA has advantages in parallel processing, each accelerate device has maximum capability of 10 Gbps that can process data from several preceding servers at the same time.
- (3) The image quick-look server receives the sampled image data and the interpretation results from the FPGA device, displays the pictures on the screen and reports the image errors to the users.

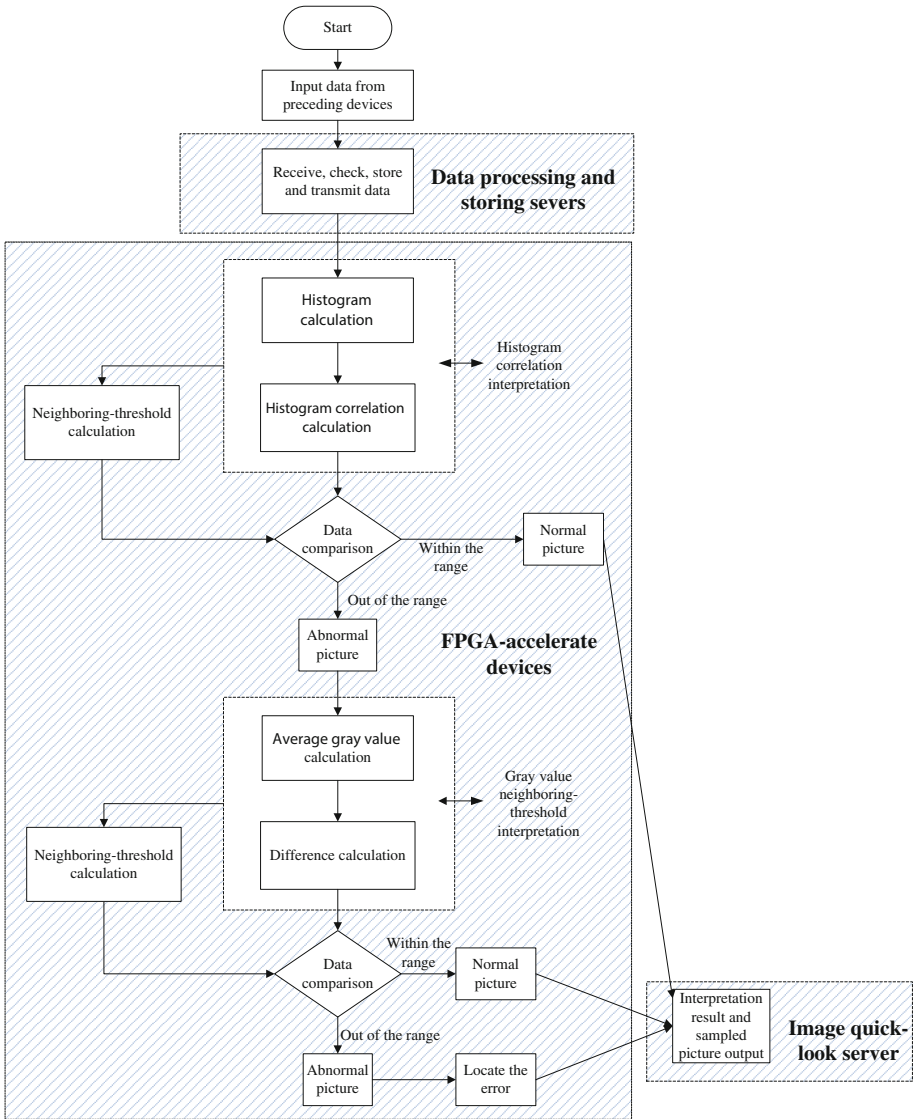


Fig. 4. The data flow of the remote sensing image interpretation system

The data flow of the system is shown in Fig. 4 from the top to the right-bottom corner of the figure. There are 3 striped rectangles in the figure corresponding to the data processing and storing servers, the FPGA-acceleration devices and the image quick-look server.

4 Experiment Results

To meet the high-speed processing requirement, the authors construct a powerful system including 16 data processing and storing servers, 8 FPGA-accelerated devices and 1 quick-look server in 2016. The practical results proved that the system has capabilities of more than 50 Gbps and 95% error detection rate is achieved. The user interface is shown in Fig. 5.

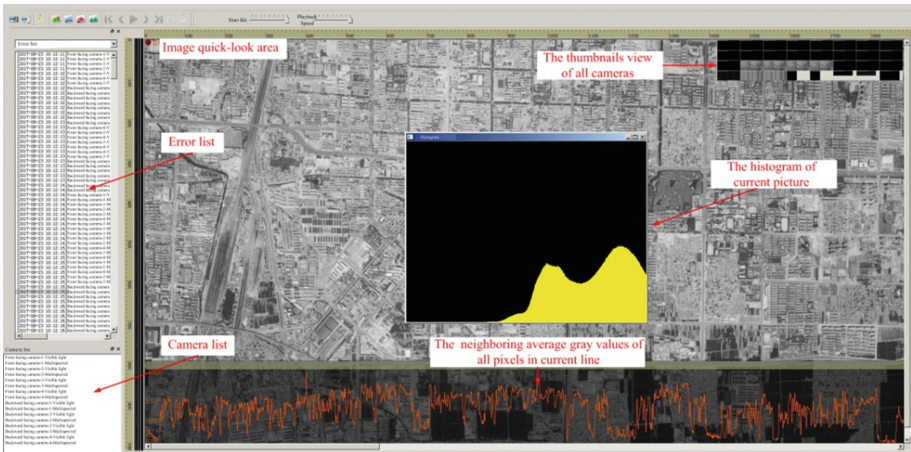


Fig. 5. The user interface of the remote sensing image interpretation system

The central part of Fig. 5 is the image quick-look area which shows the pictures from the camera in detail. The thumbnails view of the pictures from other cameras is on the top right of the figure while the users can switch the quick-look area to another camera by clicking the specific area of the thumbnails view. It is optional to display the histogram value of the current picture. On the bottom of Fig. 5, the curve of the neighboring average gray values changes continuously as new pictures are processing. The error list and camera list are on the left of the figure. All the detected errors are reported in the error list and also stored in the logbook of the system. The users can double-click any single error warning and then the quick-look area will jump to the abnormal picture with the error in it.

Besides the picture mode of the image quick-look area, there are two more modes as the curve mode and the hexadecimal mode to give the users different perspectives to analysis the remote sensing data as shown in Fig. 6.

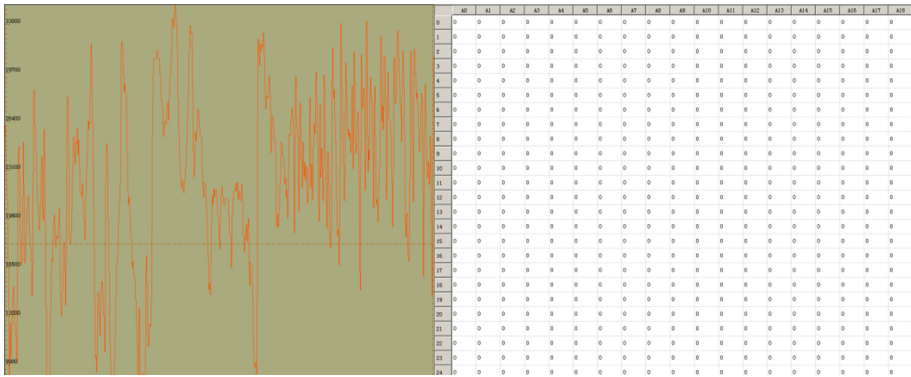


Fig. 6. The curve mode and the hexadecimal mode of the system

The capabilities of 64 channels (in another word the system can process the data from 64 sensors concurrently), high processing speed of more than 50 Gbit/s and 95% error detection rate are achieved. The system is proved to be an automatic, efficient and high interpretation quality system by real experience.

5 Conclusion

This paper introduces a two-step feature parameter correlation calculation algorithm to interpret the images. The authors design and implement a universal server+FPGA structure system to realize high-speed and automatic interpretation of images from remote sensors. The application scenario of the system is AIT phase of satellite in which the images are mostly simple and formalized.

Future developments should focus the following aspects:

- (1) More interpretation algorithm focusing on different image features could be extended into the system.
- (2) The database for recording and classifying the errors could be established.
- (3) The function of offline interpretation could be provided to realize Big Data analytics and improve the quality of interpretation.

References

1. Zhao, L., Ni, L., Liu, Y.: Design and implementation of data processing platform for remote-sensing satellite payload. *Spacecr. Eng.* **23**(4), 77–83 (2014)
2. Kasai, R., Oniyama, A., Fukunaga, T.: User-oriented operation of ASNARO. In: *SpaceOps 2016 Conference*, pp. 2016–2581, Daejeon (2016)
3. Xia, Q., Wang, D., Zhang, L.: Realization of fast-view system for high-speed multi-channel remote sensing camera. *Opt. Precis. Eng.* **21**(1), 158–166 (2013)

4. Liu, L.: Design of real-time processing platform for satellite remote sensing data based on MPI. *Spacecr. Eng.* **22**(3), 130–134 (2013)
5. Liu, X.: “Fast-View” technology and its implementation for high-resolution image. *Radio Eng. Chin.* **36**(3), 20–22 (2006)
6. Garg, N., Singla, S., Jangra, S.: Challenges and techniques for testing of big data. *Procedia Comput. Sci.* **85**(2016), 940–948 (2016)
7. Kaur, K., Khatri, S., Datta, R.: Analysis of various testing techniques. *Int. J. Syst. Assur. Eng. Manage.* **5**(3), 276–290 (2014)
8. Navarro, V., Pollock, A.: One-click data analysis software for science operations. In: *SpaceOps 2016 Conference*, pp. 2016–2399, Daejeon (2016)
9. Miller, S., Grant, K.: Adding a mission to the Joint Polar Satellite System (JPSS) Common Ground System (CGS). In: *AIAA SPACE 2015 Conference and Exposition*, pp. 2015–4470, California (2015)
10. Tschan, C., Ludwig, J.: How intelligent is your satellite or ground system. In: *AIAA SPACE 2013 Conference and Exposition*, San Diego, pp. 2013–5483 (2013)
11. Firtz, M., Winter, S., Freund, J., Pflueger, S., Zeile, O., Eickhoff, J., Roeser, H.: Hardware-in-the-loop environment for verification of a small satellite’s on-board software. *Aerosp. Sci. Technol.* **47**, 388–395 (2015)
12. Bester, M., Roberts, B., Lewis, M., Marchant, W.: NuSTAR ground systems approach – lessons learned. In: *AIAA SPACE 2013 Conference and Exposition*, San Diego, pp. 2013–5405 (2013)
13. He, F., Shan, Q.: Design of remote sensing data transmission baseband system based on blade system. *Comput. Meas. Control* **19**(3), 668–672 (2011)
14. Tian, L., Fan, S.: Application and discussion of remote sensing data processing based on multi-payload. *Spacecr. Eng.* **18**(3), 77–83 (2009)

An Improved Cloud Detection Method of Optical Remote Sensing Image

Yang Gao^{1(✉)}, Hao-tian Zhou², and Liang Chen²

¹ China Academy of Space Technology, Beijing 100094, China
35794664@qq.com

² Radar Research Lab, Beijing Institute of Technology, Beijing 100081, China

Abstract. The effect of cloud cover on the quality of remote sensing data becomes an unavoidable problem when dealing with a large amount of remote sensing data obtained from satellite sensors. As an important meteorological element, cloud plays a vital role in all areas of atmospheric science. In this paper, we propose a cloud detection method based on multi-feature hierarchical judgement. First, the gray histogram of the object to be interpreted is extracted and the histogram is intercepted to remove the singular value. Then, five types of feature are employed in feature extraction. After that, the objects to be interpreted is divided into single type and mixed type, and mixed type can be further divided into certain mixed type and uncertain type. Finally, threshold method and support vector machine(SVM) are employed to classify these types. Experiment has shown good performance of the proposed method.

Keywords: Cloud detection · Remote sensing · Feature extraction

1 Introduction

The remote sensing technology firstly began in the early 1960s. With the development of aerospace industry and the continuous maturation of satellite technology, the remote sensing technology is widely used in various fields such as weather analysis, environmental monitoring, resource exploration and military reconnaissance. The effect of cloud cover on the quality of remote sensing data becomes an unavoidable problem when dealing with a large amount of remote sensing data obtained from remote sensing sensors. As an important meteorological element, cloud plays a vital role in all areas of atmospheric science. However, in the optical remote sensing, the existence of the cloud inevitably affects the satellite imaging, identification and application. Therefore, the elimination of cloud cover is critical to satellite remote sensing processing.

Detection of the clouds is the first step to eliminating the cloud coverage. At present, the common cloud detection technology can be divided into threshold method, cluster analysis method and artificial neural network method.

The threshold method is proposed the earliest and it is also the easiest to implement. Consequently, it is widely used in the cloud detection. Its basic principle is to compare the brightness and reflectance of the pixel with the pre-set threshold.

The clustering analysis was proposed by Diday and began to be used in the study of cloud detection in the 1980s. The common clustering methods are histogram clustering, dynamic threshold clustering, adaptive clustering and so on.

The Artificial neural network method uses the sample learning and simulates the working principle of the human brain system to construct the neural network with the recognition ability of the cloud.

Saunders and Kriebel proposed an improved method for detecting clear sky and cloudy radiances based on Advanced Very High Resolution Radiometer (AVHRR) data, named AVHRR Processing Over Land cLOUD and Ocean (APOLLO) method in 1988 [1] and Stowe proposed a cloud detection method based on threshold classification method in 1991 [2]. In 1993, Rossow proposed a cloud detection method using satellite measurements of infra-red and visible radiances for The International Satellite Cloud Climatology Project (ISCCP) [3]. The CO2 flake method was firstly proposed by Wylie in 1994 [4].

In this paper, we propose a cloud detection method based on multi-feature hierarchical judgement. First, the gray histogram of the object to be interpreted is extracted and the histogram is intercepted to remove the singular value. Then, the objects to be interpreted is divided into two parts. One part is named single object which contains only one type of object. The other part is named mixed object which contains two or more types of objects. For mixed objects, the variance standard ratio is used as a threshold for a certain or an uncertain object. For the single type object and certain object, the threshold method is used for classification, and SVM classifier are used for uncertain objects. The variance standard ratio is the ratio of the gray variance to the gray standard deviation of the interpretation object. The larger the variance standard ratio is, the closer the uncertainty is. Experiment result has shown the good performance of our method. The rest of this paper is organized as follows. The proposed method is described in Sect. 2. Experiment results are presented in Sect. 3 followed by the conclusions drawn in Sect. 4.

2 Proposed Method Based on Multi-feature Hierarchical Judgement

In this paper, we propose a cloud detection method based on multi-feature hierarchical judgement. The thought of classification isn't employed in the basic idea of the proposed method, which means not classify the cloud and non-cloud, but classify the objects into accurate results and fuzzy results. The advantages of this method are: (1) The judged accurately objects will not be affected by the judged fuzzily objects. (2) The classifier can be designed aimed at the judged fuzzily objects. (3) The judged fuzzily objects can be refined by the judged accurately objects. The flow of the proposed method is in Fig. 1. The individual steps of the method are briefly presented in the following discussions.

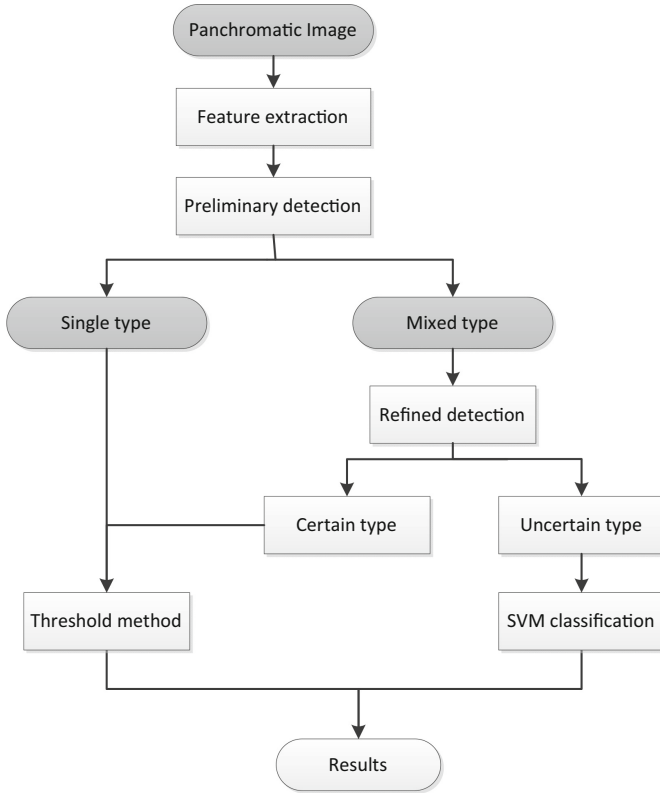


Fig. 1. The work flow of the proposed method.

2.1 Data Preprocessing Based on Gray Histogram Extraction

In the first step of our method, we first deal with the input panchromatic remote sensing images. Usually, in the beginning, we divide the input image into patches. Then the histograms of the patches are extracted. After that, in order to increase the robustness of subsequent histogram-derived parameters, such as mean and variance, for some singular values, the histogram needs to be intercepted to remove singular values.

In this stage, a critical step is to intercept the gray histograms. In the proposed method, we achieve it by the following algorithms. First, the pixels are accumulated from the highest level of the gray scale, stops when the sum of the accumulated pixels reaches M , and the gray level when the sum of the accumulated pixels reaches M is defined as F_h . Similarly, the pixels are accumulated from the lowest level of the gray scale, stops when the sum of the accumulated pixels reaches N , and the gray level when the sum of the accumulated pixels reaches N is defined as F_l . The subsequent step is to deal with the pixels between the upper bound F_h and lower bound F_l , which are the results after removing the singular value. The results after removing singular value can more accurately reflect the characteristics of the object.

2.2 Feature Extraction

The application background of cloud detection has put forward a lot of special requirements for the extraction and selection of cloud features. First, due to the diversity of the clouds and the complexity of the features, the cloud area often needs to be discriminated by various features to increase the separability of the cloud and the land in the feature space. Second, cloud detection often need to sample the remote sensing data in order to reduce the amount of calculation. This requires that the feature will not be changed by the multi-scale images, which requires the feature to meet the scale invariance.

In the proposed method, we use five features in this stage. These features can be divided into two categories, the first is to measure the degree of object mixing characteristics, including gray variance, non-zero width of histogram and variance standard ratio; the second category is to measure the features of the single type, including gray mean and cloud coverage ratio.

The extraction of gray mean is to accumulate the gray value of pixels between the upper bound F_h and lower bound F_l , and calculate the average, which can be modeled as:

$$\bar{\rho} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N f(i,j), F_l \leq f(i,j) \leq F_h \quad (1)$$

The extraction of gray variance is to calculate the gray variance of pixels between the upper bound F_h and lower bound F_l , which can be modeled as:

$$\sigma = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [f(i,j) - \bar{\rho}]^2, F_l \leq f(i,j) \leq F_h \quad (2)$$

The non-zero width of the histogram is the difference between the upper bound F_h and lower bound F_l . The cloud coverage ratio is extracted using the double experience threshold of gray value and Sobel edge information, and the threshold is very tight when we use.

The extraction of variance standard ratio can be divided into two steps. First is to calculate the gray variance standard value. This standard value is the gray value that is calculated when the pixels between the upper bound F_h and lower bound F_l are evenly distributed, which can be modeled as:

$$H = (4096 - M - N) \times \left(\frac{F_h - F_l}{2} + 1\right) \times \left(\frac{F_h - F_l}{2} + 2\right) \quad (3)$$

The second step of the extraction of variance standard ratio is to calculate the ratio between gray variance and gray variance standard value.

2.3 Preliminary Cloud Detection

In this stage, we use the extracted features to discriminate the single type and the mixed type. Gray variance and non-zero width of the histogram are used to measure the number of gray levels in the object to be interpreted, and the type contains more grayscale is closer to the mixed type.

During this stage, the principle of the threshold selection is try to make the single type more accurate. It is because the pattern recognition method aimed at the mixed type is more complicated, and easier to deal with complex situations. So the uncertain situation is discriminated as mixed type as far as possible.

2.4 Refined Cloud Detection

For the objects of mixed type, the variance standard ratio is employed as threshold to divide the objects into certain mixed type and uncertain mixed type. The certain mixed type means although the objects to be interpreted contains the mixture of cloud and land, but more biased towards cloud or land. The uncertain mixed type means the proportion of cloud and land in objects to be interpreted is quite equal. For the single type and the certain mixed type, the threshold method is employed in reclassification. And the uncertain mixed type can be classified by the trained classifier.

For the single type, the gray mean value and cloud coverage can be used as threshold to achieve classification, and the classification results can be divided into absolute clouds, thick clouds, thin clouds, land objects and ocean.

For the mixed type, the variance standard ratio can be used to measure the relative intensity between cloud and land objects. When the proportion of cloud and land in objects to be interpreted is equal, the variance standard ratio is largest. If one's proportion is larger than other, the variance standard ratio is less. So the variance standard ratio can be used as threshold to divide mixed type into certain mixed and uncertain mixed. When variance standard ratio is smaller than the threshold, the patch possibly contains a small part of cloud or land, and can be identified as cloud or land, which is defined as certain mixed type object. When variance standard ratio is larger than the threshold, these patches are defined as uncertain mixed type. In this situation, because of their approximate features, threshold method doesn't have good performance. In the proposed method, the SVM classification is employed in the refined classification due to its simple implementation method, superior judgment effect and less training data.

3 Experiment

In experiment, to validate the effectiveness of the proposed method, we used panchromatic remote sensing images from GF-1 to do the test experiment. The proposed method was implemented in Matlab R2016a, Windows 10. And the computer configuration is Intel® Core(TM) i7-6500U CPU @ 2.50 GHz, 4.00 GB RAM. The detection results are shown in Fig. 2.

The experiment also compared our method with other methods, such as Gray-level co-occurrence matrix(GLCM) method and Pixel-level Bag of Visual word(BOV)

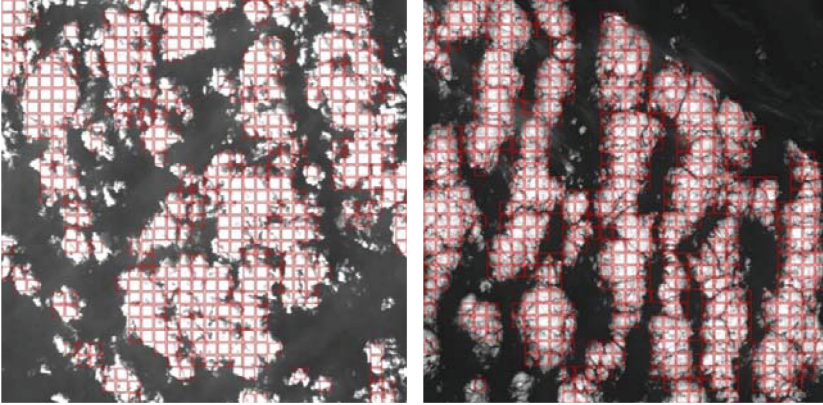


Fig. 2. Cloud detection results.

classification. To evaluate the effectiveness of these method, precision and recall are employed as performance metrics, defined as

$$\textit{Precision} = \frac{\textit{Ture Positive}}{\textit{True Positive} + \textit{False Positive}} \quad (4)$$

$$\textit{Recall} = \frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Negative}} \quad (5)$$

The detailed results are shown in Table 1. Experiment has shown the good performance of our proposed method. Compared with other methods, the proposed method applies different measures for different types of targets. The mixed type and the single type targets are dealt according to their degree of difficulty. This is the reason for the good performance of the proposed method.

Table 1. Performance of our proposed method and the state-of-the-art methods.

Method	Precision	Recall
GLCM method	61.86%	81.43%
Pixel-level BOV method	74.43%	86.91%
Proposed method	81.12%	91.07%

4 Conclusions

In this paper, we propose a cloud detection method based on multi-feature hierarchical judgement. The method contains two detection steps. In the first step, the objects are divided into single type and mixed type. If mixed type, the variance standard ratio is employed to discriminate certain mixed type and uncertain type. The single type and certain mixed type can be classified by the threshold method, and the uncertain mixed

type is discriminated by SVM classifier. Experiment results has shown good performance of our method.

References

1. Saunders, R.W., Kriebel, K.T.: An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sens.* **9**(1), 123–150 (1988)
2. Stowe, L.L., McClain, E.P., Carey, R.: Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Adv. Space Res.* **11**(3), 51–54 (1991)
3. Rossow, W.B., Garder, L.C.: Cloud detection using satellite measurements of infrared and visible radiances for ISCCP. *J. Clim.* **6**(12), 2341–2369 (1993)
4. Wylie, D.P., Menzel, W.P., Woolf, H.M.: Four years of global cirrus cloud statistics using HIRS. *J. Clim.* **7**(12), 1972–1986 (1994)
5. Taylor, V.R., Stowe, L.L.: Reflectance characteristics of uniform earth and cloud surfaces derived from Nimbus7 ERB. *J. Geophys. Res.* **89**(D4), 4987–4996 (1984)
6. Gu, Z.Q., Duncan, C.N., Renshaw, E.: Comparison of techniques for measuring cloud texture in remotely sensed satellite meteorological image data. *IEE Proc. F Radar Signal Process.* **136**(5), 236–248 (1989)
7. Wang, A.Z., Jin, J., Liang, J.: A new cloud removal algorithm for multi-spectral images. In: *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6043, pp. 60430W–60430W-11 (2005)
8. Jedlovec, G.J., Haines, S.L., Lafontaine, F.J.: Spatial and Temporal Varying Thresholds for Cloud Detection in GOES Imagery. *IEEE Trans. Geosci. Remote Sens.* **46**(6), 1705–1717 (2008)
9. Tseng, D.-C., Tseng, H.T.: Automatic cloud removal from multi-temporal SPOT, images. *Appl. Math. Comput.* **205**(2), 584–600 (2008)
10. Maalouf, A., Carre, P., Augereau, B.: A bandelet-based inpainting technique for clouds removal from remotely sensed images. *IEEE Trans. Geosci. Remote Sens.* **47**(7), 2363–2371 (2008)
11. Benabdelkader, S., Melgani, F.: Contextual spatio-spectral postreconstruction of cloud-contaminated images. *IEEE Geosci. Remote Sens. Lett.* **5**(2), 204–208 (2008)

A Novel Method to Analyze Dual Camera Pointing Direction Difference of Remote Sensing Satellite

Kan Cheng¹(✉), Zihao Cui², Tao He¹, and Mengjie Shi¹

¹ Beijing Institute of Spacecraft System Engineering, Beijing, China
ckjack_007@163.com

² Wuhan University, Wuhan, China

Abstract. This paper proposes a new method for analyzing dual camera pointing direction difference of remote sensing satellite. This method can calculate relative pointing direction difference without reference to any Ground Control Points (GCPs). This enables the analysis of direction change in long image time. The theorem of the method is demonstrated, including initial calibration for dual-cameras. An experiment is also carried out based on one Chinese remote sensing satellite. Results from different images are presented to demonstrate this method can apply to random orbit. This method is widely applied on evaluation of satellite performance and improvement of image quality.

Keywords: Dual camera · Pointing direction difference
Remote sensing satellite

1 Introduction

The quality of image is one important index for performance of Remote Sensing Satellite. It can also provide important information of how satellite performance change by time. This information can also help to develop methods for further improvement on quality. Chinese Space Industry has developed several remote sensing satellites equipments with dual-camera in recent years. They provide wider coverage with higher resolution thanks to dual-camera approach. These satellites have been proven to enhance earth observation ability.

The ideal product of such satellite is one stitching image from both cameras. Traditional image stitching methods have been studied by many researchers and can be generally grouped into the image-space-oriented and the object-space-oriented [1, 2] methods. The latter approach aims at establishing the mapping relation between the mosaic image and the original images by rigorous geometric models. There are also approach based on the mapping relation between geometric models of the virtual CCD and real CCDs [3], and acquired seamless mosaic images of multi-CCDs cameras [4]. However, these methods only considered the multi-CCDs within one camera. They are incapable to process dual-cameras stitching problem.

However, due to instability nature of camera mounting structure, a drifting of two cameras pointing direction is observed in many of these satellites. This makes it

impossible to use traditional method to stitching images in processing. Either method available is not capable of stitching image from two images without lost geometric accuracy.

To develop a stitching method fit for image, the shift pattern of dual camera pointing direction differences have to be known at first place. There are two question should be answered. First question is how the pointing direction of camera changes in long term, as in several months. This can be solved by existing on orbit geometric calibration method. With help of high accuracy GCPs, the interior orientation of both cameras can be obtained and compared. This helps to determine the frequency of calibration. Another question is how cameras change its pointing direction during a short period of time, as in one image mission. This change has much bigger impact in image processing, because if there is significant time-sensitive direction change, the geometric model in image processing has to be time-related. Traditional camera pointing methods need GCPs to determine the direction of each pixel. These method is not applicable in this case for there are not enough GCPs to satisfy the time span of image.

This paper proposed a novel method to evaluate dual camera pointing direction difference, which is tested in satellite images. This method does not depend on GDPs to calculate the pointing difference. This enables analyzing of all images of one orbit and any orbit of any time. With this merit, this paper demonstrated that the pointing direction difference of two cameras in one satellite shifts about 2 pixels at period of 5 min.

2 Geometric Image Properties of Dual Cameras

Dual-cameras are applied to expend the field of view of remote sensing satellite. As in most cases, two cameras of satellite are mounted on one main structure. The fields of view of two cameras are lined across flight direction to realize wider coverage. The projection of two cameras on the ground in relation of flight direction is shown in Fig. 1.

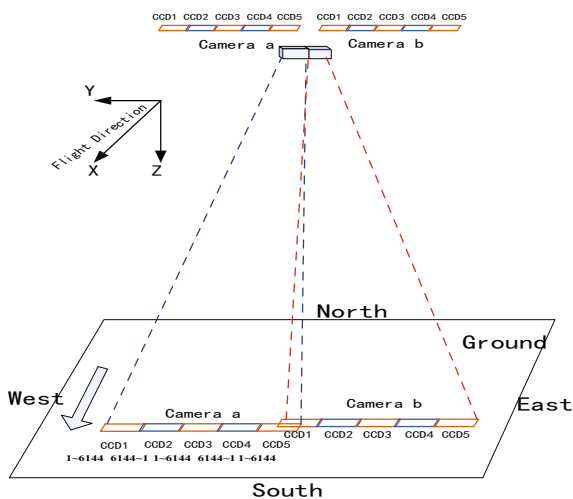


Fig. 1. The projection of dual cameras' field of view of dual-camera remote sensing satellite

There are overlapping in fields of view of two cameras, which equals to 990 m of length on the ground. And there is 0.98 m in mounting position along the flight direction. Therefore, for one subject on the overlapping area on the ground, two cameras take the images in sequence. The time difference should be one integration time.

There are several possible causes for shift of dual camera pointing direction difference. All of these caused related to in orbit environment change. For example, because two cameras image the same object at the overlapping area separately, the stability of satellite platform can also cause stitching error. Another factor is the stability of mounting structure of two cameras. This factor has been considered during the design of this structure. High precision temperature control is employed to limit the influence of environment on the structure. Simulation shows it effectively reduced the shift in direction between two cameras, camera and star trackers. There is no quantified analysis on exact cause of pointing direction difference yet. To solve this problem, it is necessary to evaluate the amount of changes in orbit.

3 Analysis of Dual Camera Pointing Direction Difference

3.1 Theorem of Image Based Dual Camera Pointing Direction Difference Calculation Method

As demonstrated before, the aim of this analysis is show the change of dual cameras pointing direction in short period time. For the purpose of this analysis, only the relative direction is needed. Therefore, the object of this analysis concentrates on overlapping area of images from dual cameras.

Based on the rigorous geometric model of remote sensing satellite imaging, there is:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{WGS84} = \begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix}_{WGS84} + mR_{orbit2wgs84}R_{body2orbit}R_U R_{camera2body} \begin{bmatrix} x - x_0 - \Delta x \\ y - y_0 - \Delta y \\ f \end{bmatrix} \quad (1)$$

Equation 1 shows the rigorous geometric model of linear array push-broom sensors, where $(X \ Y \ Z)_{WGS84}^T$ is the object position vector in the geocentric earth-fixed coordinate system and $(X_s \ Y_s \ Z_s)_{WGS84}^T$ is the position of the satellite with respect to the geocentric earth-fixed coordinate system. Furthermore, m denotes the scaling factor, $R_{camera2body}$ denotes the rotation matrix for converting the sensor coordinate system to the satellite body coordinate system, which were determined before launch. $R_{body2orbit}$ is based on satellite attitude data and defines the transformation between the satellite body coordinate system and the geocentric inertial coordinate system, whereas $R_{orbit2wgs84}$ represents the rotation matrix for converting the geocentric inertial coordinate system to

the geocentric earth-fixed coordinate system. (x, y) is the image coordinates, (x_0, y_0) is the principal point position, f is the focal length, and $(\Delta x, \Delta y)$ denotes CCD distortions. R_U represents the exterior orientation, interior orientation is represented with pointing direction model.

The initial pointing direction is calibrated with relative calibration. Conventional geometric calibration methods compensate for systematic errors of isolated individual camera without consideration of the relative geometric relation between multi-cameras onboard one satellite [6, 7]. To recover the relative geometric relation of dual-cameras onboard one satellite, the common errors of dual-cameras and the specific errors of each camera should be distinguished based which a relative calibration model can be established.

According to Eq. 1, there are many errors affecting the positioning accuracy, such as the attitude and orbit errors in measurement, installation errors and CCD distortions. Meanwhile, the errors of attitude and orbit are identical for dual-cameras since the two cameras are mounted on the same satellite, but installation errors and CCD distortions of dual cameras are different. Therefore, it is essential to recover the relative geometry relation of dual cameras to calibrate installation errors and CCD distortions of each camera.

Previous studies have demonstrated that orbit errors are equivalent to attitude errors under the condition of high flight altitude and narrow field of view [7]. In other words, it is unnecessary to model the orbit errors, for calibration. In addition, installation angle errors are equivalent to attitude errors in geometric positioning. Therefore, just how attitude errors affect geometric positioning is needed to be analyzed.

In the dual camera case, both cameras have independent interior orientation parameters. These parameters can be calibrated by on orbit geometric calibration method. The interior orientation of camera is considered relative stable on orbit. The change of interior orientation is measured by period of month. Therefore, in short a short time, the effect of interior orientation can be omitted.

For the exterior orientation, both cameras are mounting on the same structure. Therefore, their relations with satellite platform are same. The difference of two cameras lies in the geometric relation between mounting structure and camera, which is desired in this study. This relation can be obtained by comparing the geographic coordinates in the overlapping area. The positioning error Δy across the track can be calculated:

$$\Delta y = \frac{f}{\lambda_{ccd} \cos \psi} \Delta \omega \tag{2}$$

where λ_{ccd} denotes the size of the CCD detector. As shown in Eq. 2, positioning errors caused by roll errors are related to pixel's view angle ψ . It can be comprehended that a roll angle error may cause varying positioning errors at different detectors. Similarly, It can also illustrate the positioning errors caused by the pitch angle error, which can also

be regarded as a translation error according to the above derivation. The yaw angle error causes positioning errors modeled as rotation errors:

$$\begin{aligned} \Delta x &= s(1 - \cos(\Delta\kappa)) \\ \Delta y &= s \sin(\Delta\kappa) \end{aligned} \tag{3}$$

where s denotes the image row and $\Delta\kappa$ is the yaw angle error.

The CCD distortions of each linear push-broom CCD array can be compensated by following model: [8]

$$\begin{cases} \tan(\psi_x) = a_0 + a_1s + a_2s^2 + \dots + a_5s^5 \\ \tan(\psi_y) = b_0 + b_1s + b_2s^2 + \dots + b_5s^5 \end{cases}, i, j \leq 5 \tag{4}$$

where $\tan(\psi_x) = \frac{x-x_0-\Delta x}{f}$, $\tan(\psi_y) = \frac{y-y_0-\Delta y}{f}$

If Eq. 4 is adopted to compensate for CCD distortions, a_0 and b_0 can eliminate both pitch and roll errors in installation angles. In addition a_l and b_l can compensated for yaw errors in installation angles. In other words, Eq. 4, which determines the relative geometric relation of dual-cameras, can eliminate both installation errors can CCD distortions of each camera. Furthermore, the offset matrix R_u can be used to eliminate attitude and orbit errors in the measurement thanks to the equivalence between orbit errors and attitude errors. So the relative calibration model can be written as follow:

$$\begin{cases} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_A = \begin{pmatrix} X_S \\ Y_S \\ Z_S \end{pmatrix}_A + m_A \left(R_{orbit2wgs84} R_{body2orbit} R_U R_{camera2body} \begin{bmatrix} a_0 + a_1s + a_2s^2 + \dots + a_5s^5 \\ b_0 + b_1s + b_2s^2 + \dots + b_5s^5 \\ 1 \end{bmatrix}_{CCD_m} \right)_A \\ \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_B = \begin{pmatrix} X_S \\ Y_S \\ Z_S \end{pmatrix}_B + m_B \left(R_{orbit2wgs84} R_{body2orbit} R_U R_{camera2body} \begin{bmatrix} c_0 + c_1s + c_2s^2 + \dots + c_5s^5 \\ d_0 + d_1s + d_2s^2 + \dots + d_5s^5 \\ 1 \end{bmatrix}_{CCD_m} \right)_B \end{cases} \tag{5}$$

where R_+ and a_i, b_j, c_k, d_l ($i, j, k, l \leq 5$) are the parameters to be solved CCD_m denotes the index of CCDs. It's notice that a_i, b_j, c_k, d_l should be solved for each CCD unit in dual-cameras. After calculating a_i, b_j, c_k, d_l ($i, j, k, l \leq 5$) of cameras A and B using GCPs, the relative geometric relation of dual-cameras can be recovered precisely. As the result, the further compare of dual-camera pointing direction difference is carried on with relative relation from one calibration.

3.2 Pointing Direction Difference Analysis Procedure

The method proposed by this paper is based on matching points in overlapping area of two images. Then calculate the geographic positions of each matching points with interior orientation model of each camera. By comparing the geometric positions, the pointing direction of each camera and relative difference is obtained.

The procedure of this method is shown in Fig. 2:

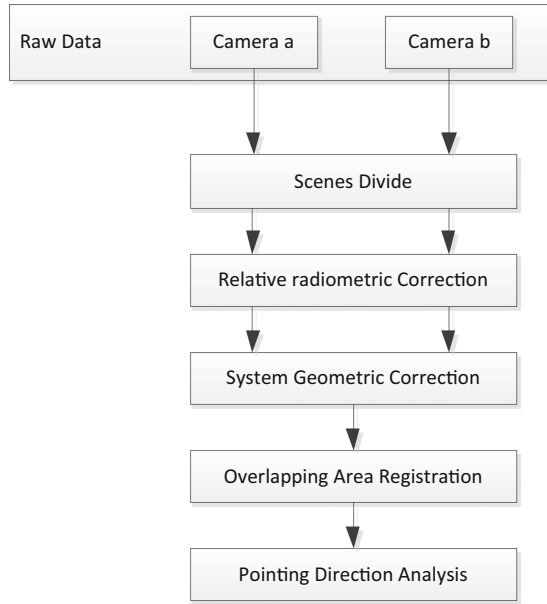


Fig. 2. Procedure of dual camera pointing direction analysis

Firstly, the raw data from both Camera a and Camera b are divided into a series of scenes. This is only to reduce the data processing in future steps. The scene divides based on the image time in auxiliary data. This is to make two scenes can have as many as matching points as possible [9].

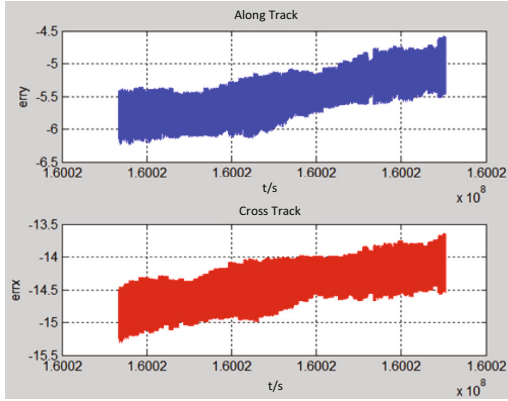
Secondly, the data is processed with standard relative radiometric correction and system geometric correction. Radiometric correction is to eliminated the radiometric different between pixels, improve the success matching chance. System geometric correction use both exterior and interior orientation model. As explained in last section, this step is to eliminated any influence from satellite platform, and relate the pixel to pointing direction of each camera.

Third step is to overlapping area registration. This paper employs the high density registration method to identify as many matching points as possible. And influence related registration and least square are employed to eliminate the matching points with larger error. Finally, matching points with RMS error less than 0.5 pixels are obtained.

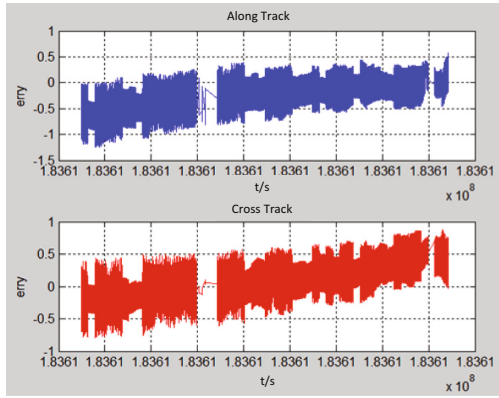
Pointing direction analysis is implemented to each matching points. The difference between dual camera pointing directions is calculated use followed equations.

$$\begin{cases} \Delta x_k = x_{ik} - x_{jk} \\ \Delta y_k = y_{ik} - y_{jk} \end{cases} \quad (6)$$

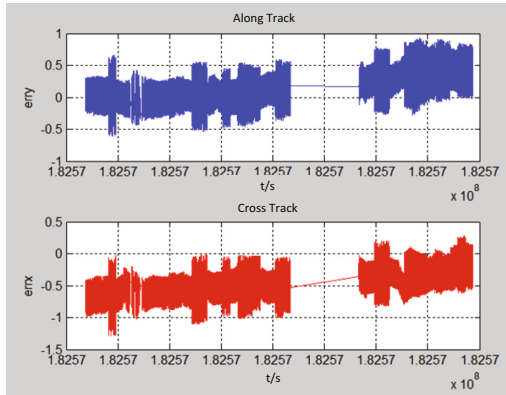
Cross reference the position of matching points and their geographic coordinates to obtain the pointing direction difference of two cameras.



(a) Pointing difference of image in orbit 2383



(b) Pointing difference of image in orbit 6247



(c) Pointing difference of image in orbit 6424

Fig. 3. The pointing direction difference from different orbit

4 Experiment and Result

To verify this method, dual-camera images of one Chinese remote sensing data from several orbits are analyzed. The data are from 2383, 6247, 6424 orbit. The exterior and interior orientation models are from calibration based on data from 6424 track. Following the procedure described above, the dual camera pointing direction differences are calculated and shown in Fig. 3.

The direction difference of two cameras is shown in along-track and across track. The differences are valued by pixels. There are significant data vacuum in 6247 and 6424. This is due to large area of cloud in the image which makes it impossible to find matching points. Nevertheless, it is obvious that the shift of pointing direction is constant change during image. There are changes in both directions in the field of view. This experiment demonstrated that this method can calculate pointing direction difference of dual cameras. The results of this analysis can help to determine the reason for direction change in dual cameras and develop method to compensate for this change.

5 Conclusions

In this paper, a method for calculating the pointing difference of dual camera is demonstrated. Based on structure of one satellite's dual camera mounting property, this method can analyze the pointing direction difference without help of GCPs. As the result, this method can apply to any image at any length. It can help to determine the reason for direction change in dual cameras and develop methods to compensate for this change.

References

1. Tang, X.M., Hu, F., Wang, M., Pan, J., Jin, S.Y., Lu, G.: Inner FoV stitching of spaceborne TDI CCD images based on sensor geometry and projection plane in object space. *Rem. Sens.* **6**, 6386–6406 (2014)
2. Zhang, G., Liu, B., Jiang, W.S.: Inner FOV stitching algorithm of spaceborne optical sensor based on the virtual CCD line. *J. Image Graph.* **17**, 696–701 (2012)
3. Poli, D., Toutin, T.: Review of developments in geometric modeling for high resolution satellite pushbroom sensors. *Photogram. Rec.* **27**, 58–73 (2012)
4. Radhadevi, P.V., Muller, R., d'Angelo, P., Reinatz, P.: In-flight geometric calibration and orientation of ALOS/PRISM imagery with a generic sensor model. *Photogram. Eng. Rem. Sens.* **77**, 531–538 (2011)
5. Breton, E., Bouillon, A., Gachet, R., De lussy, F.: Pre-flight and in-flight geometric calibration of SPOT5 HRG and HRS images. *Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci.* **34**(Part 1), 20–25 (2002)
6. Furukawa, H., Jeminous, M., Cui, G., Laks, H., Sen, L.: In-flight CCD distortion calibration for pushbroom satellites based on subpixel correlation. *IEEE Trans. Geosci. Remote Sens.* **46**, 2675–2683 (2008)

7. Takagi, M., Shimoda, H.: *Handbook of Image Analysis*. Tokyo Press, Tokyo (2004)
8. Jiang, Y.H., Zhang, G., Tang, X.M., Li, D.R.: Geometric calibration and accuracy assessment of ZiYuan-3 multispectral images. *IEEE Trans. Geosci. Remote Sens.* **52**, 4164–4172 (2014)
9. Leprince, S., Barbot, S., Ayoub, F., Avouac, J.P.: Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. *IEEE Trans. Geosci. Remote Sens.* **45**, 1529–1558 (2007)

Application of Wi-Fi Wireless Network on Attitude and Orbit Control System of Satellite

Jinpeng Wang^{1(✉)}, Yue Wang², Yi Zhan¹, Mingyu Xie¹,
and Jianzhao Ding¹

¹ Beijing Institute of Control Engineering,
No. 104, Youyi Road, Beijing 100094, China
kingofwang@126.com

² China Academy of Space Technology,
No. 104, Youyi Road, Beijing 100094, China

Abstract. This article analyzes the feasibility of applying Wi-Fi wireless network into Attitude and Orbit Control System (AOCS) based on the previous research of commercial wireless network. Moreover, this article analyzes the wireless transmission delay of WI-FI. And analyzed results are verified by running a specified experiment. In addition, this article proposes a new scheme of wireless AOCS and develops an AOCS with Wi-Fi network by employing modified spacecraft products with six newly developed modules. Furthermore, the simulation results are illustrated and critically analyzed.

Keywords: Wi-Fi wireless network · Attitude and Orbit Control System Satellite

1 Introduction

The continuing development on aerospace technology and thousands of application on ground wireless technology make it possible for wireless network to be used in AOCS of Spacecraft. AOCS within wireless network is more flexible in data transmission and data interaction. In addition, by employing wireless transmission technology, transmission cabling can be dramatically decreased which will lead to weight reduction of satellite and also make the ground tests much easier. Hence, the characteristics described above make the application of Wi-Fi wireless network on AOCS of satellite have the advantage of rapid reconfiguration, agile manufacturing and dynamic expansion.

Generally, Intra-Spacecraft wireless network consists of optical wireless networks and Radio Frequency (RF) wireless network. Compared to optical communication technology, RF communication technology utilizes less equipment and has higher technology maturity. According to the report published by Consultative Committee for Space Data Systems (CCSDS) in 2009 emphasizes, the application of wireless technology in aerospace becomes the trend in the field of information exchange [1]. Spacecraft Onboard Interface Services Area (SOIS), a branch organization of CCSDS,

has published the Green Book named ‘Wireless Network Communication Overview for Space Mission Operation (800.G-2)’, which summarized the development and application of wireless technology in Intra-Spacecraft.

Nowadays, researchers and engineers from the aerospace field believe that the satellite control system could be more flexible under lower cost by utilizing the software and hardware system from the Commercial Off-The-Shelf (COTS) [2]. Therefore, application of commercial protocol on aerospace research could improve the performance and reliability, and also be beneficial to the updating and development of communication standards.

The essential factors that need to be considered in choosing and evaluating communication protocols of AOCS include: real time performance, data transmission speed, reliability, networking capacity, transmission range, power consumption, and system complexity, etc. The oversea institutions, including NASA [3, 4], Tokyo Denki University [5, 6], Delft University of Technology [7], University of Surrey [8], and the domestic institutions [9] use protocols, such as Zigbee, Bluetooth and ultra wideband (UWB), to carry out the research on the application of wireless technology in satellites. Haerbin Institute of Technology uses the user-defined protocol with commercial wireless device nRF2041 as the way to implement communication among products of the control system. And transmission speed is 250 kbps [10, 11].

IEEE 802.11 combines the technologies of Multi-Input Multi-Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) to increase the transmission rate of WLAN up to 600 Mbps. Compared to Zigbee and Bluetooth, the Wi-Fi wireless network has a higher transmission rate and has no limitations in device capacity in the cellular network, which makes it more suitable in dealing with large data transmission in AOCS.

The paper is organized as follows: In Sect. 2, the Wi-Fi transmission delay is introduced, analyzed and verified by running an experiment. The scheme of wireless AOCS is described in Sect. 3. And tests on a working wireless AOCS are presented in Sect. 4 with simulation results illustrated. Section 5 summarized the paper and simply presented future work research on wireless AOCS.

2 Analysis and Verification of Wi-Fi Transmission Delay

2.1 Analysis of Wi-Fi Transmission Delay

Because the existing wireless communication standards currently do not consider the real-time characteristic of signal transmission, the researcher analyzed and tested the transmission delay before constructing the wireless control system.

Figure 1 illustrates the Distributed Coordination Function (DCF) accessing mechanism based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) with Request to Send (RTS)/Confirm to Send (CTS).

When the data frames is 64 kBytes and the wireless transmission rate is 150 Mbps, the time needed for completing a handshake protocol is:

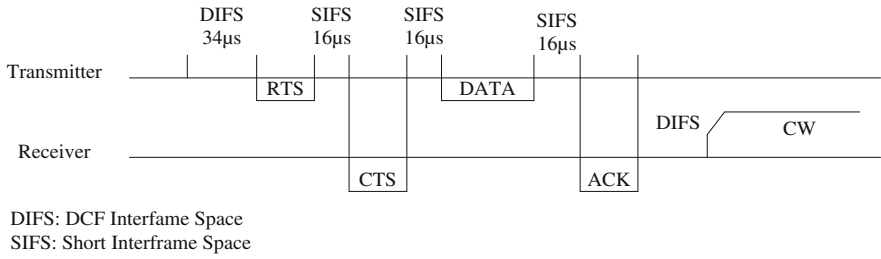


Fig. 1. DCF accessing mechanism based on CSMA/CA

$$T = DIFS + SIFS * 3 + \frac{64\text{ kBytes} + 112\text{ bits} + 160\text{ bits} + 112\text{ bits} + 192\text{ bits} * 3}{150\text{ Mbps}} = 3501.73\ \mu\text{s} \quad (1)$$

When the data frames is 0 and the wireless transmission rate is 150 Mbps, the minimal time needed for completing a handshake protocol is:

$$T = DIFS + SIFS * 3 + \frac{0 + 112\text{ bits} + 160\text{ bits} + 112\text{ bits} + 192\text{ bits} * 3}{150\text{ Mbps}} = 88.4\ \mu\text{s} \quad (2)$$

2.2 Verification of Wi-Fi Transmission Delay

Connecting Host A and Host B with wireless network adaptor, where Host A is the Access Point (AP) terminal and Host B is the client terminal, the researcher has

No.	Time	Source	Destination	Protocol	Length	Info
99	21.495789	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
100	21.495712	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
101	21.497769	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
102	21.497788	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
103	21.497788	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
104	21.497792	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
105	21.497798	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
106	21.497802	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
107	21.497805	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
108	21.497809	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
109	21.497812	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
110	21.499879	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
111	21.499883	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
112	21.499897	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
113	21.499901	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
114	21.499905	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
115	21.499909	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
116	21.499912	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
117	21.499916	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
118	21.499920	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
119	21.499924	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
120	21.502378	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
121	21.502392	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
122	21.502396	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
123	21.502400	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
124	21.502404	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi
125	21.502408	192.168.1.100	192.168.1.103	UDP	74	Source port: 57462 Destination port: irdmi

▶ Frame 33953: 74 bytes on wire (592 bits), 74 bytes captured (592 bits)
 ▶ Ethernet II, Src: Dell_cc:05:76 (18:03:73:cc:05:76), Dst: AskeyCom_df:b6:4d (08:24:d2:df:b6:4d)
 ▶ Internet Protocol Version 4, Src: 192.168.1.100 (192.168.1.100), Dst: 192.168.1.103 (192.168.1.103)
 ▶ User Datagram Protocol, Src Port: 57462 (57462), Dst Port: irdmi (8000)
 ▶ Data (32 bytes)

Fig. 2. Transmission time of signals from Host A to Host B

monitored the interval between data packages when AP terminal was transmitting data package in 32 bytes under UDP protocol. The transmission time is illustrated above in Fig. 2.

When data packages in 32 bytes are being transmitted continuously, the time delay of each data packet is less than $10\ \mu\text{s}$ in some time sequences (for example, the time delays between the time sequences (1) 102 and 103, (2) 104 and 105, (3) 105 and 106, are $5\ \mu\text{s}$, $4\ \mu\text{s}$, and $6\ \mu\text{s}$, respectively). However, in another time sequence, a long time delay may occur (for instance, $2\ \text{ms}$ – $3\ \text{ms}$ in time sequence 101 and 120). The average delay is around $257\ \mu\text{s}$ which can satisfy the requirement of prototype system that no longer than $1\ \text{ms}$. However, it is still slightly higher than the theoretical analytical data.

3 Scheme of Wireless AOCS Based on WIFI

Figure 3 illustrates a traditional wired satellite closed-loop control system, which includes wired AOCS and ground dynamics module. In order to make the satellite closed-loop control system in wireless, it is essential to develop a wireless AOCS.

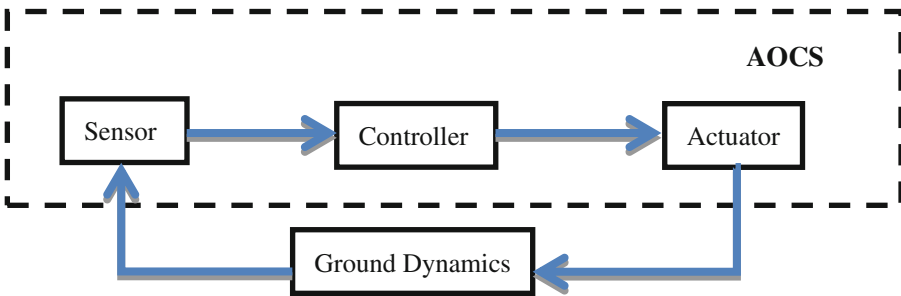


Fig. 3. Wired Satellite Closed-loop Control System

Currently, the frequency of data interaction between Attitude & Orbit Control Computer (AOCC) and sensor/actuator is $4\ \text{Hz}$ when an AOCS of satellite utilizes cables to transmit signals. The topology of wired AOCS is shown in Fig. 4. In addition, the attitude stability of satellite is smaller than $5\text{e-}4\ \text{degree/s}$, which can satisfy the attitude stability requirement of most satellites. By modifying the AOCS with employing wireless configuration, feasibility of wireless network application on satellite control system can be verified.

Electronic products utilized in satellite control system not only have interfaces connected with each single unit on satellite, but also have interfaces connected to ground-test system for implementing closed-loop tests. Authors hope to build an entire wireless close-loop testing system which leads to each equipment become wireless so that the on-satellite system and satellite-ground system in the future can be made in wireless. Therefore, the entire wireless data transmission system works as two wireless networks: “wireless network 1” transmits data between the satellite and the ground (so

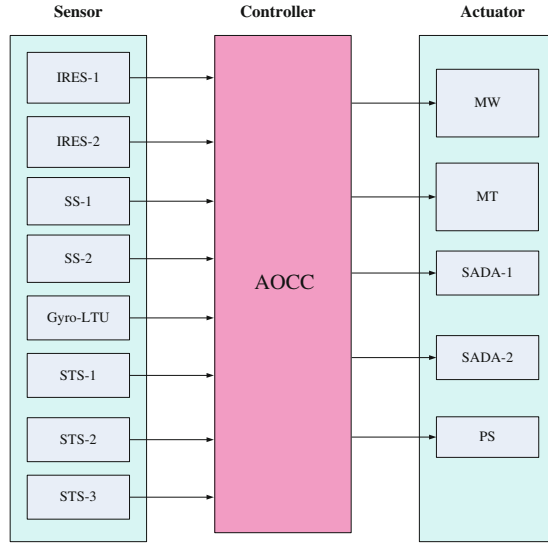


Fig. 4. Configuration of AOCS on a satellite

called satellite-ground communication), while “wireless network 2” transmits data among internal electronic products of the satellite (so called on-satellite communication). Connection of devices in the system is shown in Fig. 5.

The wireless system consists of six newly developed modules: ground-controlled wireless transmission module, satellite-controlled wireless transmission module, on-satellite Infrared Earth Sensor (IRES) wireless transmission module, on-satellite

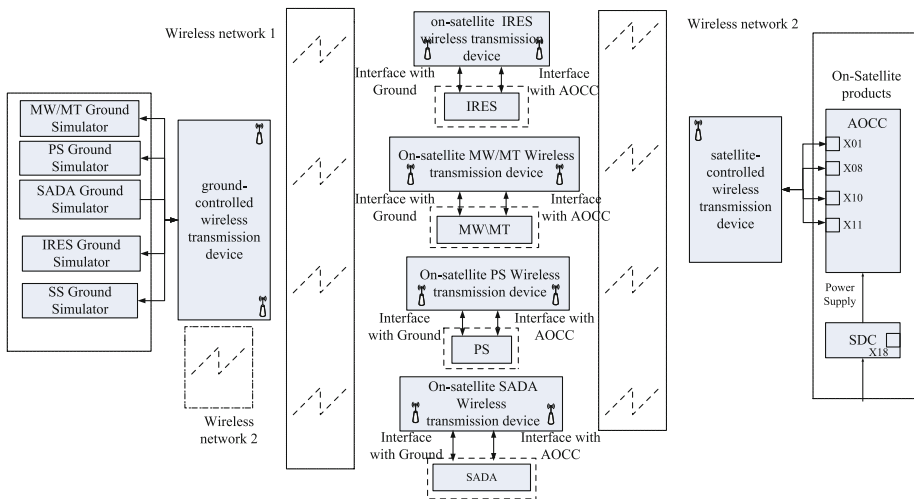


Fig. 5. Connection of devices in the system of AOCS with wireless network

Momentum Wheel (MW)/Magnetic Torque (MT) wireless module, on-satellite Propulsion System (PS) wireless transmission module and on-satellite Solar Array Drive Assembly (SADA) wireless transmission module.

Signals transmitted among internal electronic products of the satellite are presented as follows: 7 analog signals, 12 RS422 signals, 12 pulse signals, and 10 BOOL signals. Analysis shows that effective data transmission speed required by the on-satellite communication network should be no smaller than 500 kbps.

Signals which transmitted between the ground and the satellite are of the following types: 3 analog signals, 12 RS422 signals, 16 pulse signals, and 12 BOOL signals. In order to guarantee that the closed-loop test works properly, the data transmission speed of satellite-ground wireless network should be greater than 500 kbps.

In this application, The ATHEROS AR9331 is a highly integrated System-on-Chip (SoC) for wireless local area network (WLAN). In a single chip, the AR9331 integrates a MIPS 24 K processor and an 802.11n 1 × 1 MAC/Baseband/radio with internal power amplifier and low noise amplifier. It supports 802.11n operations up to 150 Mbps for 40 MHz channel.

Figure 6 illustrates the connection between on-satellite products and satellite controlled wireless transmission device. Combination of these two modules become a wireless AOCC, which is one part of wireless AOCS. As shown above, the black box is AOCC which is the center of control system, and the white one is satellite controlled wireless transmission device which is illustrated in Fig. 5. Two green cables are connected to the AOCC as the power supply, while the information signals are transmitted to the satellite controlled wireless transmission device by employing five white cables. The two blue 1553B links connect AOCS with other sub-systems for telecommunication; however, these two cables do not belong to the closed-loop control system.

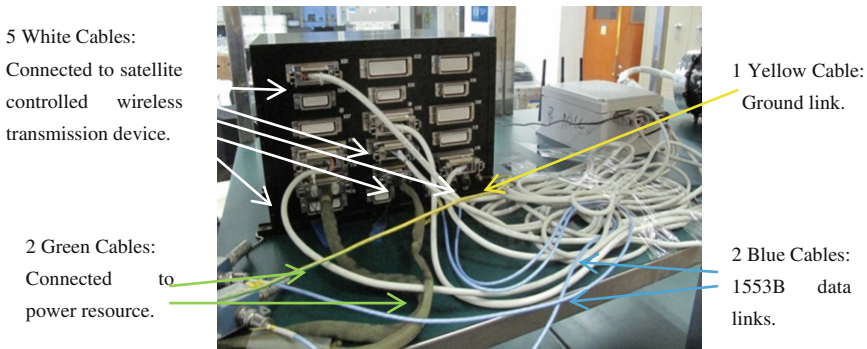


Fig. 6. How the cables are connected after the AOCC is made in wireless

4 Tests Based on Wireless Network

As illustrated in Figs. 7 and 8, the angular velocity of AOCS with Wi-Fi employed is about $5e-4$ degree/s. It is higher than that in AOCS with wired transmission configuration, which is around $2e-4$ degree/s. This degradation of performance is mainly caused by signal distortion and time delay during the wireless transmission.

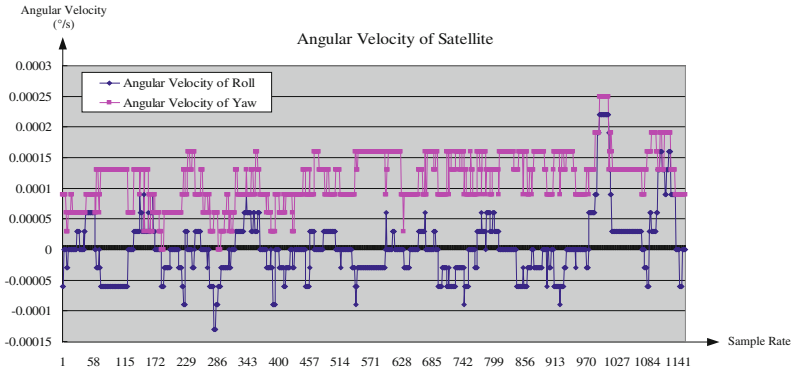


Fig. 7. Angular velocity curve with wired transmission

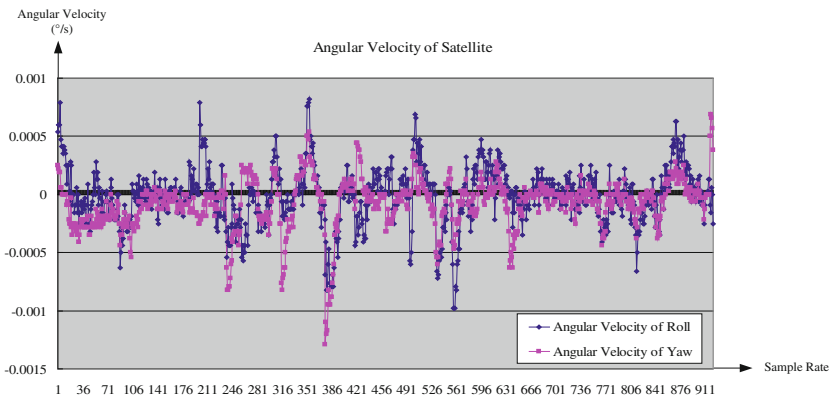


Fig. 8. Angular velocity curve with wireless transmission

As shown in the above curves resulting from running closed-loop test, the system can operate properly after modifying the existing system by employing 802.11 into signal transmission link circuit.

5 Summary and Future Work

This article has proposed a wireless AOCS with WI-FI transmission technology, which contributes to the building of an entire wireless satellite closed-loop control system. Employing WI-FI technology has been analyzed and verified as a feasible solution for constructing an ideal wireless satellite closed-loop control system. Simulation results obtained from the newly developed wireless AOCS have shown a good performance that can satisfy the stability requirement of current mainstream satellite.

The two important indexes of wireless network in control system are discussed in this article, which are transmission rate and time delay, respectively. By analyzing the indexes above, it has shown that 802.11n can satisfy the requirement of prototype system on attitude stability in the mainstream short distance wireless transmission technologies. The size of 802.11 wireless module is around 2 cm*2 cm. The stand-by power consumption is around 2 mW, while its working power consumption is less than 0.5 W. The customer can not only find industrial level of 802.11 chips in the market, but also buy military level of chips provided by Broadcom and TI, which makes it practical to choose this technology.

However, there are still some problems in the special circumstances of space technology that need to be resolved in applying wireless technology into satellite control system. For example, the signal fadeout and signal interference, hidden terminals and exposed terminals among the joints of the wireless internet. These problems can lead to future work on optimization of network joints, topology configuration and network protocol.

References

1. Rodger, M., Patrick, P., Inma, H.: Current activities and status of the ESA/ESTEC-industry wireless onboard spacecraft working group. In: The 4th Space Internet Workshop SIW-4, USA, pp. 41–47 (2004)
2. Thom, S., Richard, A., Jaren, B., et al.: A viable COTS based wireless architecture for spacecraft avionics. IEEE (2012). ISBN: 978-1-4577-0557-1
3. NASA/CANEUS Workshop, NASA/JSC/ES6/Veorge Studor, Fly-by-Wireless A Revolution in Aerospace Architectures for Instrumentation and Control, 27 March 2007
4. Robert, M.: Avionics Division, NASA JSC, NASA Bluetooth Wireless Communications, 28 May 2007
5. Shinichiro, H., Atsushi, T., Tomoaki, T., Takehiko, K.: Wireless Connections within Spacecrafts to Replace Wired bus. IEEE (2013)
6. Akihisa, M., Atsushi, T., Tomoaki, T., Takehiko, K.: Experimental evaluation of ultra wideband propagation and transmission within a spacecraft for replacing wired interface buses. In: Loughborough Antenna & Propagation Conference (2010)
7. Rouzbeh, A., Eberhard, G., Georgi, G.: The challenges of intra-spacecraft wireless data interfacing. In: 58th International Astronautical Congress, pp. 56–63 (2007)
8. Jean, R.P., Tanya, V.: Design of a wireless link for SpaceWire networks. In: SpaceWire-Conference (2010)

9. Hexiang, T., Zheng, Y., Bin, L., Xiangqi, Z., Youjun, L.: Feasibility test and analysis of BlueCore chip in space application. *Transducer Microsyst. Technol.* **28**(4), 63–71 (2009)
10. Limin, D.: Research on Inter-Satellite/Intra-Satellite Communication Technology, Ph.D thesis. Harbin Institute of Technology (2012)
11. Zhaowei, S., Lei, X., Guodong, X., Guocheng, F.: Wireless RF bus design for an intra-satellite. *J. Harbin Eng. Univ.* **33**(7), 881–886 (2012)

Design and Implementation of the GEO Remote Sensing Satellite for Intelligent Applications

Fengjing Liu^(✉), Ning Liu, Xiang Li, and Guo Li

Beijing Institute of Spacecraft System Engineering, Beijing 100094, China
13661045050@qq.com

Abstract. The orbit control of remote sensing satellites is the core part of satellite system application. Its performance directly affects the efficiency of satellite system. Aiming at studying the characteristics of Geosynchronous (GEO) remote sensing satellite imaging, this paper proposes the scheme for mission planning and visualization of remote trajectory remote sensing, which is based on in-depth analysis of mission planning objectives and modeling requirements. It adopts the user oriented control system architecture and branch and bound algorithm to complete the mission planning scheme. By using 2D and 3D visual presentation environment, the proposed scheme can display the result of the mission planning. Finally, an application example is given to evaluate the performance of the system. The evaluation results show that the planning scheme is fast and effective for satellite on orbit applications.

Keywords: Mission planning · Visualization · Control system

1 Introduction

With the development of space technologies and the increasingly frequent space activities, how to complete the task space with low cost, low risk and high efficiency has become an important topic of research in the field of aerospace [1]. Among the key technologies in this field, the earth observation task planning for remote sensing satellite is based on the target characteristics, satellite payload characteristics, satellite platform and load requirements and other constraints. It adopts a certain optimization strategy to determine the work mode, work sequence and ground resources work plan of the satellite, and achieves the optimal allocation of satellite and ground resources [2]. Therefore, the rationality and correctness of the satellite mission planning program will be directly related to the rational utilization of satellite resources in the full application process of the satellite on orbit.

With the extensive application of visualization technology and cross-platform interaction technology in computer, it is necessary to develop an intelligent control system for satellite applications. Based on the characteristics of remote sensing satellite imaging, this paper studies the task planning method based on branch and bound algorithm, and develops the task planning and control system based on STK (Satellite Toolkit), which lays the foundation for satellite on-orbit operation simulation,

verification and on-orbit control. Based on the STK software, the mission planning is conducted according to the practical tasks, the utilization of satellites, constraints, and task scheduling and planning algorithms. By the rapid and convenient implementation of the satellite observation mission, the proposed scheme will be able to give a more intuitive picture of the satellite's mission performance on the orbit, giving the user and policy makers a better visual understanding.

2 System Characteristics and Modeling Requirements Analysis

The most important feature of geosynchronous remote sensing satellites is the extremely high time resolution. It must respond quickly to user's emergency ground observation mission. Therefore, the traditional control mode based on instruction template is not only complex in operation and control and high in operation cost, but also difficult to meet the corresponding requirements of the rapid task.

The main goal of the mission planning and visualization control system for geosynchronous remote sensing satellite is to meet the relevant constraints according to the task demand, the current state of the system, the environmental conditions and so on. By the reasonable working mode selection, the system should generate an optimized control sequence [3] of actions from the initial state to the target state in the shortest possible time, and convert them into control instructions for satellite applications. At the same time, it should use visualization tools and demonstrate the specific actions of task execution, which can provide effective decision suggestions for decision maker.

Satellite mission planning and visual control system needs to address the following aspects of the problem:

- (1) Mission setting model: The mission setting model for geosynchronous remote sensing satellite includes orbit model, payload model, ground station model and mission information model (single target, multi target and regional target), etc.
- (2) Basic service model: Basic service model includes satellite orbit prediction model, ground light model, satellite access information model and so on. By combining ground observation mission, it outputs the load configuration parameters.
- (3) Using constraint model: It establishes a relatively complete system constraints model, fully expressing clear logical order constraints and satellite utilization constraints among actions of spacecraft, and expressing the knowledge in machine language.
- (4) Imaging task sequence: It is calculated by satellite mission planning algorithm. The satellite load switch time and state, sway moment and angle, data transmission time and transmission type need to be visually displayed;
- (5) Correctness test: According to operation rules, logical relations, timing constraints and other requirements of the satellite, correctness test is conducted to verify the result of mission planning and analyze whether the system can correctly implement target imaging.

- (6) Imaging task result: It is calculated by the satellite mission planning algorithm. The task response timelines, imaging results information, target coverage, image stretching and spatial resolution degradation needs to be visualized. It has a certain effect analysis function in addition to quantitative assessment function.

3 The Overall Structure of the System

The mission planning and visual control system for geosynchronous remote sensing satellite are based on STKX technology and C # language. STKX technology is a new development technology provided by STK software. Compared with the traditional STK/Connect module, STK simulation module can be integrate into self-developed software seamlessly without starting STK software. At the same time, it can respond to user's operation, and its interface is clear and convenient for system development [4].

The proposed system provides a convenient and friendly mission-oriented control interface, so that the user can directly input the required imaging tasks without having to pay attention to satellite implementation details. According to the imaging task plan developed by user, the system generates task execution sequence and related parameters by using mission planning algorithm and combining scene library with the model library, Then the implementation process of the satellite mission and imaging results are displayed in 3D or 2D visual display, which will assist the mission arrangement staff to make final decision. After the completion of the task, the instruction block is output for direct application of the satellite.

4 Mission Planning

4.1 Mission Planning Process

Mission planning of remote sensing satellite is usually divided into four phases [5]. In the task pre-processing phase, the main task is to decompose the user task requirements express them in a unified form; In the model construction stage, the mission planning objective function and constraints are determined according to the specific requirements to construct task planning model; In the algorithm solving phase, it is based on the task planning model, combined with intelligent solution algorithm, to generate task planning schedule; The last stage is mainly to evaluate the mission plan and imaging effect. If the scheme is reasonable and feasible, the system will output the whole task sequence.

Imaging model of geosynchronous remote sensing satellites is different from traditional low earth orbit remote sensing satellite. As its orbit is still relative to the ground, it's more suitable to use staring imaging. Therefore, the core feature of mission planning is to map observation area represented by latitude and longitude into effective view plane suitable for satellite observation to form mapping image through intelligent mathematical modeling, and use the image from payload view to form a coverage matrix divided by payload view graph. Then the system determines whether each unit of coverage matrix need to achieve coverage through the geometric method. According

to satellite imaging ability, attitude maneuverability, and data transmission capability, the process is based on branch and bound algorithm to search and counts imaging plan, flexibility plan and antenna transmission plan in accordance with the optimization indicators and use constraints (such as load utility constraints, data balance constraints and energy balance constraints, etc.). Finally the work sequence and related evaluation results are generated from the system. The key process of the main algorithm is shown in Fig. 1. Single target area imaging mode represents imaging of a single continuous area target, and multiple target patrol mode represents imaging of multiple discontinuous area targets separately.

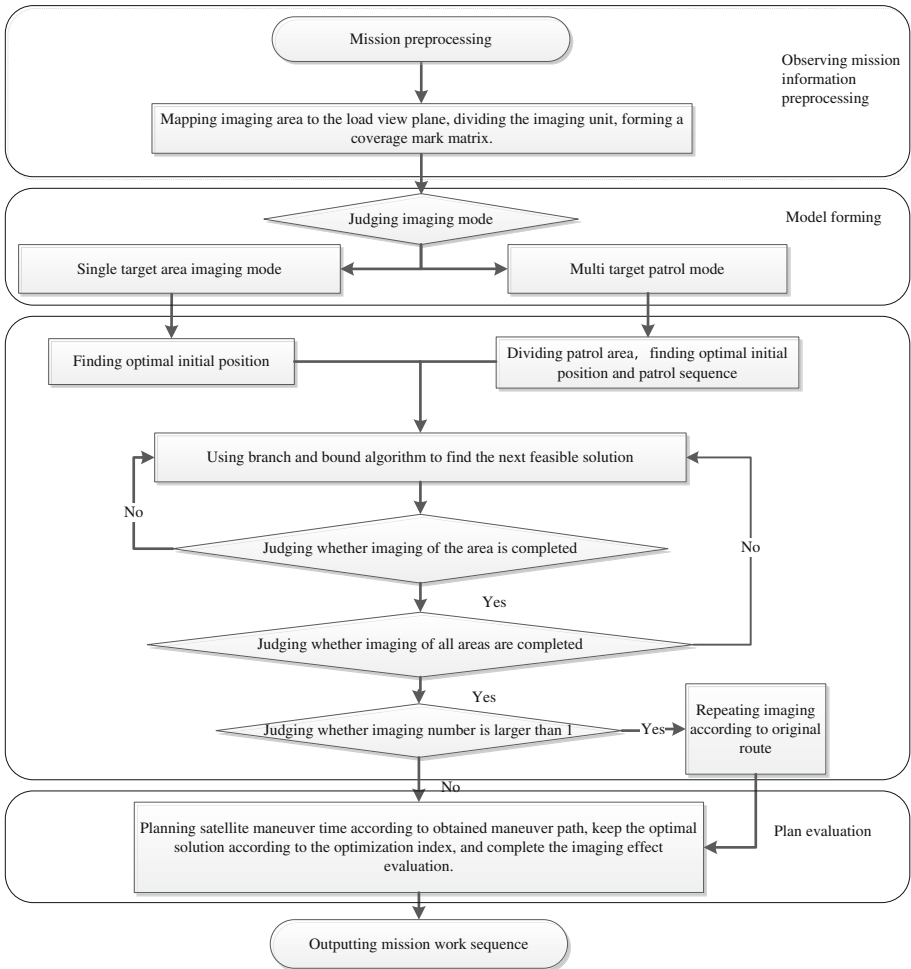


Fig. 1. The proposed mission planning algorithm for geosynchronous remote sensing satellites

4.2 Example of Mission Planning

Taking imaging task of Hainan Island as an example, concerning the illustration convenience and confidentiality requirements, the instruction sequence of the task plan is illustrated in Fig. 2, where the black border area is a user-specified imaging task, the red curve is the maneuvering path of satellite after mission planning and blue square is load over ground when imaging after mission planning.

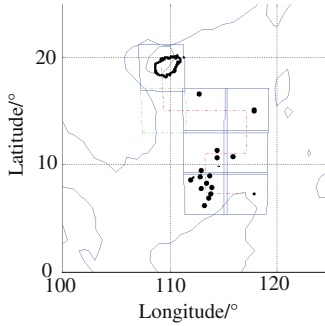


Fig. 2. Example I of mission planning results of Hainan island

5 Visual Display Design

The visualization program for mission planning and visual control system of geosynchronous remote sensing satellite is based on 2D and 3D controls provided by STKX. Their interaction is shown in Fig. 3. The 2D controls (AGI Map Control) embedded in software can respond to all kinds of user operation such as mouse click, mouse movement and so on. It can display the track of sub-satellite point, payload movement trajectory and ground observation of payload, etc., displaying execution process in a more sophisticated way. In addition to the two-dimensional scene, the 3D controls (AGI Globe Control) can also display satellite's attitude change, sunlight illumination situation, etc., which are not able to be displayed in some two-dimensional scenes.

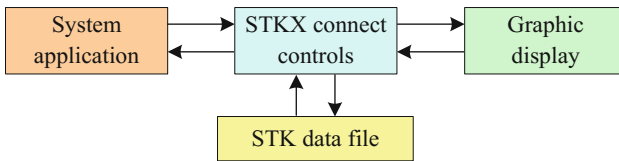


Fig. 3. Interface connection between STK and application

The following mainly describes implementation of key modules for visual display:

(1) Satellite model establishment

In general, in order to enable the developed system to have better and more realistic visual representation of the whole process of task execution, the satellite model is expected to be added in the system. Model file is a hierarchical structure, consists of entities and components. Components contain entities which define component (such as a polygon or a cylinder), parameters that describe certain content (such as color and brightness), or entities that are referenced by other components [6].

(2) Orbital model establishment

Satellite orbital information determines many constraint properties of spacecraft application model design such as mission time window, target visibility, lighting conditions, external interference, etc. These constraints play an important role in satellite design and planning.

(3) Payload model establishment

The payload model is also one of the key core elements of visual display. Key parameters include angle of view, imaging state and other related parameters.

(4) Satellite attitude control design

Satellite attitude control is another key element of visual display. Due to the diversity of user observation tasks, satellite-to-ground observation becomes a very complex maneuvering process. Usually, an observation task needs to set multiple attitude maneuver segments. For this complex process, STK does not provide multiple attitude data control interface, so the usual practice is to use STK/Connect connection interface module. By modifying command parameters, attitude control segment can be added, deleted and modified.

6 System Application Analysis

Based on the above architecture design and key technical analysis, we take an application example to analyze operation of the system after the completion of task planning and visual control system. First, we use hypothetical parameters for system configuration. Main hypothetical parameters of the system are shown in Table 1. Then we input satellite observation task into the system, which is regional imaging for Sichuan Province, Guizhou Province and Qinghai Province at 12:00:00 AM on Dec 10th, 2016. The mission planning results are shown in Table 2. From the analysis of the operation results of the system, it can be seen that all the requirements are met according to the planning objectives, and the ground observation efficiency of the geosynchronous orbit remote sensing satellite has been greatly improved.

Visualized results of the system are shown in Figs. 4 and 5, in which the colored irregular closing curve represents target area for imaging, the blue solid line box indicates ground coverage situation of payload view field during each imaging, and the dark yellow dotted line indicates attitude maneuvering path after satellite planning.

Table 1. System hypothetical parameters

Parameter name	Value
Satellite's fixed point position	104°E
Payload view angle	0.8°
Adjacent image overlapping area	10%
Load data volume	1.6 Gbps/Frame
Data transmission capability	450 Mbps
Attitude manoeuvrability	0.8°/30 s 6°/240 s

Table 2. Application model simulation results

Planned project	Required value	Planned value
Target area	Sichuan Province + Guizhou Province + Qinghai Province	17 images cover three provinces
Imaging time	12:00:00 AM, Dec 10th, 2016	12:00:00 AM, Dec 10th, 2016
Imaging mode	/	Imaging mode: single target area
Mission planning time	Less than 5 min	141.3 s
Mission response time	Less than 5 min	145 s
Imaging duration	Less than 15 min	445 s
Ground station work duration	Less than 5 min	95 s
RAM	Less than 10%	0.11%

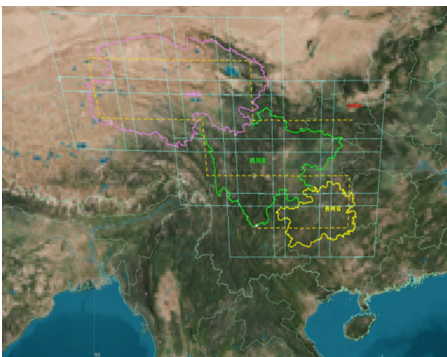


Fig. 4. System planning results visual display (2D)

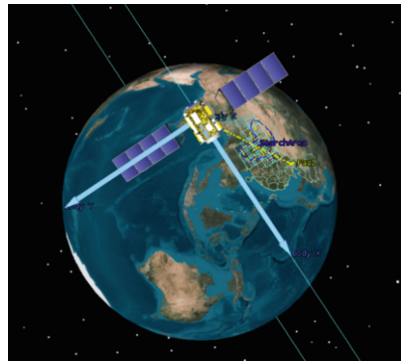


Fig. 5. System planning results visual display (3D)

The yellow flag is the location of ground station, and the yellow dotted line indicates that image data is in transmission, which proves the rationality of the proposed system.

7 Conclusions

In this paper, a mission planning and visualization control system for geosynchronous remote sensing satellite is proposed and developed for mission planning and satellite control of high-orbit remote sensing satellites. This system takes full advantage of STK's model management, computing capability, graphics display and other characteristics. Through the STKX module combined with 2D and 3D scene display, imaging mission plan of satellite can be presented intuitively to staff in form of animation. The task development, instruction arrangement, ground verification integration and visualization are realized accordingly. The system can effectively reduce the difficulty of manipulating complex satellite and improve efficiency and correctness of decision for mission planning.

Combining with current research basis and application requirements, future work will focus on the following aspects. One is to strengthen the study of multi-mission planning algorithms including genetic algorithm, greedy algorithm and tabu search algorithm, while establishing algorithm library to meet the application requirements of different remote sensing satellites. Another is introducing the HLA architecture to enhance the system compatibility for different types of remote sensing satellites.

References

1. Gu, H.: Design on a remote sensing satellite scheduling architecture based on SOA. *Radio Eng.* **44**(12), 43–47 (2014)
2. Zhang, Z.: Multi-satellite control resource scheduling based on ant colony optimization. *Expert Syst. Appl.* **41**(6), 2816–2823 (2014)
3. Jiang, W., Hao, C., Li, Y.: Review of task scheduling research for the earth observing satellites. *Syst. Eng. Electron.* **9**(35), 1878–1885 (2013)
4. Liu, X., Chen, Y.: Application of Kriging surrogate model to optimization of earth observation satellite system. *Acta Automatica Sin.* **38**(1), 121–122 (2012)
5. Zhu, X.: Qos-aware fault-tolerant scheduling for real-time tasks on heterogeneous clusters. *IEEE Trans. Comput.* **60**(6), 800–812 (2011)
6. Li, J., He, C., Guo, Y.: Visual simulation technique for the task planning system of imaging satellites. *Comput. Eng. Sci.* **29**(6), 39–42 (2007)
7. Zhang, W., Chen, J., Shen, L.: A visual simulation system for spaceflight mission based on STK/VO. *Comput. Simul.* **22**(10), 82–85 (2005)
8. Thummala, B.: Spacecraft autonomy. In: *MSc in Astronautics and Space Engineering Individual Research Thesis*, pp. 7–10. Cranfield University (2003)
9. Analytical Graphics Inc. F.: *STK User's Manual*. 1st edn. Analytical Graphics Inc., USA (2003)

Application of Computer Simulation Technology in Observing Effectiveness Analysis of the Satellite

Yunhe Liu^(✉), Fengjing Liu, Kuai Yu, Jian Liu, Yongheng Zou,
and Guo Li

Beijing Institute of Spacecraft System Engineering, Beijing 100094, China
lyh88127678@163.com

Abstract. With the development of computer and information technology, simulation technology has been promoted worldwide and is playing a more and more important role in the space mission. In this paper, the main requirements of spacecraft's observing effectiveness test during the general analysis were first analyzed. And the design of based simulation and test platform based on STK (satellite tool kit) was proposed. By using user-oriented system architecture, building up the high precision coordinate transforming relations and using domain decomposition algorithm, the path planning scheme focused on transformation from geographic coordinate space to the view plane were achieved. And the plan results were further tested using the two and three-dimension visual environment in the typical cases to provide an example for using STK simulation technology in engineering project.

Keywords: STK · Simulation technology · Observing effectiveness analysis

1 Introduction

With the development of space technology, space activities are on the rise. It has become an important subject on how to accomplish space tasks with high efficiency but low cost and risk. The ground simulation and test can meet the above requirements of space technology [1], especially computer simulation technology. With its characteristic of economic efficiency and fast speed, computer simulation technology has gained rapid development and played an irreplaceable role in the past few years. However, most research only focused on the algorithm level. Domestic research on the ground simulation and test are not deep enough [2, 3]. Furthermore, there are few integrated simulations that can combined the work mode design with observing effectiveness analysis.

STK can quickly analyze complex aerospace tasks and provide two and three-dimension visual dynamic scene, charts, reports and many other tools [4, 5]. The basic principle of using STK to simulate the development is to import the STK component into the form control program written in the C# language through the form of the control module. The STK component is connected with the STK simulation software analysis engine, and then the STK simulation scene and data analysis

capabilities integrated into their own software, not only enhance the robustness of the application and the control of the simulation process, but also make the simulation more refined [6].

In this paper, we first analyzed the main simulation requirement of the spacecraft's observing effectiveness and then proposed the simulation and test platform based on STKC and C++ language. And according to different work modes of static orbiting remote sensing satellite, such as regional imaging model and mobile inspection mode, an optimal imaging path display algorithm focused on transformation from geographic coordinate space to the view plane was explored base on the foundation of high precision two-dimension earth observation mode. The simulation of the spacecraft's observing effectiveness of satellite attitude was successfully achieved and the by using the underlying intelligent planning algorithm.

2 Simulation Requirement Analysis of Observing Effectiveness

The observing effectiveness of the spacecraft is closely related to its work mode design. The main goal is to optimize the solution with reasonable work mode to generate an action sequence from the initial state to the target state in shortest time. And the work must meet the task requirement, the current state of the system, the environmental conditions and other relevant elements [7]. With the diversification of the spacecraft's tasks and increase of the restrictions, the difficulties of simulation and test in observing effectiveness has greatly increased. The difficulties mainly include the following two points.

2.1 The Knowledge of the Space System Is Quite Complex

With the development of space technology, the function of the sub-system of the spacecraft is becoming stronger and stronger in order to accomplish complicated tasks. The various application mode leads to the complicated coordination between the sub-systems, which needs to consider the working mode, working status and types and quantity of available resources and so on. It acquires huge and complex knowledge combination to establish a relatively perfect system model on the premises of full consideration of the difference of each sub-system and then describe it. So it will be difficult to test the rationality of the simulation [9].

2.2 The Logical and Time Constraints Between Actions Are Complex

The action set of the satellite is huge resulted from complex task requirement and various work mode. There are definite logistic constraints between each action and developer need strictly following the rules of operation. For example, when it need imaging at a particular location, satellite maneuver should be in place before opening the load for imaging. [10] And meanwhile each action has a time window, ignoring the action of the implementation process will lead to system failure or abnormal. So in the process of modeling and system performance simulation, the logical relationship,

timing constraints, time window between each action must be described as the operational rules.

3 Project Design of Mission Planning Display

Satellite mission of earth observing mainly include the geographical location information of the observation area that required that preliminary definite as the latitude and longitude information collection of the closed area of N , expressed as $[\lambda_l, \varphi_l]$ and the starting point of the observing task. In system simulation, the first step is to process the original data, translate the irregular closed area into the coordinate planes constituted by the pitch angle and the roll angle by geometric relation, divide the complex setting area and use the results as the input of satellite mission planning.

3.1 High Precision Model Setting and Its Coordinate System Conversion

To map the discrete latitude and longitude information of the setting area into the effective view plane and demonstrate the results in the displaying interface, the first step requires building up the transformation relationship between the geocentric longitude and latitude of the observing point T and the Euler angle of the satellite S . Transformation relationship of four coordinate system is involved in it showed in Fig. 1.

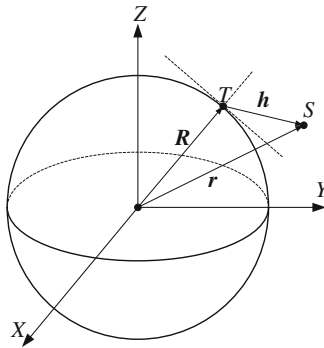


Fig. 1. The geometrical relationship in static satellite imaging

The first step is to build up transformation relationship between the geocentric longitude and latitude (λ_l, φ_l) and the position coordinates $(x_{fix}, y_{fix}, z_{fix})$ of the observing point T in the GEO-fixed coordinate system.

$$\begin{cases} x_{fix} = R_\varphi \cos \varphi_l \cos \lambda_l \\ y_{fix} = R_\varphi \cos \varphi_l \sin \lambda_l \\ z_{fix} = R_\varphi \sin \varphi_l \end{cases} \quad (1)$$

The local earth radius R_φ of the location where the observing point T places shows as follows (Equatorial radius $R_E = 6378.140$ km, earth oblateness $f_E = 0.0033528$):

$$R_\varphi = R_E \frac{1 - f_E}{\sqrt{1 - f_E(2 - f_E) \cos^2 \varphi_l}} \quad (2)$$

Next step is to calculate the location $r(t)$ and speed $v(t)$ of satellite S according to the transformation relationship between inertial coordinate system J2000 and GEO-fixed coordinate system. The location and speed are respectively \mathbf{r} and $\mathbf{v}_i = \mathbf{v} + \omega_e \times \mathbf{r}$. ω_e is the rotational angular velocity vector of the earth. Then the unit vector of the satellite in the three direction of the orbital coordinate system is:

$$\mathbf{u}_z = \frac{-\mathbf{r}}{\|\mathbf{r}\|}, \mathbf{u}_y = \frac{-\mathbf{r} \times \mathbf{v}_i}{\|\mathbf{r} \times \mathbf{v}_i\|}, \mathbf{u}_x = \mathbf{u}_y \times \mathbf{u}_z \quad (3)$$

The orbital coordinate system definite as: z axis is from the center of mass to the center of the earth, y axis is upward perpendicular to the orbital plane, x axis is determined by right-hand screw rule. The coordinate system moves as the satellite and the angular speed is ω_n . The $+x$ axis of the ontology in satellite's effective load imaging directs to the due east and $+y$ axis to the due south, the direction of the $+z$ axis is the same as the position vector \mathbf{h} of the observing point T , $\hat{\mathbf{h}}$ is the unit vector of \mathbf{h} , L_{ob} is the transfer matrix from the ontology coordinate system to the orbital coordinate system. φ, θ, ψ is the Euler angle during satellite observation (respectively rolling angle, pitch angle, yaw angle).

$$\mathbf{L}_{ob} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = [\mathbf{u}_x \quad \mathbf{u}_y \quad \mathbf{u}_z]^T \hat{\mathbf{h}} \quad (4)$$

$$\mathbf{L}_{ob} = \begin{bmatrix} \cos \psi \cos \theta & -\sin \psi \cos \theta & \sin \theta \\ \sin \psi \cos \varphi + \cos \psi \sin \theta \sin \varphi & \cos \psi \cos \varphi - \sin \psi \sin \theta \sin \varphi & -\cos \theta \sin \varphi \\ \sin \psi \sin \varphi - \cos \psi \sin \theta \cos \varphi & \cos \psi \sin \varphi + \sin \psi \sin \theta \cos \varphi & \cos \theta \cos \varphi \end{bmatrix} \quad (5)$$

$$\begin{aligned} \theta &= \arcsin(\hat{\mathbf{h}} \cdot \mathbf{u}_x) \\ \varphi &= \arctan 2(-\hat{\mathbf{h}} \cdot \mathbf{u}_y, \hat{\mathbf{h}} \cdot \mathbf{u}_z) \end{aligned} \quad (6)$$

The longitude and latitude information of toward point of satellites' effective loading can also be calculated in the same method.

3.2 Decomposition of the Observing Area and Calculation of the Symbolic Matrix

Since the target area set by the user cannot be completely covered by the full field of view of the payload, in order to complete the observation task, it is necessary to

segment the regional target first, and divide the large area into small areas with the same pay field. Not only the simple geometric cutting should be taken into account in the design of the regional segmentation method, but also the imaging field of the payload should be satisfied, and take the optimization process of the satellite posture maneuver into consideration. According to the latitude and longitude information collection of the observing area given by the users (λ_l, φ_l) to get the required parameters of every covering area at the beginning of imaging by using effective loading view, definite the pitch angle, rolling angle (θ_0, φ_0) , the field angle in the corresponding direction $(d\theta, d\varphi)$, latitude and longitude informational point (λ_l, φ_l) , the corresponding array (λ, φ) and the identity matrix M . The specific algorithm is as follows:

- (a) According to the above formulas to get the corresponding Euler angle array collection (θ, φ) from the POI(point of information) collection (λ_l, φ_l) . Meanwhile, command $\varphi_0 = \min(\varphi)$, $\theta_0 = \min(\theta)$, $d\varphi = \max(\varphi) - \varphi_0$, $d\theta = \max(\theta) - \theta_0$
- (b) Definite an identity matrix M including N_R line and N_C row, $N_R = \text{ceil}(d\varphi/\alpha)$, $N_C = \text{ceil}(d\theta/\alpha)$ (ceil means the smallest positive integral no less than the result from the value in the bracket). Definite the array M as M_{nk} The coordinates of the four points of M_{nk} which arrange anti-clockwise direction in the square region in coordinate plane.

$$p = \begin{bmatrix} \theta_0 + (k - 1)\alpha & \varphi_0 + (n - 1)\alpha \\ \theta_0 + k\alpha & \varphi_0 + (n - 1)\alpha \\ \theta_0 + k\alpha & \varphi_0 + n\alpha \\ \theta_0 + (k - 1)\alpha & \varphi_0 + n\alpha \end{bmatrix} \tag{7}$$

- (c) Decide whether the unit needs to be covered by Euler angle array collection (θ, φ) . Divide the coordinate plane $\theta-\varphi$ into 9 subdomain by the square boundary line that obtained by central point. Set a identifier f_i to each subdomain, and make sure if the array (θ, φ) is included in any subdomain. If it is included, f_i is 1, otherwise f_i is 0. For example, when $f_5 = 1$, the corresponding area of the matrix unit needs to be covered, that is $M_{nk} = 1$. But if $f_5 = 0$ and the other 8 identifier are all 1, that means the unit is surrounded by the target area, then the unit should also be covered and the term should be added that when $\sum_{i=1}^9 f_i = 8$, $M_{nk} = 1$ (Fig. 2).

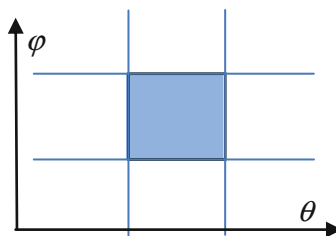


Fig. 2. Division illustration of the coordinate plane $\theta-\varphi$

The divided results of the area are showed as Fig. 3, the black closed area is the imaging area set by the users, the blue square made of the solid line is the unit to be covered, each square represents the width range when the effective loading target the location on the ground.

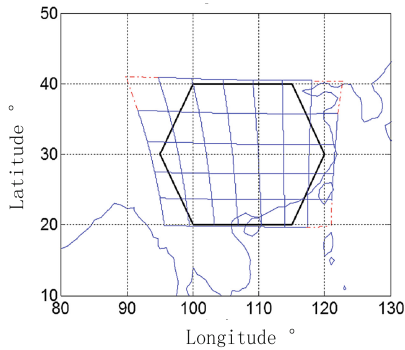


Fig. 3. Division Illustration of the observing area set by users—hexagon

4 Summarization of Simulation Technology Based on STK

4.1 Design of the Task Input Interface

When using STK for spacecraft application mode design and simulation, 2D controls and 3D controls that STK provides can be fully applied as the main application of user input video interface. They can real response to the user's operation and can be directly controlled by the developer. Embedded software in the 2D control (AGI Map Control) can respond to mouse click movement and other user operations, very user-friendly as a simulation task input interface, you can use the world map as the background display user input observation point target, strip target, Target and so on, and have two-dimensional high-precision latitude and longitude point of view, but also shows the satellite star under the trajectory, payload trajectory and payload in the ground observation, so as to more sophisticated control simulation process. 3D control (AGI Globe Control) in addition to the two-dimensional scene can display all kinds of information, but also can display the attitude of the satellite changes, the sun and other circumstances of the two-dimensional scene can't display the information.

4.2 Controls Transfer of STK

When writing in C# language, there are two ways you can control the STK controls. One is the traditional method, which needs using Connect command string and writing Connect command in according to the format specification of the command. In this paper, we mainly recommended another control method, which can be achieved by

using the interface provided by STK. This method is the process making use of convenient, direct and fast logic application programs that can greatly simplify the complex task, but for developers who are not familiar with STK, it will take a lot of time to find related classes and interfaces.

Establishment of the spacecraft object. The information of the satellite orbit determines the various constraints of spacecraft's work mode design, such as task time window, visibility of the target, light condition and external interference. These constraints play an important role in work mode design and simulation, and directly determine the output results. STK has powerful function of orbit analysis and simulation, which can flexibly select the coordinate system to design the initial orbit parameters of the spacecraft, obtain the real-time attitude information of the track in the simulation, and show its flight status in real time.

The focus of creating spacecraft objects is the flexible application of interfaces such as IAgSatellite. The orbital dynamics model is the key to the satellite object. The most commonly used orbit prediction model is the two-body model Two Body and the high-precision orbit prediction model HPOP. When the requirements of the orbit prediction model are high and other effects such as percussion are should be taken into consideration, high-precision orbit prediction model HPOP is usually used.

Establishment of payload model. The payload is a key subsystem of remote sensing satellite's work mode design and simulation, which directly determines the rationality of the work mode design and simulation results. The establishment of the payload is usually arranged after the establishment of the spacecraft model. The core parameters involved include the viewing angle, imaging status, imaging styles and some other related parameters.

Setting of satellite attitude. The key factor of spacecraft's work mode simulation is the attitude control. With the diversification of users' observation tasks, the design of spacecraft application model is more and more complicated. Simple attitude control cannot meet the requirements of intelligent spacecraft. At present, STK cannot provide the planning algorithm for intelligent attitude tasks. Additional development of intelligent attitude planning module is needed to coordinate with STK to complete the simulation demonstration.

Observation of the spacecraft target is a very complex maneuvering process, it will undergo multiple maneuver controls during observation. And usually an observation task requires the setting of multiple gesture segments.

The first step of the process is to receive data of the intelligent attitude planning module, and then use the STK/Astrogator module to perform the simulation calculation of the track, attitude and flight time to generate the simulated flight scene and send the STK-generated simulation data file to simulation platform. The simulation platform takes on all the logic judgment and process control work, and finally will get the sequence of instructions and data passed to the simulation control module to achieve the spacecraft application mode simulation.

5 Performance Analysis Platform Based on STK

In order to simulate the implementation process of the on-orbit mode of the spacecraft in the simulation system, and test the rationality of the algorithm and task design scheme. The system platform should fully integrate the functions of the spacecraft database, mission planning and scheduling, orbit calculation and analysis, scene display, under the command of the simulation control program, according to the actual operation and external environment changes, automatic implementation of the work mode updates, resource attributes change, task scheduling and other functions.

The simulation system developed simulation control program by using Visual C++/C# to integrate the relevant software module [6]. Considering the practical application, the simulation system needs to simulate the whole process of spacecraft mission execution, including user simulation requirements, task and resource definition, task data acquisition, simulation control, scheduling, receiving anomaly, re-scheduling, scene display, data display and analysis, Simulation data records and management and other important links. Application performance simulation system framework were shown as below (Fig. 4).

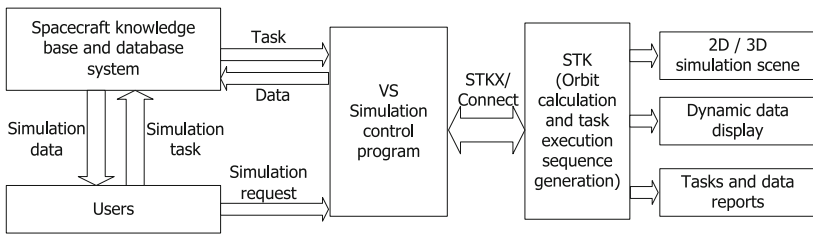


Fig. 4. Task simulation system framework design

The spacecraft knowledge base and the database system mainly store important information such as spacecraft information, load information, intelligent planning algorithm, simulation environment configuration parameters, various tasks and resource data, simulation process data and so on. The control platform integrates the simulation sub-module, carries on the user input information verification, the task execution sequence generation, the simulation process control and the state monitoring, and realized the certain degree of autonomous operation. STK is responsible for orbital computing, operational data analysis, report generation, 2D/3D scene display, task execution sequence simulation. The system can dynamically update the task, change the resource attributes and re-schedule according to the actual operation of the spacecraft and the change of the external environment.

After users' simulation requirements being verified, you can define tasks and resources according to the pre-agreed format, and enter the task data to the database. The control program simulates the environment initialization, reads the task, the resource definition and the attribute from the database, the STK task interaction interface, and uses the background intelligent programming algorithm to generate the

task scheduling execution plan. Finally, STK performs the entire process simulation and stores the task data. When the space environment changes, changes in tasks such as abnormal circumstances, will automatically enter the task and resource verification. The task simulation flow chart was shown as below (Fig. 5).

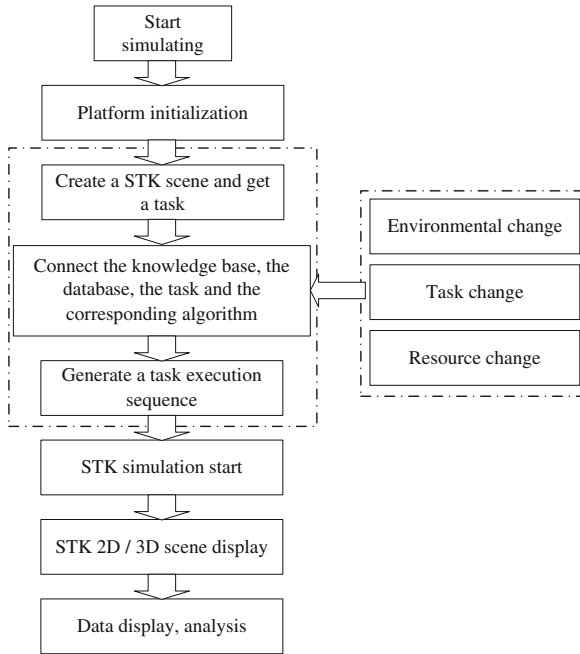


Fig. 5. Simulation flow chart

6 An Example of Observing Effectiveness Simulation

According to the above introduction, using STK technology, through the flexible setting of the spacecraft, payload, attitude control interface and command parameters the observing effectiveness simulation of the spacecraft can be achieved.

In this paper, the algorithm was tested in the imaging tasks of several different areas in our nation (showed in Table 1), and the feasibility and effectiveness was effectively verification. The parameter of the orbit employ the static track of which the eccentricity is 0 and the location of the position point is 108°E, the original directional position is the central position in our country.

The algorithm is accomplished by VC++ and the results of the task planning can be demonstrated visible by using the mode in STK and related components and application programming Interface function (showed in Figs. 4 and 5). The red mark is the original direction of the satellite effective load, and the irregular closed curve in various colors set by users are areas that need imaging, and the light green solid square demonstrates the real visible region covered in the imaging, and the deep yellow dotted line demonstrates the path of the direction of the effective load.

Table 1. Results of the work model simulation

Track type	Spacecraft application mode	Imaging area	Number of images	Total time of imaging	RAM	Platform running time
Sun synchronous return orbit	Multiple with splicing mode	150 km × 70 km	15 Bands	240 s	0.10%	45.95 s
Geostationary orbit	Regional imaging mode	Sichuan Province + Guizhou Province + Qinghai Province	17 Amplitude	445 s	0.11%	21.36 s

As showed in Table 1, the simulation results of the task planning can be showed in the dimensional, three-dimensional display system in about 10 s when the imaging area set by users is small. And the resource occupied rate is very small. Compared with this, if the imaging area selected by users is relatively dispersed, it will require more time and resources. Even though the resources occupied by the computer will increase by linear growth, and the time cost can be accepted (Figs. 6 and 7).

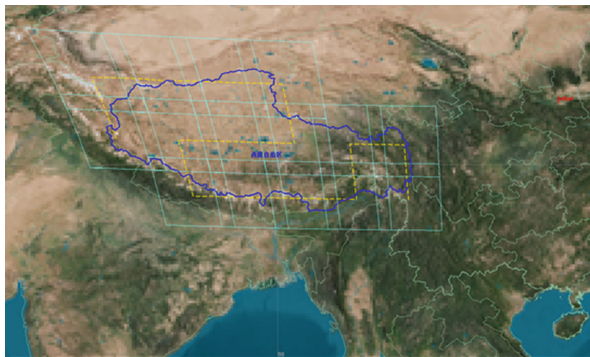


Fig. 6. Results of regional imaging path planning

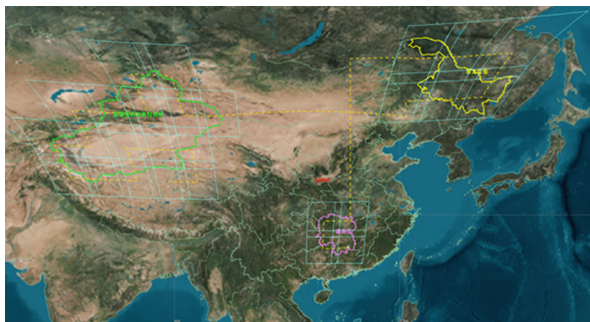


Fig. 7. Results of irregular region

7 Conclusion

In this paper, STK-based simulation technology was studied according to the main requirements of the simulation of the spacecraft's application performance. First, the problem of simulation and test of application performance is analyzed. Second, the method of monitoring STK controls is studied, and the setup of the spacecraft and payload and the control method of spacecraft attitude is analyzed. Finally, the method of integration with Visual C++/C# is used to analyze the application of simulation method. This paper designs an application simulation platform that can interact with STK. The technical details of simulation system frame design, simulation process control and data communication are explored, and the design task instance is verified. The main purpose of the article is to deepen the research of simulation technology in the field of space from the angle of engineering, and to provide reference and example for the realization of simulation technology in engineering. In addition, on the basis of this article, the user can achieve a variety of flexible simulation by using STK technology according to their actual engineering requirements.

References

1. Ni, S., Li, G.: Study on the space vitality based on the numerical simulation technology. In: Proceedings - 2014 6th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA, pp. 607–609 (2014)
2. Xu, Z., Shi, P., Zhao, Y.: Design of electromagnetic spacecraft formation simulation system based on STK. *Comput. Simul.* **32**(5), 95–99 (2015)
3. Du, Y.: Design of STK-based real time visualization simulation system for satellites. *Aerosp. Control Appl.* **35**(2), 60–64 (2009)
4. Ding, Z., Zhang, C.: Spatial simulation model based on STKX component. *J. Sichuan Ordnance* **30**(10), 141–143 (2009)
5. Chen, B., Zhang, G., Wang, N.: Key visualization simulation technology based on STKX component. *Comput. Eng.* **37**(19), 261–263 (2011)
6. Dang, C.: Based on STK and matlab satellite formation's configuration design and simulation. *Adv. Mater. Res.* **823**, 411–416 (2013). Computing, Control and Industrial Engineering IV
7. Bhargava, T.: *Spacecraft Autonomy*, Cranfield University, 7–10 (2003)
8. Hao, L., Feng, Y.: The research of visual simulation technology for orbit maneuver based on STK. *Trans. Shenyang Li Gong Univ.* **28**(1), 57–61 (2009)
9. Hoermann, H., Johannes, B., Petrovich Salnitski, V.: The “Digital Friend”: a knowledge-based decision support system for space crews. *Acta Astronaut.* **63**, 848–854 (2008)
10. Simonet, A., Simonet, M.L.: Conceptual and physical structuring of the object space in a knowledge and data management system. In: Proceedings of the International 2nd Conference on Computer Science and its Applications, CIAA 2009 (2009)

Investigation on the Electromagnetic Field Leakage Effect Dominated by the Cable Penetration for a Spacecraft

Ran Li¹(✉), XiaoYong Yang¹, Xiang Li¹, and XuYang Du²

¹ Beijing Institute of Spacecraft System Engineering, Beijing, China
lir501@126.com

² Institute of Telecommunication Satellite, Beijing, China

Abstract. The electromagnetic shielding effectiveness (SE) is essential to a spacecraft, since the electronic equipment of it may be severely interfered by the electromagnetic interference. SE is determined by both the aperture and the cable penetration except the structure material. Former research is mostly on the impact of the aperture. In this paper, both simulations and lab tests are done to investigate the impact of the cable penetration. The results show that the impact of the cable penetration is obvious and non-neglectable, and it is up to 20 dB for the frequency band lower than 1 GHz. The shielding operation can be deployed to reduce the electromagnetic leakage and then SE can be improved by 10–15 dB for the frequency band of 200 MHz–3 GHz. The results also show that the impact is not related with whether a signal is transferring or not in the cable. If there is an original cable perforation and then an additional cable perforation near it does not impact SE obviously. The impact is dependent on the locations of the built-in radiated sources.

Keywords: Shielding effectiveness · Cable penetration
Electromagnetic compatibility

1 Introduction

Nowadays, the sensitive electronic products of a spacecraft are often interfered in the self-compatibility test, because of the unintentional radiated emission of the modern electronics products which have high speed digital process [1]. Usually, a spacecraft deck is made of aluminum honeycomb panel with honeycomb core and aluminum skins or mixed honeycomb panel with honeycomb core and carbon skins. This kind of panel can provide superior shielding effectiveness.

But the shield shell of a spacecraft cannot be designed as a hermetically sealed one like a perfect Faraday cage, since there are apertures, slots and cable penetration which are used to fulfill the requirements of communication [2] and other operations. These openings create electromagnetic energy coupling paths and thus allow the electromagnetic waves to travel through them [3]. Due to this reason, the shielding effectiveness is seriously degraded. As a result, the apertures, slots and cable penetration in the shield shell can influence the effectiveness of the shielding enclosures significantly [4].

In the past, the research mostly focuses on the apertures and slots, and several analytical methods and numerical models have been purposed to analyze the influence of them [5–8]. Less research work is done on the cable penetration, though the impact of the cable penetration on shielding effectiveness (SE) is more serious than apertures and slots in theory. That's because these cable penetrations, unfortunately, create electromagnetic energy coupling paths that allow inside electromagnetic waves to get out of the enclosure, thus it degrades the shielding effectiveness. Therefore, simulations and experimental researches are needed to investigate the leakage effect dominated by the cable penetration.

This paper presents that the cable penetration is the key element which impairs the SE of spacecraft greatly. The electromagnetic leakage from the cable penetration of a typical spacecraft and the effect of shielding operation were studied experimentally. The investigation can be used to give a guide to the perforation design and improve SE of the enclosure during the initial phase of the spacecraft development. Then the Electromagnetic Interference (EMI) aroused by the cable penetration is reduced.

2 The Calculation and Simulation of the Electromagnetic Field Leakage Dominated by the Cable Penetration

The EMC problems caused by the electromagnetic wave penetration into a cavity with apertures and cable penetrations are very important [9, 10]. The over-sized cavity theory is used to assess the electromagnetic fields inside an enclosure [3].

According to this, the ambient fields in the cavity are roughly homogeneous and isotropic. The distance between the units is not a relevant parameter, and the orientation and polarization of a given leaking point do not matter.

A given amount of RF power injected into the cavity is uniformly attenuated; the power picked up by a receive probe is independent of the probe location and orientation. This fixed attenuation is called the cavity insertion loss X_c . The cavity insertion loss X_c is a combination of a radiated component X_p and a conductive component X_σ with (1), (2) and (3). Where δ is the skin depth and S_0 is the total surface of cable penetrations in (4).

$$\frac{1}{X_c} = \frac{1}{X_p} + \frac{1}{X_\sigma} \quad (1)$$

$$\frac{1}{X_p} = 8 \times \pi \times \frac{S_0}{\lambda^2} \quad (2)$$

$$\frac{1}{X_\sigma} = \frac{8 \times \pi^2}{\lambda^3} \times (S - S_0) \times \delta \quad (3)$$

$$\delta = \frac{1}{\sqrt{\pi \times f \times \mu \times \sigma}} \quad (4)$$

Then the simulation of the electromagnetic field leakage dominated by the cable penetration is performed. A simple metallic square enclosure is assumed to be the shielding structure. The dimension is 1 m. There is a circle perforation on the top panel, its diameter of the circle is 30 mm, and the circle is perforated by a metallic cylinder. The diameter of the cylinder is 20 mm. All sides of the enclosure are assumed perfectly conducting with an infinitesimal thickness. The receiving probe outside the enclosure is independent of the orientation and polarization. The simulation result is described in Fig. 1.

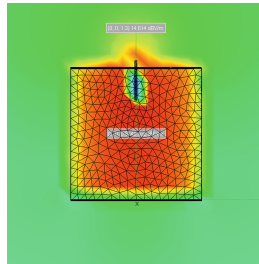


Fig. 1. The simulation result of cable penetration dominations

It's observed obviously that the electromagnetic field outside the enclosure varied severely with the excitation of the inner radiated source. SE declined at least 20 dB within the frequency band of 100 MHz–1 GHz.

The simulation result illuminates that the cable penetration is the key element which affects SE.

3 Experimental Configuration According to the Structure Design of a Spacecraft

The above simulation is based on a simplified assumption. It cannot provide sufficient accuracy to meet the design specifications, since the shielding structure of the spacecraft is geometrically more complex than a cube.

Then a typical rectangular enclosure designed according to the structure of spacecraft is established. The geometry of this enclosure is shown in Fig. 2. The enclosure was composed of five pieces of 2.5-cm-thick aluminum honeycomb panel with honeycomb core and aluminum skins and one piece of 0.1-cm-thick aluminum board for the face installing the radiated source. Six cable penetrations (CP) are distributed on the four panels, and they are marked with CP1 to CP6 in Fig. 2. The inside dimension of the enclosure was 110 cm × 41 cm × 118 cm. For every measurement, the enclosure was fed with 3 antennas through 3 subminiature version A (type-SMA) flanges, which were peripherally bonded to the aluminum board.

The positions of the radiated sources are shown in Fig. 3, A-A, B-B and C-C are the identifiers of the three cross sections. P is the abbreviation of position, and P1–P9 are the positions of the nine radiated sources. The unit of the distances in the figure is millimeter (mm).

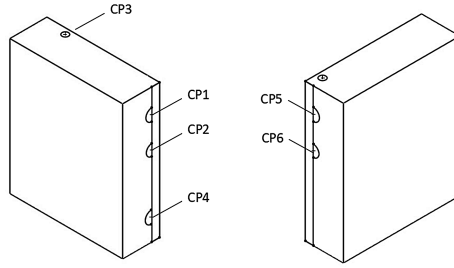


Fig. 2. A typical rectangular enclosure with aperture and cable apertures

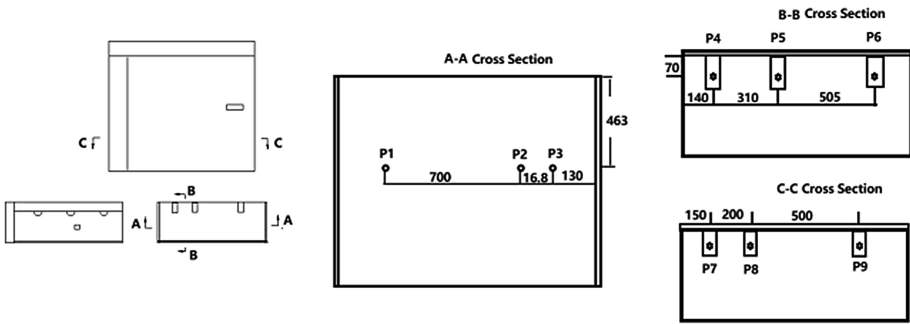


Fig. 3. The position of radiated sources

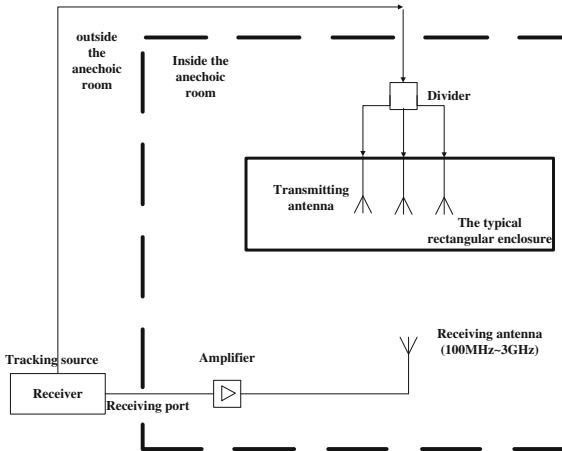


Fig. 4. Experimental setup

The radiated emission measurements were performed in a 10-m anechoic chamber, as is shown in Fig. 4. Two-port parameters were employed with an R&S ESU40 receiver. Port 1 was connected to the radiated source, and Port 2 was connected to a receiving antenna 1 m away from the enclosure. For all measurements multiple antenna was used in the receiving system. The receiver was placed outside the anechoic room to avoid influencing the measurement result.

4 The Impact of the Cable Penetration

To investigate the impact of the cable penetration, 6 test cases have been designed in Table 1.

Table 1. Illuminations of 7 test cases

Test case	Illuminations
1-1	There was no cable penetration in the structure
1-2	A cable without signal was set to pass through CP1
2-1	Apply the signal generator to generate the signal of 600 MHz and 0 dBm to the cable
3-1	Set 5 cables pass through CP1
3-2	On the basis of case 3-1, set another 5 cables pass through CP2
4-1	Set two different locations of the radiated sources
5-1	The cable and CP1 were shielded by using aluminized polyimide film

4.1 The Cable of Perforation Without Signal

It's shown in Fig. 5 that the cable penetration causes huge electromagnetic leakage especially at low frequency band. The SE is worsened by 10–20 dB within 100 MHz–1 GHz frequency band. Because the cable penetration acts as a functional antenna, the electric surface current of the cable is increased and the electromagnetic field leakage from the perforation is also increased.

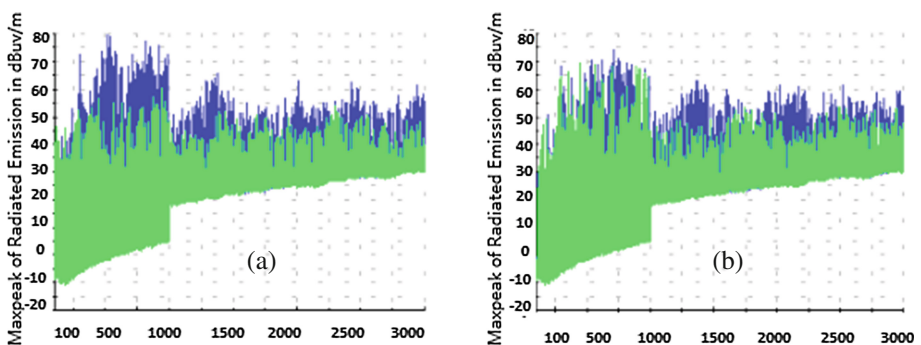


Fig. 5. Vertical (a) and horizontal (b) polarization for case 1-2 (blue) and case 1-1 (green)

4.2 The Cable of Perforation with Signal

Figure 6 shows no difference on SE whether there's signal in the cable or not. The illumination is that it's the electric surface current of the cable from perforation which resulted in decrease of SE, not the signal in the cable.

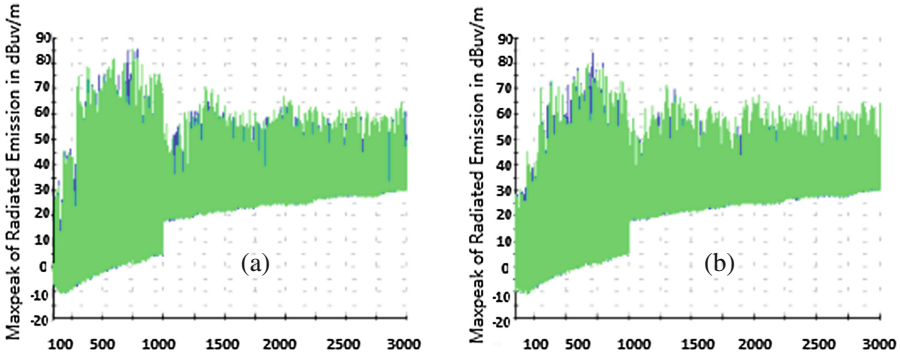


Fig. 6. Vertical (a) and horizontal (b) polarization for case 1–2 (blue) and case 2–1 (green)

4.3 Two Cable Penetrations

It's shown in Fig. 7 that if additional cable penetration is applied to the existing cable penetration, SE would not be influenced much. When SE of a shielding enclosure is influenced by more than one element, such as slot, aperture, cable penetration and etc., the one which make the SE worsen most severely is conclusive.

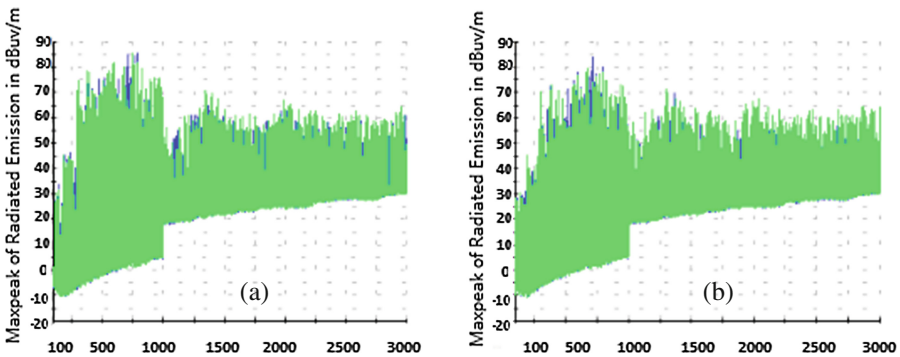


Fig. 7. Vertical (a) and horizontal (b) polarization for case 3–1 (blue) and case 3–2 (green)

4.4 Multiple Cable Penetrations

Generally, multiple cable penetrations exist in the spacecraft structure. In this experiment, it is assumed that there are 6 cables passing through CP1–CP6 separately.

The signal transmitting system is calibrated to generate the signals of 43 dBuV/m @600 MHz and 36 dBuV/m@1.3 GHz from the 3 radiated sources simultaneously. Then the locations of the 3 radiated sources are changed to observe the variation of SE. There are two test cases, for test case 2, the locations of the radiated sources (P2&P4&P7) are closer to the cable penetrations than the test case 1 (P3&P5&P8).

Table 2. The variation of SE with changing the locations of radiated sources

Test Case	The locations of the radiated sources	SE at 600 MHz (dB)	SE at 1.3 GHz (dB)
1	P3&P5&P8	21.5	11
2	P2&P4&P7	17.3	9

It's shown in Table 2 that the electromagnetic leakage of the enclosure aroused by the 6 cable penetrations varies with the locations of the radiated sources. Compared with the two test cases, the electromagnetic leakage is more serious for test 2 and SE is thus declined as well. This hints that the electromagnetic leakage from the nearby radiated resources is greater.

5 The Verification of the Shielding Operation

In engineering, some cables have to pass through the panels of metallic enclosure without filters or flanges. Thus some necessary work should be performed to reduce the electromagnetic leakage and improve the SE of the structure. Aluminized polyimide film is a kind of recommended shielding material.

It's shown in Fig. 8 that the SE could be improved 10–15 dB with the shielding operation. The shielding material and the shielding method can be applied to solve the EMI problem caused by cable penetration.

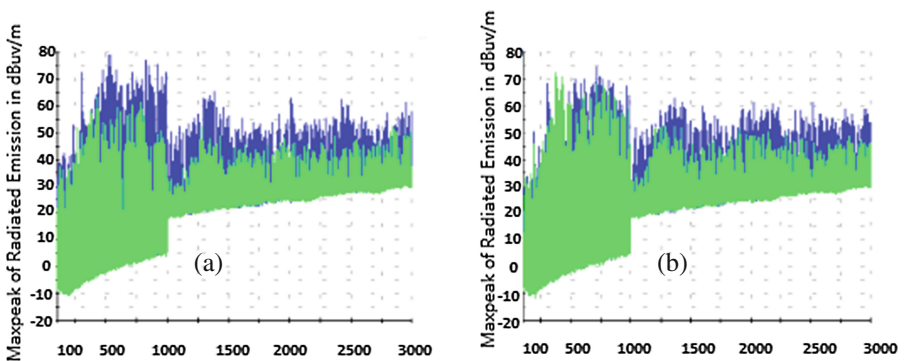


Fig. 8. Vertical (a) and horizontal (b) polarization for case 1–2 (blue) and case 4–1 (green)

6 Conclusion

The question of the electromagnetic field leakage from the cable penetrations for a shielding enclosure was investigated with experiments and simulation. A typical rectangular metallic enclosure with built-in radiated sources was used to study the electromagnetic field leakage effect due to the cable penetrations. The results of the experiments and simulation match each other.

In summary, the cable penetration is the key elements which dominate the SE of spacecraft. The max deterioration reached 20 dB within the low frequency band. The radiated emission of the enclosure varied with the location of the built-in radiated source.

The aluminized polyimide film is suggested to be used for shielding the cable penetration. The shielding operation can improve SE 10–15 dB within 200 MHz–3 GHz. What followed is that the EMI problem is solved.

References

1. Chen, J., Liu, G.: Analysis of electromagnetic shielding with a incompletely sealed metal plane. *Electron. Eng.* **34**(7) (2008)
2. Ghandehari, M.B., Moghadasi, M.N., Attari, A.R.: Improving the shielding effectiveness of a rectangular metallic enclosure with aperture by using extra shielding wall. *Prog. Electromagnet. Res. Lett.* **1**, 45–50 (2008)
3. Li, R., Han, Y.Z., Jiang, H., Wang, X.C.: Analysis on e-field environment inside the communication satellite. In: *The 5th International Symposium on Electromagnetic Compatibility (EMC 2017)*, Beijing (2017)
4. He, M., Liu, G.B., Hu, Y.A., Jiang, L.Q.: Influence of apertures in the electronic equipment of missile on electromagnetic shielding effectiveness. *J. Astronaut.* **27**(2), 262–267 (2006)
5. Xue, M.F., Yin, W.Y., Mao, JF.: Shielding effectiveness characterization of metallic rectangular cascaded enclosures. In: *Asia-pacific Microwave Conference*, pp. 1–4 (2007)
6. Scogna, A.C., Antonini, G., Orlandi, A.: Analysis of radiated emissions and shielding effectiveness for a metallic enclosure with shielding springs. In: *IEEE International Symposium on Electromagnetic Compatibility*, pp. 1–6 (2007)
7. Zhan, Q., Yang, R.Q., Gao, X.: Research the shielding effectiveness of aperture array on the cavity based on FDTD method. *J. Guizhou Univ. (Nat. Sci.)*, **29**(4) (2012)
8. He, H., Fu, J.M.: Numerical investigation on the radiation leakage from the built-in pulse in the target cavity through the aperture. *Chin. J. Radio Sci.* **19**(z1), (2004)
9. Park, H.H., Eom, H.J.: Electromagnetic penetration into a rectangular cavity with multiple rectangular apertures in a Conducting Plane. *IEEE Trans. Electromagnet. Compat.* **42**(3), 303–307 (2000)
10. Kaiser, K.L.: *Electromagnetic Compatibility Handbook*. CRC Press, Boca Raton (2012). ECSS-E-HB-20-07A

Big Data Workshop

Research on Impact of LTE RSSI Based on Network Data Correlation Analysis and Optimization Practice

Mingxin Li^(✉), Tianbiao Tang, Juanjuan Tan, Hao Guo,
and Hongxi Liao

No. 192 Yuzhou Road, Yuzhong District, Chongqing, China
limx87@chinaunicom.cn

Abstract. Regarding the topic of LTE uplink interference, this article mainly discussed the impact of LTE network on RSSI, and the impact of RSSI on network quality and user perception. Furthermore, this article classified the suggested RSSI thresholds based on the extent of impact, and utilized these thresholds, in order to guide optimization practice and improve user perception. Finally, after taking a cell of an indoor distribution system on campus with high RSSI level as an example, user perception rate has been found significantly increased after the cell splitting.

Keywords: LTE · Network load · RSSI · KPI · Optimization practice

1 Preface

Along with the continuous evolvement of Mobile Communication Network, the co-existence of multiple networks built by different Telecommunications Operators is inevitable, and the reuse rate of frequency has been continuously going up [1]. With promoting reduction of excessive capacity, and co-constructing and -sharing, co-site of outdoor macro stations and sharing antenna transmission system of indoor distribution systems becomes more and more common [2], and network interference has also grown considerably.

In LTE networks, RSSI (Received Signal Strength Indicator) is a measurement of power present in a signal received by the base station. It is often used as a key indicator for analyzing uplink interference [3]. Insistent high RSSI level indicates that uplink signal is too strong and uplink interference is overwhelming. High RSSI level might cause the decrease of access success ratio and handover success ratio of a LTE cell. It may also cause low network speed [4], and thereby creates impact on user perception.

This article, based on network-oriented big data analysis [5], mainly investigated the impact of LTE network load on RSSI, and the impact of high RSSI on network quality. According to the degree of impact of network quality, the RSSI level was classified into different intervals, which can be utilized to guide RSSI optimization process.

2 Analysis of the Impact of RSSI

Basically, there are approximately three main categories of reasons that lead to high RSSI level: intra-system interference, inter-system interference [6] and network overload. Common reasons for intra-system interference include: unqualified construction quality, antenna transmission system malfunction, and parameter configuration error. Based upon formation mechanism, frequency band and source, inter-system interference could be categorized as barrage jamming, spurious emission interference, and intermodulation interference [7]. Since previous literatures have considerably discussed the first two reasons, this article only focused on the impact of network load on RSSI.

In this study, the hourly KPIs (Key Performance Indicator) and KQIs (Key Quality Indicators) was generated from China, Chongqing's LTE cells. The KPIs included: cell's RSSI level (the average RSSI of all RRUs of a cell), uplink and downlink PRB utilization ratio, uplink and downlink bearing perception rate per user, uplink and downlink data usage, access success ratio, handover success ratio, and LTE drop ratio. KQIs mainly chose relevant indicators of video and web page. In order to ensure the data was valid, the original extracted data was furtherly cleaned up, e.g., deleting anomalies, blank records, records with low average daily data-usage cells (might still be in construction period, high data volatility) and so on so forth, then statistical analysis method was introduced for analytic purpose.

2.1 The Impact of Network Load on RSSI

PRB (Physical Resource Block) utilization ratio is often employed to evaluate network load. The higher the number of access users and throughput is, the higher the PRB utilization ratio would be. PRB utilization ratio can be assorted as uplink PRB utilization ratio and downlink PRB utilization ratio.

Based on the statistical analysis on the co-relationship between uplink PRB utilization ratio and downlink PRB utilization ratio, as shown in Fig. 1, the uplink PRB utilization ratio is positively correlated with the downlink PRB utilization ratio; hence study on the co-relationship between RSSI level and downlink PRB utilization ratio could be used as the indicator of network load on RSSI level.

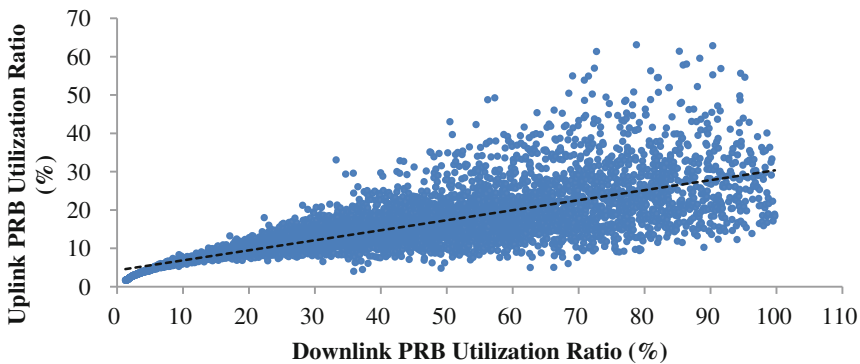


Fig. 1. The relation between uplink and downlink PRB utilization ratio

To avoid the impact of interference on RSSI, this study, from the purified data then, eliminated cells in which interference exist when idle. Preliminarily, a cell is considered interfered when RSSI level is greater than -95 dBm. According to the statistical data of Chongqing’s LTE network, the network idle time is mainly at around 5 a.m.; therefore, records of cells with RSSI level higher than -95 dBm at around 5 a.m. were deleted. As shown in Fig. 2, RSSI level rises as the PRB utilization ratio increases, with or without the interfered cells. Furthermore, natural logarithm relation is found between RSSI level and downlink PRB utilization ratio. When downlink PRB utilization ratio reaches 50%, RSSI approximately increases by 6 dB; When downlink PRB utilization ratio reaches 100%, RSSI increases by about 8 dB.

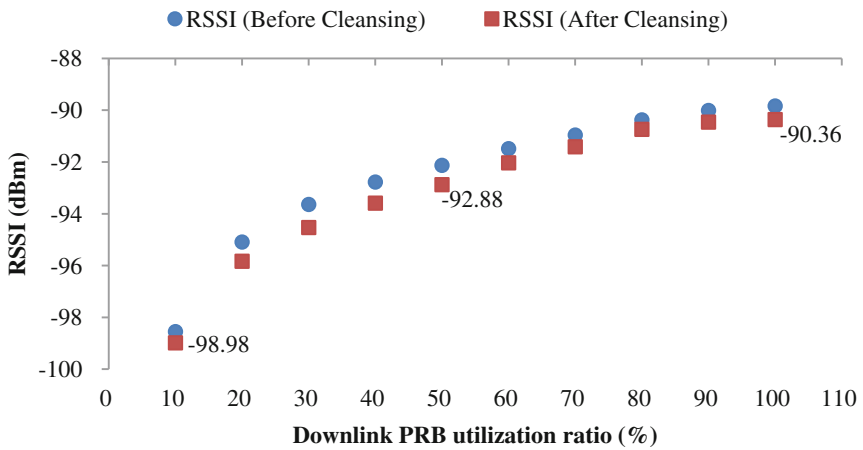


Fig. 2. RSSI level rises as the PRB utilization ratio increases

Based on the figure above, this study identified a regression relationship between RSSI level and downlink PRB utilization ratio. Assuming this relationship can be reflected with function (1), a non-linear regression analysis [8] was adopted (based on logarithm model).

$$RSSI = A \ln(PRB) + K \tag{1}$$

Once the summarized model, and analysis result of variance and coefficients [9] were obtained as shown in Tables 1 and 2, and the coefficients got applied into above formula, the regression function was identified, as function (2):

Table 1. Model summary

R	R-square	Adjusted R-square	SEE
0.998	0.996	0.995	0.195

Table 2. Analysis of variance and coefficients

	Sum of squares	df	Mean square	F	Sig.	Unstandardized coefficients		t	Sig.	
SSR	68.101	1	68.101	1794.5	0	B	SE			
SSE	0.304	8	0.038			A	3.753	0.089	42.362	0
SST	68.405	9				K	-90.108	0.093	-964.995	0

$$RSSI = 3.753 \ln(\text{PRB}) - 90.108 \tag{2}$$

In addition, to further investigate the co-relationship between the RSSI level of high RSSI cells and downlink PRB utilization rate, all cells with RSSI level higher than -95 dBm around 5 a.m. were analyzed. If there is interference, the upward tendency of RSSI level gradually levels off while the PRB utilization rate is rising. When the downlink PRB utilization ratio reaches 50%, RSSI level rises about 3 dB; When down-link PRB utilization ratio furtherly increases, RSSI level gradually becomes stable.

2.2 The Impact of RSSI on Network Quality

Based on the RSSI intervals, this study conducted statistical analysis on cleaned KPI records in order to measure user perception rate, and thereafter, identify the relationship between the user perception rate and RSSI. As illustrated in Fig. 3, user perception rate declines while RSSI is going up, thus it can be concluded that high RSSI level had consequences on cells' throughput rate.

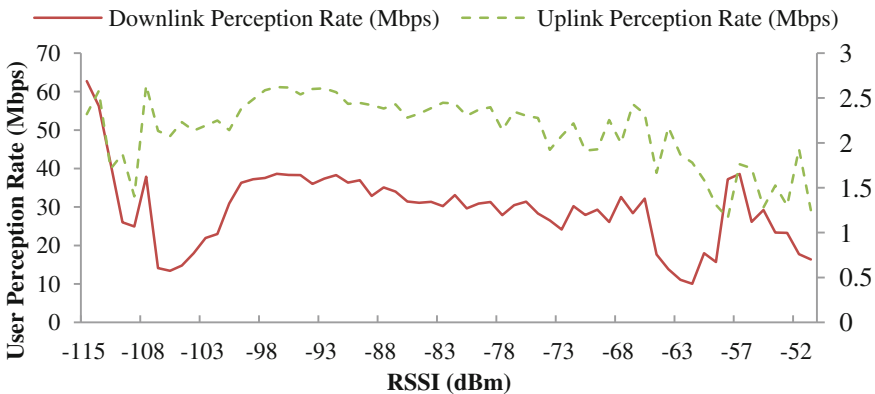


Fig. 3. User perception rate declines with RSSI increases

In addition, it could be found in Fig. 4 that cells' access success ratio and handover success ratio falls as the RSSI level increases. When RSSI is greater than -95 dBm, access success ratio and handover success ratio drops to 99.7%; When RSSI is greater than -85 dBm, access success ratio and handover success ratio decreases distinctively. Similarly, LTE dropping ratio rises as RSSI level inclines. When RSSI is greater than -85 dBm, LTE dropping ratio is over 0.1%. Therefore, effective maintenance of RSSI level is a crucial key for better network quality and service perception.

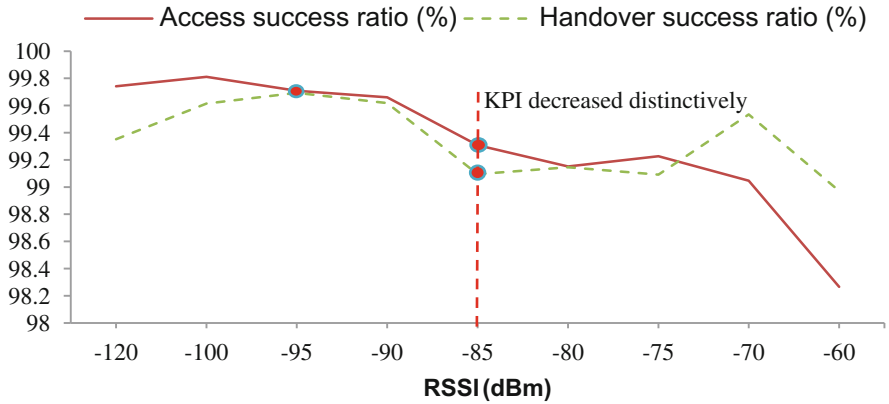


Fig. 4. Access success ratio and handover success ratio falls as the RSSI level increased

According to the measurement report (MR is the short form), this study extracted all cells' coverage ratio (the ratio of which RSRP was greater than -110 dBm), RSSI, and uplink user perception rate in this LTE network. The median of coverage ratio was used as the standard for partitioning cells into two groups: cells with good coverage and cells with bad coverage. This article also compared the impact of RSSI on user perception rate between user at far point and user at near point in the graphs below (see Fig. 5). It could be found that the elastic coefficient between RSSI and uplink user

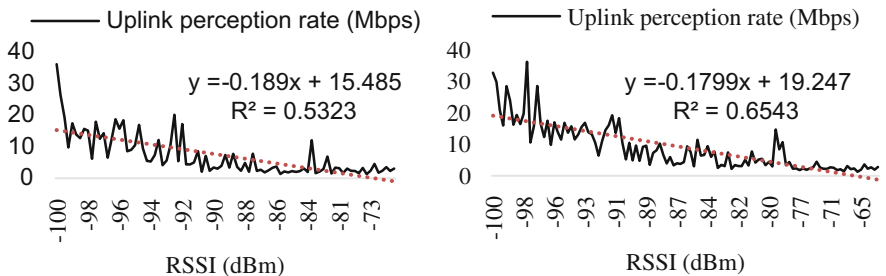


Fig. 5. The relation between uplink perception rate for far/near users (left/right) and RSSI

perception rate is 0.1799 for users nearby, and 0.189 for users far away. It shows that user perception rate at far point is more sensitive to the fluctuation of RSSI level.

In conclusion, when downlink PRB utilization ratio was greater than 50%, average RSSI of the network would rise to about -95 dBm; when RSSI was greater than -95 dBm, LTE uplink/downlink perception rate and access/handover success ratio began to decrease; when RSSI was greater than -85 dBm, KPIs such as access/handover success ratio and LTE dropping ratio deteriorated demonstrably. This proves that effective control of uplink interference of LTE cells is especially important. It is recommended to keep RSSI below -95 dBm for all RRUs in LTE network.

3 RSSI Optimization Practice

Along with the booming of 4G users and widespread of 4G terminals, network capacity problem appears to be more critical. Since March 2017, China Unicom has launched a series of large data plan such as “Tencent King Card” and “Ice Cream Plan”. In the meanwhile, 4G network throughput has risen sharply, and the quantity of high PRB utilization ratio cells has increased significantly. As showed above, high PRB resource utilization would lead to the uplift of RSSI, which would lead to undesirable network quality and user perception. Moreover, user complaints would in-crease accordingly. For cells with high RSSI due to network overload, it is common to apply carrier expansion or cell splitting to expand network capacity, thereby enhancing the user network perception. In the following paragraphs, this article will take a real-life optimization project as an example to analyze the changes in RSSI before and after expansion, as well as the improvement of network KPIs.

Cell 1 was a cell of an indoor distribution system in some university, and the cell had a 3-RRU cascade connection which distributed in building No. 12, building No. 14 and building No. 16 in the dormitory area. The graph above shows how Cell 1’s RSSI level changed in a day (see Fig. 6). It could be found that during idle time (around 6 a.m.), the cell’s RSSI was around -98 dBm. With the rise of downlink PRB

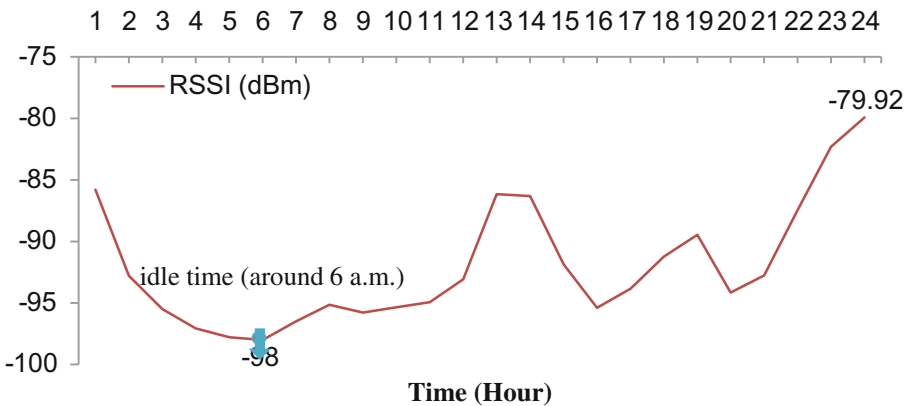


Fig. 6. Cell 1’s RSSI level changes in a day

utilization ratio, RSSI sharply drops to -80 dBm, and the service perception of users has gradually been limited, as shown in the following graph (see Fig. 7). Subsequently, the 3 RRUs of Cell 1 were split, and the original Cell 1 was divided into 3 separate cells (Cell 1, Cell 2, and Cell 3). After the split, the original cell's downlink PRB utilization ratio decreases by 60%, RSSI decreases by 5–8 dB, downlink perception rate increases significantly, from the original 5 Mbps to 42 Mbps (see Fig. 8).

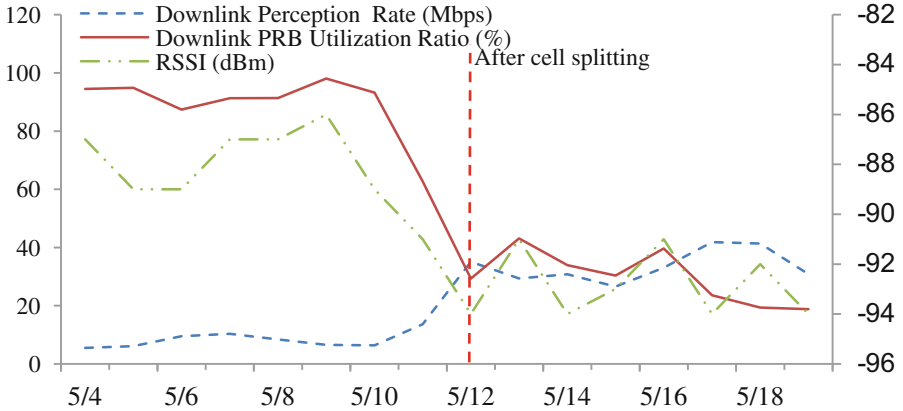


Fig. 7. Downlink perception rate, downlink PRB utilization ratio and RSSI

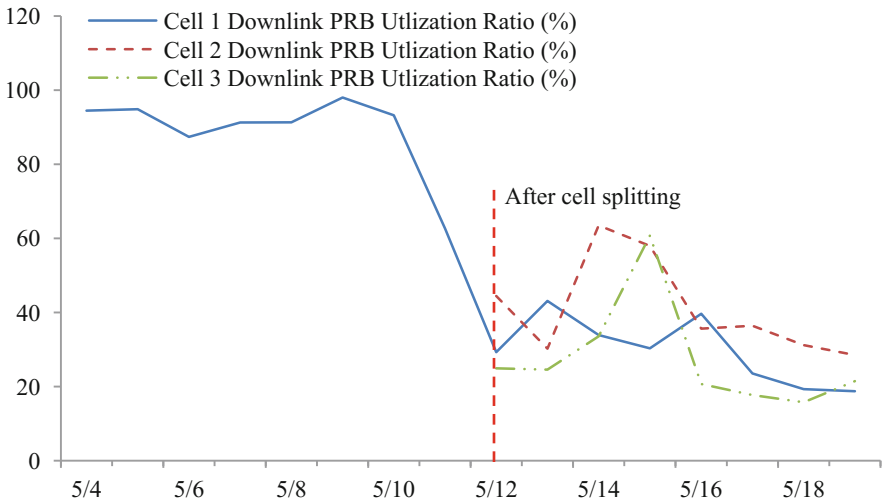


Fig. 8. Downlink PRB utilization ratio before/after spitting of cell 1

4 Conclusion

The uplink interference of LTE network will be more and more prominent. LTE network overload lifts RSSI besides the influences of inter-system interference and intra-system interference. This article analyzed and discussed the impact of LTE network load on RSSI through big data of LTE network. As the LTE network load increased, the RSSI generally rises 3–8 dB. When RSSI is greater than -95 dBm, LTE network quality and user perception begin to worsen, and network quality and perception were obviously decreases when RSSI is greater than -85 dBm. With the future LTE network load increases gradually, the phenomenon of high RSSI will become more and more common because of network overload. In order to ensure good service perception, it is recommended that every RRUs' RSSI level of LTE cells should be controlled below -95 dBm. In view of the cells with high RSSI due to overload, it is necessary to conduct cell splitting, carrier expansion or load balancing [10] in time so as to enhance the user network perception.

References

1. Li, Z.-N., Wei, G.X.: Analysis on uplink interference of LTE system. *Commun. Technol.* **12** (43), 47–70 (2010)
2. Xu, L., Luan, Y., Cheng, X., Cao, X., Chao, K., Gao, J., Jia, Y., Wang, S.: WCDMA data based LTE site selection scheme in LTE deployment. In: 1st International Conference on Signal and Information Processing, Networking and Computers, pp. 249–260. CRC Press, Taylor & Francis Group, Beijing (2015)
3. Deb, S., Monogioudis, P.: Learning-based uplink interference management in 4G LTE cellular systems. *IEEE/ACM Trans. Netw.* **23**(2), 398–411 (2015)
4. Lin, Q.: Impact of uplink interference on FDD-LTE upload rate. *Chin. New Telecommun.* **18** (15), 68 (2016)
5. Xu, L., Luan, Y., Cheng, X., Fan, Y., Zhang, H., Wang, W., He, A.: Telecom big data based user offloading self-optimization in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)
6. Guo, X.-R., Zhang, T.: Research on intermodulation interference between 2.3 G TD-LTE and China unicom WCDMA systems. *Mob. Commun.* **12**, 15–20 (2015)
7. Tang, Q.J.: LTE interference identification method based on PRB noise floor statistics. *Mob. Commun.* **40**(20), 1006–1010 (2016)
8. Liangping, H., Gao, H.: *Nonlinear Regression Analysis and Intelligent Realization of SAS*. Publishing House of Electronics Industry, Beijing (2013)
9. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Publishing House of Electronics Industry, Beijing (2004)
10. Xu, L., Cheng, X., Liu, Y., Chen, W., Luan, Y., Chao, K., Yuan, M., Xu, B.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: 16th IEEE International Conference on Communication Technology, pp. 806–812. IEEE press, Hangzhou (2015)

Customer Churn Analysis for Telecom Operators Based on SVM

Runsha Dong^(✉), Fei Su, Shan Yang, Xinzhou Cheng,
and Weiwei Chen

China Unicom Network Technology Research Institute, Beijing, China
dongrs@dimpt.com

Abstract. Customer churn prediction is important for telecom operators to retain valuable users. Accurate features that can characterize customer behaviors, as well as efficient extraction method are key factors in constructing the customer churn analysis model. In literature, Support Vector Machine (SVM) has shown its applicability to the problem of customer churn analysis. This paper identifies the main features that influence the customer churn model from telecom experts' viewpoints, and proposes a suitable one based on Support Vector Machine (SVM). An experimental results also is illustrated to verify reasonableness of the proposed models.

Keywords: Customer churn · Telecom operators · SVM

1 Introduction

With the arrival of the mobile internet era, competition among telecom operators has been fierce. The market is almost saturated, and the pool of available customers is very limit. Telecom operators also have to face the external impact of Internet companies. For stable profit, telecom operators pay more attention to customer relationship management, especially to retain valuable customers. A customer churn prediction model with reasonable accuracy is important for decision of customer retention. Many data mining methods have been applied to formulate the problem, such as decision tree, logistic regression [1], naive Bayesian classifiers and neural networks. Some literatures [2–5] even combine several different techniques for churn prediction. It is noted that, the constructed model based on these methods above cannot guarantee prediction precision and generalization ability [6]. This paper constructs a framework of customer churn prediction from a professional viewpoint and also considers the support vector machine (SVM) method for better precision.

The paper is organized as follows. In Sect. 2, the process of customer churn prediction is proposed. In Sect. 3, feature extraction method SVM is introduced. In Sect. 4, experiment and results are exhibited. The paper is concluded in the final section.

2 Problem Formulation

In the process of analysis for customer churn prediction, there are several key steps: problem formulation, algorithm selection, data preparation, result assessment and model output. The flow chart is shown in Fig. 1.

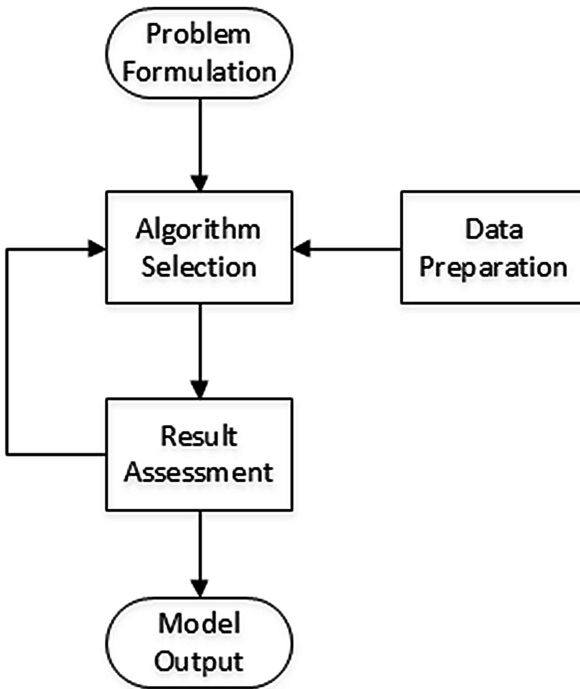


Fig. 1. The flow chart of customer churn prediction.

Problem Formulation: From a business perspective, the understanding of experts from telecom companies is instructive to decide what factors are crucial in customer churn prediction model. The related understandings include: the definition of churn; which kind of customer churn: active churn or passive churn. If mobile telecom customers are not satisfied with the current service (or competitors provide a better service), they may stop the current service actively. Some other customers may passively churn since they go to work located in a different place due to graduation or promotion. Customer churn analysis is usually more concerned about the active churn of valuable users who are profitable to telecom operators.

Algorithm Selection: When basic concepts are identified, the model of customer churn prediction is generally constructed based on selected data mining method. Original data sets should be prepared. And what customers' features will be used for constructing models should be determined. At the same time, the original data set is usually split into two parts. One part is the training set which is used to construct the model, the other part is test set to verify the efficiency of the proposed model.

Data Preparation: The status in business of telecom operators show that, there is around 2.5% churned customers per month. Therefore, customer churn prediction also face the problem of imbalance data set [7]. In addition, user data in telecom operators even for one city has reach to a mount of hundreds of thousands or even millions. Training time is difficult to meet the requirements if all data is trained to construct customer churn prediction model. Moreover, since the original data sets come from different fields and departments of telecom operators, many important features are missing or even wrong, which influence the model accuracy. Based on the above, data should be well-prepared.

Result Assessment: After a model for customer churn prediction is constructed, result will be assessed according to the performance of selected algorithm. The test data set will quantify the performance of models generally by verify the correct rate. There are also other criterions to assess the constructed model such as cover rate [8]. The effect of actual implication is also a key standard for assessment.

3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised linear classification algorithm in machine learning, which was first published in 1995 by Cortes and Vapnik. SVM has been extensively researched and applied in text classification, especially for binary classification tasks. In the later research papers, SVM was generalized to solve multi-classification problem. The method supposes that ordinary customer data and potential churned customer data can accurately be separated by a hyperplane when the appropriated parameters are selected. SVM solves the classification problem by transforming samples in low dimensional space which is linearly inseparable into ones in a higher dimensional space. Take an example, there is a linearly inseparable problem in two dimensional space (See Fig. 2, where linear hyperplane is a dot).

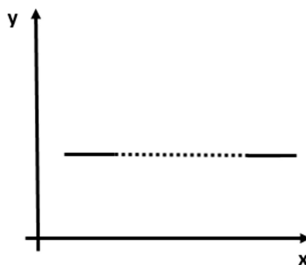


Fig. 2. Straight line and dotted line are linearly inseparable in two-dimensional space.

However, there exists a nonlinear function in two dimensional space described by

$$g(x) = c_0 + c_1x + c_2x^2. \quad (1)$$

It can divide straight line and dotted line into two categories by judging that the line is above or below the non-linear classification function (See Fig. 3).

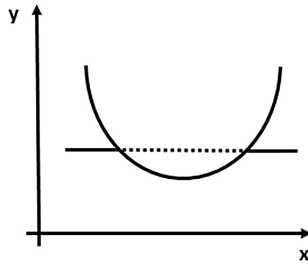


Fig. 3. Straight line and dotted line are linearly separable in four-dimensional space.

If it supposes that,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}, \tag{2}$$

we have

$$g(x) = f(y) = ay. \tag{3}$$

Then, a linearly inseparable problem in two dimensional space is solved in four dimensional space. In a classification problem for SVM, the training set can be described by

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, x_i \in \mathbb{R}^n, y_i \in \{-1, +1\}, \tag{4}$$

with each sample

$$D_i = (x_i, y_i). \tag{5}$$

It is noted that each sample may be in high dimensional space.

4 Model Illustration

In this paper, we consider the original datasets provided by a mobile telecom operator and describe subscribers of one city. User data include bill data, call detail records, data detail records and user terminal information. The subscribers are activated for at least 3 month. Too much features of customers usually bring some troubles, such as over-fitting, or larger memory requirements. The most important and informative factors which influence the potential churned customer are chosen by experts in telecom operators and given as follows (See Table 1).

Table 1. Key factors which influence potential churned customer.

Item	Data	Purpose
Bill	<i>city_id</i>	Subscribers' basic and informative features
	<i>user_id</i>	
	<i>vip_level</i>	
	<i>in_network_time</i>	
	<i>charge_already</i>	
	<i>pay_month</i>	
	<i>age</i>	
	<i>sex</i>	
Detail records	<i>call_duration</i>	Linkage is established between terminal information with bill by TAC and IMEI
	<i>call_times</i>	
	<i>total_traffic</i>	
Terminals	<i>device_value</i>	Handset model and manufacturer information
	<i>equipmenttype</i>	
	<i>device_network</i>	
	<i>manufacturer</i>	
	<i>devicetype_merge</i>	

Where, 'user_id' in bill data is the only identification of a unique subscriber in the telecom operators. 'In_network_time' describes a date that the subscriber activated the corresponding service in telecom operator. 'Charge_already' and 'pay_month' show that whether a subscriber is charged or not, and how much is the fee. Other features are collected by connection of 'user_id'. Handset model and manufacturer information is gathered by a connection between TAC and IMEI, which are stored in telecom operator data warehouse.

Besides the factors mentioned above, there may be also other factors, such as customer complain records, customer perception of network quality, market promotions of telecom operators. Though these factors are useful to construct customer churn prediction model, they are lacking in quantitative accuracy. However, it is still worthy to interpret these factors based on practical applications in the process of customer churn prediction.

It should be noted that data mining method SVM classes the samples described by numerical data, not char data. Original data may be missing or incorrect. Therefore, in the phase of data preparation, some measures are taken in order to enhance data quality and to simplify the model of customer churn prediction. These measures include imputation of the missing values, discretization of continuously variable, transformation from one discrete data set to another and new variable derivation.

In our experiments, basic discretization and transformation are described as follows (See Table 2).

Table 2. Data discretization and transformation

Data	Discretization and transformation
<i>vip_level:A</i>	1
<i>vip_level:B</i>	2
<i>vip_level:C</i>	3
<i>vip_level:D</i>	4
<i>vip_level:other</i>	0
<i>in_network_time</i>	Transformed into a number which describe how much months has been going on since the subscriber's service was activated
<i>charge_already:Yes</i>	1
<i>charge_already:No</i>	2
<i>charge_already:other</i>	0
<i>age: <20 years old</i>	1
<i>20< age <40 years old</i>	2
<i>40< age <60 years old</i>	3
<i>age >60 years old</i>	4
<i>age: other</i>	0
<i>Sex:male</i>	1
<i>Sex:Female</i>	2
<i>Sex:other</i>	0
<i>Equipmenttype:smart phone</i>	1
<i>Equipmenttype: Feature phone</i>	2
<i>Equipmenttype: other</i>	0
<i>device_network: support 2G network</i>	2
<i>device_network: support 3G network</i>	3
<i>device_network: support 4G network</i>	4
<i>device_network: other</i>	0
<i>manufacturer/devicetype_merge: Top 10% users</i>	1
<i>manufacturer/devicetype_merge:: Top 30% users</i>	2
<i>manufacturer/devicetype_merge:: Top 60% users</i>	3
<i>manufacturer/devicetype_merge:: other users</i>	4

Except for features discretized and transformed as mentioned above, other features such as 'pay_month', 'call_duration', 'call_times', 'total_traffic' and 'device_value' are used directly. In our experiments, training data sets are produced by 5 month original data. For data preparation, random sampling from massive original dataset is needed in order to construct a relatively balanced training data set. In the paper, the amount of

original data is 403280. In which, there are 1629 churned customer. The extraction of original data is adopted, with a ratio of churned customer number to ordinary subscriber number 1:3. The most important, support vector machine algorithm applied in the paper was imported from sklearn algorithm library (sklearn.svm) [9] by adjusting suitable kernel function such as RBF. Suppose that n is the number of customers who are predictably churned and m is the number of actually churned customers, then, the prediction accuracy is calculated as n/m . Based on SVM algorithm, the prediction accuracy can reach up to 70.6%.

5 Conclusions

Customer churn analysis is critical in mobile telecom markets due to limit available users. Many data mining techniques such as regression and decision trees even hybrid techniques have been introduced in former literature. In this research, a framework for constructing the model of customer churn prediction is proposed based on Support Vector Machine method. Main factors informative for data preparation are also introduced from a professional viewpoint. However, there is still limitation for application of the model proposed. Many useful factors which may influence the churned customer are hard to quantify [10]. Therefore, there should be more complex investigations on these useful factors in the future.

References

1. Hassouna, M., Tarhini, A., Elyas, T., AbouTrab, M.S.: Customer churn in mobile markets. *Comp. Tech. Int. Bus. Res.* **8**(6), 224–237 (2015)
2. Inderpreet, S., Sukhpal, S.: Framework for targeting high value customers and potential churn customers in telecom using big data analytics. *Int. J. Educ. Manage. Eng. (IJEME)* **7**(1), 36–45 (2017)
3. Tsai, C.-F., Lu, Y.-H.: Customer churn prediction by hybrid neural networks. *Expert Syst. Appl.* **36**, 12547–12553 (2009)
4. Xu, L., Luan, Y., Cheng, X., Fan, Y., Zhang, H., Wang, W., He, A.: Telecom big data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)
5. Xu, L., Luan, Y., Cheng, X., Xing, H., Liu, Y., Jiang, X., Chen, W., Chao, K.: Self-optimised joint traffic offloading in heterogeneous cellular networks. In: *16th IEEE International Symposium on Communications and Information Technologies*, pp. 263–267. IEEE press, Qingdao (2016)
6. Xia, G.-E., Jin, W.-D.: Model of customer churn prediction on support vector machine. *Syst. Eng. Theory Pract.* **28**(1), 71–77 (2008)
7. Gui, C.: Analysis of imbalanced data set problem: The case of churn prediction for telecommunication. *Artif. Intell. Res.* **6**(2), 93–99 (2017)
8. Hung, S.-Y., Yen, D.C., Wang, H.-Y.: Applying data mining to telecom churn management. *Expert Syst. Appl.* **31**, 515–524 (2006)
9. <http://scikit-learn.org/>
10. Kai, W.: Research and Application of Data Mining in the Model of Off-grid Users. Zhengzhou University, Master Thesis (2014)

Evaluating LTE Service Performance for High-Speed Rail Cells via User Classification Model

Alexis Huet¹(✉), Mantian (Mandy) Hu², Jibin Wang¹,
and Ye Ouyang³

¹ Nanjing Howso Technology, Nanjing, China
alexis@howso.cn

² The Chinese University of Hong Kong, Hong Kong, China

³ Columbia University, New York, USA

Abstract. Development of LTE wireless network raises new ways for customers to communicate, by offering the possibility to access the network anywhere and anytime. As a result of the investments made over the last few years, this network has achieved widespread popularity. Some specialized infrastructures have been raised to address the special needs of customers. One of this need is to give the opportunity to connect to the network while journeying in a high-speed train. Mobile operators have elaborated special cells to manage high-speed train passengers, taking into account the characteristics of this environment. A key issue is to monitor the effective usage of the cells, by checking that passengers connect to those special cells (and not to the neighbor common cells), and that other users connect to common cells (and not to the special ones). For this purpose, a monitoring system of cells based on data analytics is detailed. This system identifies service performance of each cell, by pointing out common cells where high-speed train passengers are attaching to, and special cells where too many non-passengers are connecting to. The whole system helps mobile operators to elaborate strategies to improve service performance, by determining which cells should be tuned.

Keywords: Cell performance evaluation · High-Speed rail · Cell monitoring
Machine learning · Curve classification

1 Introduction

In the last few years, LTE wireless network has become more ubiquitous and has induced a major increase of infrastructure dedicated to its usage. By the end of 2016, the number of subscribers worldwide was over 915 millions [1], with an important dynamic in Asia. Meanwhile, high-speed rail rapidly developed in China, with over 1.1 billion trips in 2015 [2], making China as the world's busiest network. In this environment, special cells must be raised by mobile operators to offer a good signal quality. Cell investments for those special cells are based on railway traffic and are sized accordingly.

In this article, a *special cell* refers to a cell station specifically dedicated for high-speed train passengers. Those special cells are positioned along the high-speed train line and are identified as such. A *common cell* refers to any other cell station located around the high-speed train railway, which typically should serve other users. The union of special and common cells are the *selected cells*. In the same way, a *train user* is a passenger of a high-speed train, whereas a *common user* refers to any other customer of the network.

One central question is to monitor service quality of the selected cells. The traditional method to evaluate those cells is only related to the number of customers connecting to them, and does not distinguish different types of users. Under the environment composed of two types of cells, this method generally fails to identify a misuse of some cells, corresponding to the case when train users actually connect to common cells; or when common users connect to special cells.

The present article introduces a monitoring system designed to evaluate LTE service performance of cells located around the train railway. To the best of our knowledge, this is the first work to do so. The main aim is to classify train users and common users from analytic data collected by cell stations. Service performance evaluation is deduced from this classification. Two aspects are important to evaluate the service performance. First, we identify common users connecting to special cells. The proportion of such user should be reduced to prevent overloading of the special cells. Second, we identify train users connecting to common cells. Such train users may experience a degradation of the quality of service, induced by the rapid switches between normal and special cells.

The system is based on big data analytics and combines available data to evaluate the cells performance. This means that a large quantity of data is processed and combined to provide summarizing reports. From raw data gathered on a daily basis, we extract important characteristics for each user and apply a classification algorithm to obtain robust results. Once the model has been trained, the system is fully automatic and does not need fine tuning.

The rest of the article is organized as follows: In Sect. 2, we present a brief overview of existing research. In Sect. 3, data retrieved for cell evaluation are presented. In Sect. 4, the methodology to model data is detailed. In Sect. 5, we describe cell evaluation results. Finally, conclusion is given in Sect. 6.

2 Research Background

High-speed train cellular networks have been studied mainly through technical specifications and infrastructures. An early work to provide network service for train passengers has been developed in [3]. Then, GSM-R gradually moved to LTE-R to provide high-speed connections in trains [4, 5]. Optimization of those networks is mainly guided by characteristics of the network, for example based on handover optimization [6], packet allocation problem among services [7] or modeled scenario [8].

About curve and time series classification, several approaches are available. This field is an active area of research, and new methods are arising. The first classification

approach is to extract features from the time series before applying a classification model [9–11]. This method has the advantage to use a white box model, therefore it will be chosen in our methodology. An other approach is to derive features automatically with wavelet techniques such as dynamic time warping (DTW) [12–14]. Finally, some deep learning methods has became to appear [15], but those methods continue to behave as a black box during features detection.

3 Collected Data and Preprocessing

3.1 Raw Data

As the whole process is driven by data, the initial step is to retrieve raw data. We assume that a train line of interest and a day have been selected. Two files are collected: Information about cells of interest (*cell info* in the following) and log file of connections of users to those cells during this day (*user data* in the following).

Cell info contains the list of special cells located near the selected train line, and of common cells within one kilometer from any special cell. Each cell is identified with its position (longitude, latitude).

User data contains all connections between users and the selected cells, each connection being time stamped. Time of connections are irregular, which means that connections might not happen at a regular time interval.

In the following implementation, data has been retrieved from a major Chinese telecommunication company in April and May 2017. The number of special cells related to a train line is 50, completed with 450 common cells. The number of users connecting to those cells within one day is about 1 million, with 28 millions connections.

3.2 Inference of the Train Line

We deduce the train line from cell info, by listing positions of special cells and applying a regression algorithm [16, 17]. An exemplary result after applying this process is shown in Fig. 1. In this figure, each point in red corresponds to the position of a special cell (note that those special cells are along the train line, not inside the train). The black curve represents the inferred train line. In addition, each point in blue is a common cell within one kilometer distance from a special cell.

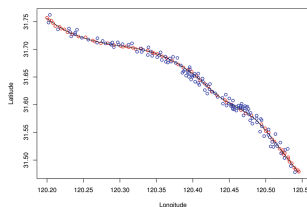


Fig. 1. Identification of a train line from spatial position of cells. Axes for longitude and latitude are degrees

3.3 Data Frame of Interest

From the representation of the train line, we deduce the distance in kilometers between each cell and the train line. From this, we get the distance between each cell and the train station along the train line.

Using this information, we obtain the data frame of interest, built from user data by adding two columns: The first column is the distance in kilometer between the connected cell and the train station; The second column is a binary value indicating whether the connected cell is a special cell or a common cell.

Exemplary rows of this data frame of interest in shown in Table 1.

Table 1. Exemplary rows of the data frame of interest.

User	Time stamp	Cell	Cell distance to station	Is high-speed cell?
User 1	08:35:40	Cell 1	31.40 km	False
User 1	10:24:14	Cell 1	31.40 km	False
User 2	20:28:39	Cell 2	1.13 km	True
User 2	20:29:29	Cell 3	3.74 km	True

4 Processing and Modeling Methodology

There are two main stages in the process: the training stage and the production stage.

The aim of the training stage is to build the classification model and is illustrated in Fig. 2: From the data frame described in Sect. 3.3, we get the behavior of each user along the train line, which is plotted for a sample of users. At this point, manual classification of the plots is performed to get a training set for our classification. Then, we extract specific features for each user from the data frame. From those, we build the classification model to automatically separate train users and common users.



Fig. 2. Training stage methodology.

The steps of the production stage is as follows: From data frame related to a new day, we use the trained model to classify users. We then summarize relationships between users and cells from this classification.

Each step of the process is detailed in the following paragraphs. The “plotting”, “manual classification” and “classification model” steps are only computed during the training stage.

4.1 Plotting

The plotting step consists of representing, for each user, its distance to the train station as a function of the time. In practice, only a sample of users is plotted (for example 4000 users). Different examples of resulting plots are shown in Fig. 3a, b, c and d. On those plots, each point represents a connection between the user and a cell. The point is red if the cell is a special cell, and black otherwise.

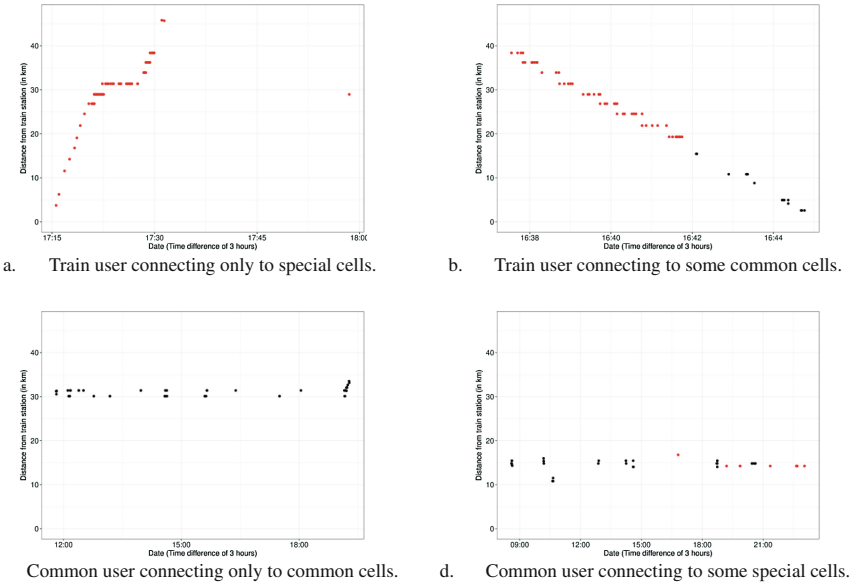


Fig. 3. Four examples of two-dimensional representation of a user along the train line through his connections to the cells.

4.2 Manual Classification

The aim of the manual classification step is to classify a sample of users. For each user of the sample, we would decide if this is a train or a common user. For this sample of users, the classification is performed manually (this means each element of the sample data is labeled). This methodology has been chosen to ensure reliability of the global model: Classification of user is not based on an arbitrary set of values, but will be learnt from this sampled classification. The manual classification is guided by the shape of the plotted figures: If the user quickly moves from one station to the other station, this user is labeled as a train user (see Fig. 3a and b). It is labeled as a common user in other cases (see Fig. 3c and d). This manual classification leaves little room for subjective interpretation, and can take into consideration all possible characteristics leading to a train user or a common user.

For example, the number of *manually-classified* users can be 4000, among a total number of 1 million users.

4.3 Features Extraction

In the features extraction step, we convert the time series related to each user (indicating its distance to the train station for irregular time stamps) into a set of features. Features are selected to collect the characteristics of the time series.

The most important feature to retrieve is the speed of the user. However, the naïve way based on instantaneous speed is not reliable: In a short interval of time, the distance to the train station (as seen in the data frame) can change abruptly, leading to instantaneous speed over 600 km/h. For example, this behavior occurs in Fig. 3c before 12:00. To circumvent this problem, we consider the average speed over different durations. Specifically, for a specific duration of time (for example 30 min), we compute the maximum distance the user traveled during this interval over the day. Those maximum distances are computed for different duration of time.

For each user, we retrieve additional features, primarily related to the relation between users and cells (such as the number of connections to any cell or the proportion of connection of the user to a special cell).

After extraction of all features, each user is described with a vector of 21 numeric features.

4.4 Classification Model

The following step is dedicated to build the model using a classification model. For this purpose, we combine: (1) Vector of extracted features and; (2) Information whether the user is a train user or a common user (from manually-classified users). Many classification algorithms exist in the literature and are generally based on machine learning tools. A convenient algorithm is random forests [18]. The reason to select this algorithm is its ability to extract important features while doing the classification (by ranking variables by their importance).

To ensure reliability of the results, we separate the manually-classified set into two parts: the training set, containing 80% of the users; and the test set. The classification model is trained on the training set. Then, the model is tested on both training and test sets. Once the model has been trained, every new user can be classified through their extracted features.

4.5 Summarizing Tables

In the production stage, the model is applied to classify each new user. The last step of the process is to obtain summarizing tables. Each table is able to monitor a key aspect of cells' service performance and is updated every day, including:

- *Incorrect service of special cells* table lists the connections to special cells by common users, and is able to highlight special cells serving common users too often;
- *Switches from special cell to common cell by train users* table, which allows to spot cells with many switches that should be improved.

5 Implementation and Results

The methodology is fully implemented with R [19] in a framework able to output summarizing tables automatically.

After having built the model, classification accuracy reaches 99% for the training set and 98% for the test set. This accuracy is high because we extracted and selected features directly connected to the behavior of users along the train line (and it is not 100% because 1. classification algorithm can misclassify; and 2. manually classified objects contain a few mistakes). The most important features selected by the random forest algorithm are related to the distance traveled in 25, 45 and 20 min.

During production, many indicators can be monitored. For example, we can be interested by spotting switches from special to common cell by train users. In the following Fig. 4, we highlight the 5 special and the 5 common cells with the most such switches. We can observe that 3 special and 2 common cells should be inspected to prevent such switching. Also, during peak period (May 1), 2 common cells are particularly overloaded.

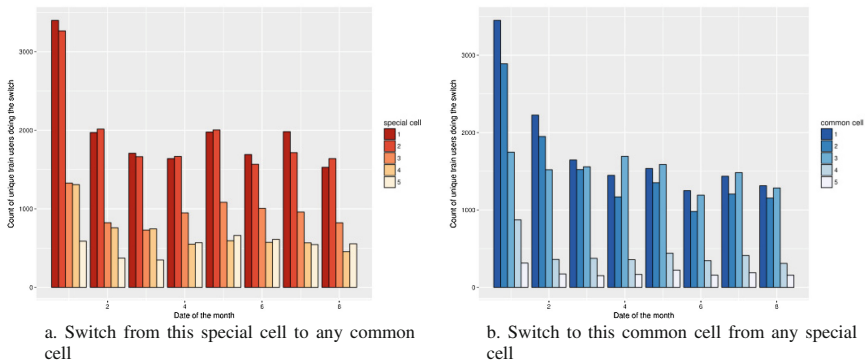


Fig. 4. Number of unique train users switching from May 1 to May 8, for the five cells with the most switches.

6 Conclusion

On the whole, the system gives robust evaluation of LTE service performance for high-speed rail cells. It leads to an efficient monitoring of special and common cells located along the train line, by initially identifying train users from other users. The classification model of users is built with machine learning tools and its efficiency is tested. The implementation is fully data-driven and is automatic once the initial model has been trained. Robustness of the system is checked during the procedure and accuracy is also tested. Connections to LTE cells are summarized, allowing to evaluate service performance of each cell and of the global network of cells. Thus, mobile carriers can elaborate strategies of tuning and development for each type of cell. In this way, cells network around train lines can be improved significantly.

References

1. Telegeography Global Comms Forecast service, telegeography.com/research-services/global-comms-forecast-service/index.html. Accessed July 13 2017
2. Xinhuanet, news.xinhuanet.com/english/2016-07/21/c_135530835.htm. Accessed 13 July 2017
3. Abrishamkar, F., Irvine, J.: Comparison of current solutions for the provision of voice services to passengers on high speed trains. In: 52nd Vehicular Technology Conference on IEEE-VTS Fall VTC 2000, vol. 5, pp. 2068–2075 (2000). IEEE
4. He, R., Ai, B., Wang, G., et al.: High-speed railway communications: From GSM-R to LTE-R. *IEEE Veh. Technol. Mag.* **11**(3), 49–58 (2016)
5. Han, T., Ansari, N.: RADIATE: radio over fiber as an antenna extender for high-speed train communications. *IEEE Wirel. Commun.* **22**(1), 130–137 (2015)
6. Yang, F., Deng, H., Jiang, F., et al.: Handover optimization algorithm in LTE high-speed railway environment. *Wireless Pers. Commun.* **84**(2), 1577–1589 (2015)
7. Xu, S., Zhu, G., Shen, C., et al.: Utility-based resource allocation in high-speed railway wireless networks. *EURASIP J. Wirel. Commun. Networking* **2014**(1), 68 (2014)
8. Ai, B., Cheng, X., Kürner, T., et al.: Challenges toward wireless communications for high-speed railway. *IEEE Trans. Intell. Transp. Syst.* **15**(5), 2143–2158 (2014)
9. Fulcher, B.D., Jones, N.S.: Highly comparative feature-based time-series classification. *IEEE Trans. Knowl. Data Eng.* **26**(12), 3026–3037 (2014)
10. Wiens, J., Horvitz, E., Guttag, J.V.: Patient risk stratification for hospital-associated C. diff as a time-series classification task. In: *Advances in Neural Information Processing Systems*, pp. 467–475 (2012)
11. Serra, J., Arcos, J.L.: An empirical evaluation of similarity measures for time series classification. *Knowl.-Based Syst.* **67**, 305–314 (2014)
12. Hills, J., Lines, J., Baranauskas, E., et al.: Classification of time series by shapelet transformation. *Data Min. Knowl. Disc.* **28**(4), 851–881 (2014)
13. Petitjean, F., Forestier, G., Webb, G.I., et al.: Dynamic time warping averaging of time series allows faster and more accurate classification. In: 2014 IEEE International Conference on Data Mining (ICDM). IEEE, pp. 470–479 (2014)
14. Sun, M., Wang, Y., Zhang, X., et al.: Feature selection and classification algorithm for non-destructive detecting of high-speed rail defects based on vibration signals. In: *Proceedings, 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 819–823. IEEE (2014)
15. Zheng, Y., Liu, Q., Chen, E., et al.: Time series classification using multi-channels deep convolutional neural networks. In: *International Conference on Web-Age Information Management*, pp. 298–310. Springer, Cham (2014)
16. Wood, S.N.: Modelling and smoothing parameter estimation with multiple quadratic penalties. *J. Roy. Stat. Soc. B: Stat. Methodol.* **62**, 413–428 (2000)
17. Hastie, T.J., Tibshirani, R.J.: *Generalized Additive Models*, vol. 43. CRC Press, Boca Raton (1990)
18. Liaw, A., Wiener, M.: Classification and regression by randomForest. *R News* **2**(3), 18–22 (2002)
19. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2016), <https://www.R-project.org/>

Research on High-Efficient Dynamic Evaluation Method of Operation Stability in Mobile Radio Network

Jian Guan¹(✉), Wensheng Li², Haina Ye¹, Jie Gao¹,
Yongfeng Wang¹, and Xinzhou Cheng¹

¹ China Unicom Network Technology Research Institute, Beijing 100048, China
guanjq9@chinaunicom.cn

² China Information Technology Designing and Consulting Institute,
Beijing 100048, China

Abstract. Traditional analysis of the Mobile Radio Network Operation Stability (MRNOS) is usually based on several indicators or human experiences, which leads to a great difficulty in carrying out a systematic evaluation. A Comprehensive Evaluation Algorithm (CEA) was proposed in order to solve this issue. However, CEA ignores the efficiency and objectivity in practical implementation, therefore, in this paper, an improved High-efficiency Dynamic Evaluation Method (HDEM) is proposed to solve three main problems of the previous method, including high complexity, non-objectivity and low efficiency.

Keywords: Mobile radio network · Operation stability · Evaluation method
Efficiency

1 Introduction

Long-term stable operation of mobile network is the premise of providing users with sustained and stable business. Radio network is responsible for the first step of users' accessing, therefore, the health of mobile radio network determines whether users can access and be able to get good perception. Compared to traditional evaluation methods based on individual indicators and parameters [1] or artificial experience [2, 3], the CEA [4] can evaluate the operation stability of base stations or mobile radio network more completely and effectively.

However, with rapid development of the network scale and improving requirement of fine management [5], there are more and more needs to reduce the operation and maintenance (O&M) cost pressures and to attend to the needs of staff members who sweated in the O&M work, the CEA exposes its deficiencies including:

- (1) *Difficulty in operation including data acquisition and maintenance.*
- (2) *Non-objectivity of weighting factor determination.*
- (3) *Inefficiency of operation in the current network.*

Focusing on these problems, this paper presents a high-efficiency dynamic evaluation method from the perspective of reducing complexity, enhancing the objectivity of evaluation and achieving the dynamic and cyclical model of evaluation.

2 Deficiencies of CEA

2.1 Difficulty in Daily Operation

The CEA needs to use a lot of basic data that related to O&M operation. All data is grouped into three types, including key Performance Indicators (KPIs), external influence factors and basic parameters. These three types of data are sequentially represented as V, W and B, as shown in Table 1.

Table 1. Three types of data in CEA.

Selected type of factors	Details	Identity
Stability KPIs	Equipment alarm indicators Network performance indicators	V
External influence factors	Standardization level of ancillary facilities Equipment consumption level Artificial maintenance condition level Emergency situation response level	W
Basic parameters	Base Station (BS) ID Cell ID BS location BS level BS region value level	B

Based on above data and its classification, the CEA has design a basic evaluation formula of MRNOS expressed as (1).

$$Score = V \times W \times B \tag{1}$$

There is nothing wrong with the formula itself, but the problem lies in the data sources of the W and B factors. Standardization level of ancillary facilities (W1), equipment consumption level (W2), artificial maintenance condition level (W3) and emergency situation response level (W4) of data W can only be obtained by manual inspection in base stations or background account checking of O&M records. This kind of situation could bring more difficulty and uncertainty to stability evaluation, especially in some VIP scenarios such as government agencies, airports, railway stations and subway stations and tunnels. Due to irresistible factors [6], manual inspection of base stations in VIP scenarios can only be implemented more than once a month (or even a quarter or two). That kind of data update cycle will reduce W's contribution to the final Score. B1 and B2 of data B are also faced with similar problems.

2.2 Non-objective Determination of Weighting Factors

In Formula (1), the calculation weight of each level is determined by the analytic hierarchy process (AHP) algorithm as well as expert scoring, as an example shown in Table 2.

Table 2. An example of the importance judgment table for experts scoring in CEA.

Content	Significance	Degree	Score (abs)	Choice (one only)
Relative importance between W_i and W_j	$W_i > W_j$	Totally agree	4	
		Partially agree	3	
		Basically agree	2	√
		Probably agree	1	
	$W_i = W_j$		0	
	$W_i < W_j$	Probably agree	-1	
		basically agree	-2	
		Partially agree	-3	
		Totally agree	-4	

AHP brings an ability to judge consistency in analysis process which helps reduce anomalies and heighten objectivity. However, expert scoring cannot help the same. The basic process of expert scoring in CEA includes three steps, as simply described in Fig. 1.

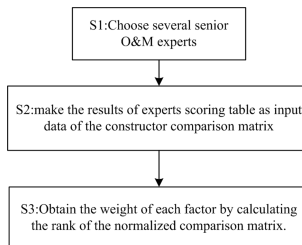


Fig. 1. Three steps of expert scoring in CEA.

So the key factor that affects the calculation process of the evaluation method is whether these chosen experts deserve their name. In other words, the experts themselves should know every detail of O&M work of mobile radio network very well and the selection process must be objective. However, due to man-made operation, this process can hardly be absolutely objective.

2.3 Low Efficiency in Evaluation

Compared to traditional analysis of MRNOS, CEA has already improved the evaluation efficiency while improving the comprehensiveness and accuracy of stability evaluation, but its evaluation efficiency is still not high enough. For example, as described in Sect. 2.1, the highest frequency of CEA implementation is once a month, which is far from the demand of dynamic cyclical evaluation.

3 High-Efficiency Dynamic Evaluation Method

In order to solve these three problems above, the HDEM is proposed in this paper from the point of view of reducing high complexity, enhancing evaluation objectivity and achieving the dynamic and cyclical model of evaluation.

3.1 Reducing Complexity

To reduce the complexity of evaluation method, first thing is try to use automatic data acquisition system instead of artificial collection. Take China Unicom for example, by upgrading and using OSS 2.0 which is an O&M system with ESB service, many kinds of data including KPIs, measurement reports, alarms etc. can be automatically and efficiently collected from the whole network into the OSS system [7], and this provides a basis for improving the frequency and efficiency of stability evaluation.

The second one is reducing complexity of W and B in Formula (1). Update the contents of W, modify four children of W in CEA into three in HDEM, represented as (W'1, W'2, W'3), and reduce the children of B from five to two, represented as (B'1, B'2), as shown in Table 3.

Table 3. Three types of data in HDEM.

Selected type of factors	Details	Identity
Stability KPIs	V'1: Equipment alarm indicators V'2: Network performance indicators	V'
Additional influence factors	W'1: Equipment max consumption W'2: Maintenance condition records W'3: Emergency situation response time	W'
Basic parameters	B'1: BS level B'2: BS region value level	B'

3.2 Enhancing Objectivity

When try to enhance evaluation objectivity in HDEM, there are basically two possible recommendations based on expert scoring method. First is increasing the minimum number of experts, hope that the increasing number of real experts could enhance objectivity. Second is to design a new calculation method of weighting factors. The concept of crowdfunding scoring is borrowed here in HDEM. The original idea of crowdfunding, as its name suggests, is a funding method where common people like you and me, henceforth the crowd, fund your personal or business project with their own money. Now crowdfunding scoring represents a scoring method using the help of the crowd.

In HDEM, the crowdfunding scoring method tries to get as much data as possible by the help of all the O&M engineers who have rich front-line O&M experience rather than less senior experts. Collection modes of such data can be very flexible and diverse, not only through the traditional way of paper questionnaire but also through social mobile apps, such as Facebook, Twitter, WeChat, Weibo etc., to get more efficiency.

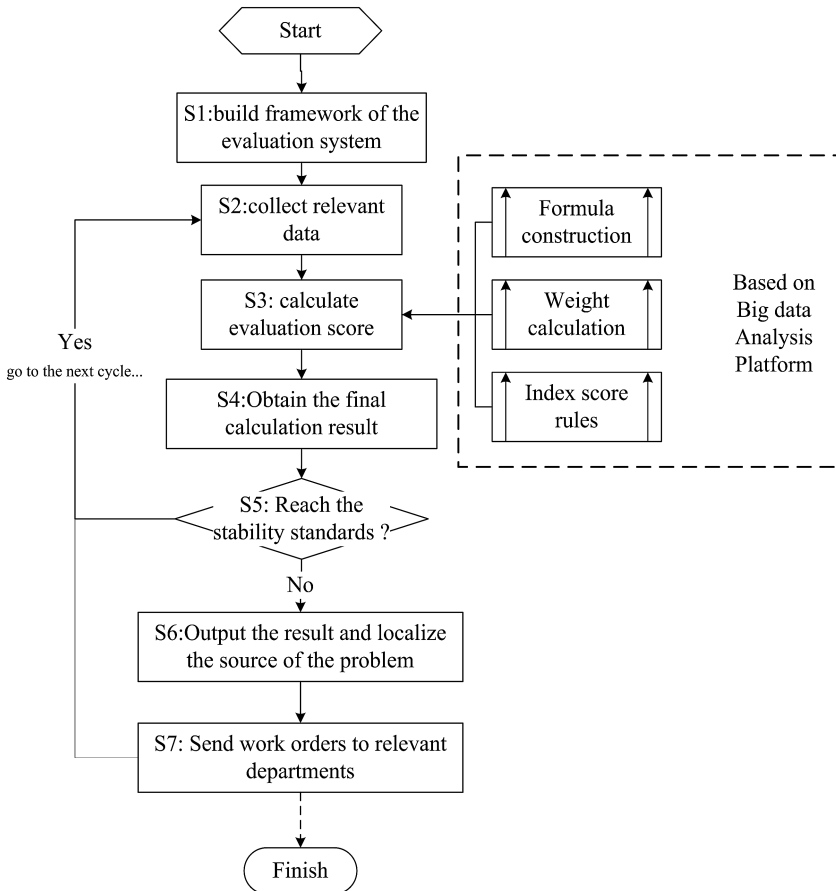


Fig. 2. An improved evaluation process of MRNOS.

3.3 Achieving the Dynamic and Cyclical Model

As described above, because of manual data acquisition, the stability evaluation work using CEA is basically a single static process with fixed period and operation procedures. There is no condition of dynamic and cyclical operation of CEA. Now in HDEM, an improved evaluation process of MRNOS is proposed as shown in Fig. 2.

To achieve the goal of dynamic and cyclical evaluation, a high performance platform should be built. Taking into account the rapid development of network size and the data volume, this paper designs a stable and efficient carrying system for HDEM based on the Big Data Analysis and Design Support Platform (BDADSP), the system architecture is shown in Fig. 3.

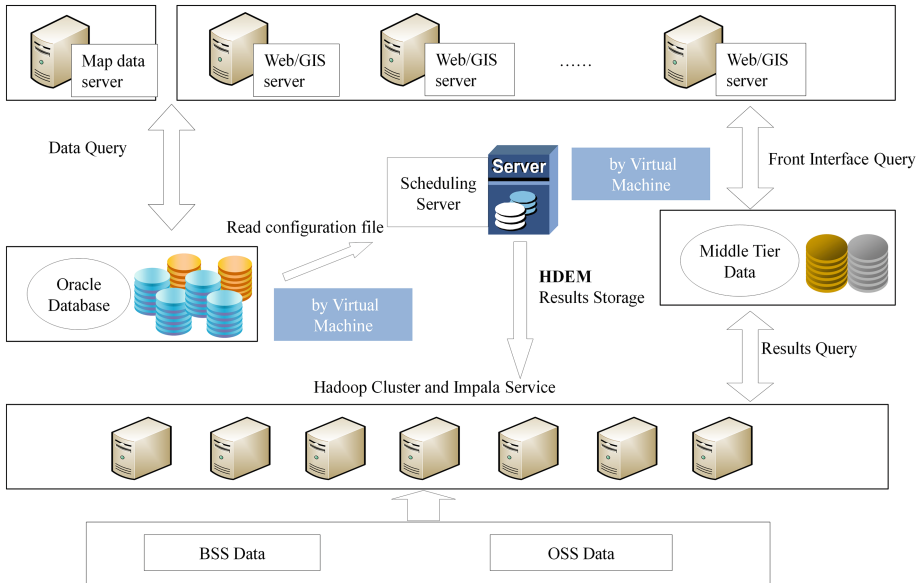


Fig. 3. Implementation of HDEM based on big data analysis platform.

As shown in Fig. 3, there is a Hadoop cluster in the BDADSP to provide Impala service [8], which is also the middle tier framework to store process data and several servers to provide map, web and GIS services [9]. The implementation of HDEM, as well as other services, are supported on this platform in form of virtual machines. The advantage of using this kind of design is that the system can be flexible with the network scale expansion and traffic data growing.

A test-running stability evaluation system using HDEM has been built in province J with configurations described in Table 4. The implementation of HDEM in the test system now is fully cyclical and automated, it meets the needs of the O&M staff in the whole province.

Table 4. The configuration of test-running system using HDEM.

Category	Quantity	CPU (cores)	RAM (GB)	Memory (TB)
Computing resources		24	300	20
Virtual machine resources	8	16	64	1

3.4 Comparison of CEA and HDEM

Choose province J in south China as a sample, make stability evaluation of WRNOS of China Unicom FDD-LTE network in scale of the whole province. The comparison of evaluation efficiency between CEA and HDEM is shown in Fig. 4.

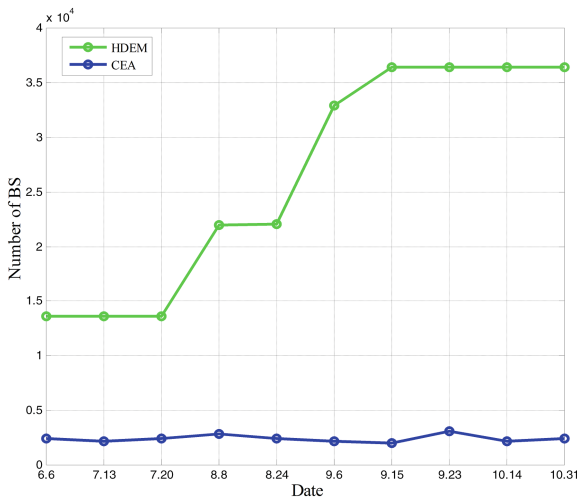


Fig. 4. Comparison of evaluation efficiency of CEA and HDEM.

It is observed from Fig. 4 that HDEM has a much higher efficiency than CEA. On one hand, the number of evaluated BS using CEA is subject to artificial data update cycle, there is little change in the date range. On the other hand, as the test-running system is getting increasingly stable, the number of evaluated BS using HDEM is growing rapidly over time and closing to the total number of BS which is about 36,000 in the province.

4 Conclusions

This paper presents an improved high-efficient dynamic evaluation method to solve three main problems of the previous one, including complexity, non-objectivity and low efficiency. The HDEM is proposed in this paper from the point of view of reducing complexity, enhancing evaluation objectivity and achieving the dynamic and cyclical

model of evaluation. Comparison of CEA and HDEM in current mobile radio network shows that HDEM has a better efficiency of stability evaluation. More measures should be taken to optimize the big data platform to ensure the HDEM system operation smoothly and efficiently.

References

1. Nunzi, G., Jactat, C., Futaki, H.: Method and system for measuring network operation related parameters in a mobile communication network. US Patent 9264921, WO (2014)
2. Cao, Y., Sun, Z.: Routing in delay/disruption tolerant networks: a taxonomy, survey and challenges. *IEEE Commun. Surv. Tutorials* **15**(2), 654–677 (2013). IEEE
3. Cao, Y., Wang, N., Kamel, G., Kim, Y.: An electric vehicle charging management scheme based on publish/subscribe communication framework. *IEEE Syst. J.*, 1–12 (2015). IEEE Press
4. Guan, J., Cao, L., Chen, W., Cheng, X., Xu, L., Cao, X.: A comprehensive method of evaluation for wireless network operation stability. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 342–346. IEEE Press, Qingdao (2016)
5. Xu, L., Cao, Y., Chai, K., Luan, Y., Liu, D.: Cooperative mobility load balancing in relay cellular networks. In: IEEE International Conference on Communication, pp. 141–146. IEEE Press, Xi'an (2013)
6. Ye, H., Cheng, X., Yuan, M., Xu, L., Gao, J., Cheng, C.: A survey of security and privacy in big data. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 268–272. IEEE Press, Qingdao (2016)
7. Chao, K., Wang, P., Xu, L., Wu, D., Cheng, X., Mu, M.: A novel big data based telecom user value evaluation method. In: 1st International Congress on Signal and Information Processing Networking and Computers, pp. 375–384. CRC Press, Beijing (2015)
8. Mayer-Schönberger, V., Cukier, K.: *Big Data: A Revolution That Will Transform How We Live, Work and Think*, 1st edn. John Murray, London (2013)
9. Cao, Y., Sun, Z., Wang, N., Riaz, M., Cruickshank, H., Liu, X.: Geo-graphic-based spray-and-relay (GSaR): an efficient routing scheme for DTNs. *IEEE Trans. Veh. Technol.* **64** (4), 1548–1564 (2015). IEEE Press

A Novel PCI Optimization Method in LTE System Based on Intelligent Genetic Algorithm

Ao Shen¹(✉), Bao Guo², Yan Gao³, Tao Xie⁴, Xiaochun Hu⁵, Yang Zhang⁶, Jinhu Shen¹, Yuan Fang¹, Guozhi Wang⁷, and Yi Liu⁸

¹ China Mobile Group Design Institute Corporation, Beijing 100080, China
shenao@cmdi.chinamobile.com

² Shanxi Branch of China Mobile Communication Corporation,
Taiyuan 030032, China

³ Henan Branch of China Mobile Communication Corporation,
Zhengzhou 450018, China

⁴ Chongqing Branch of China Mobile Communication Corporation,
Chongqing 401121, China

⁵ Shanxi Branch of China Mobile Communication Corporation,
Xian 710000, China

⁶ China Mobile Communication Corporation, Beijing 100033, China

⁷ Zhejiang Branch of China Mobile Communication Corporation,
Hangzhou 710000, China

⁸ Shandong Branch of China Mobile Communication Corporation,
Jinan 710000, China

Abstract. The PCI mod3 interference in the Long Term Evolution (LTE) system will cause degradation of radio access, handover, and quality of service, which seriously pull down user feeling. We propose a novel scheme utilizing advanced intelligent Genetic Algorithm to mitigate PCI mod3 interference which is based on data resources of Drive Test, handover, and the Measurement Report. The practical network trials demonstrate our scheme dramatically reduced the operational complexity and PCI mod3 interference in LTE system.

Keywords: Long Term Evolution (LTE) · Physical Cell Index (PCI) Optimization · Interference · Genetic algorithm · Drive Test (DT) Measurement Report (MR)

1 Introduction

Radio network interference suppression is an important study case of LTE network optimization. The interference would affect the user's voice quality, data throughout, handover performance etc. Among various types of network interference, the PCI mod3 one is an inherent systematic disease which would hardly be eradicated. As the 4th generation (4G) network is booming, the average macro-site distance is getting smaller as well as the average neighbor cells ascend. Thus, the collision risk of primary synchronized signal (PSS) of LTE system is raised up, which is so-called PCI mod3 interference [1].

Technically, there is no efficient way to resolve PCI mod3 interference so far. The conventional solution is based on geographical observation in light of engineer's

artificial experience at the expense of degenerated optimization performance and the time-consuming [2]. In other hand, vendors' software regarding this issue is always incompatible to the practical situation, which means low efficiency calculation and unsatisfactory optimization verdict [3]. A newly intelligent Genetic Algorithm based PCI mod3 interference mitigation scheme is proposed and applied in real network optimization. The trial results demonstrate its distinguished performance in terms of accuracy and efficiency.

The rest of the paper is organized as follows. Section 2 describes PCI configuration rules in the 4G network. Sections 3 and 4 constructs primary objective optimization model and illustrates the details of advanced intelligent Genetic Algorithm, respectively. There are some demonstrations of practical network solution in Sect. 5 and Sect. 6 concludes the paper.

2 PCI Configuration Principle

The cell search process for 4G user equipment (UE) includes two sequential important steps, which is the acquisition of primary synchronization signal (PSS) and secondary synchronization signal (SSS), respectively. The allocated PCI is from 0 to 503 by network, which is calculated by $\text{SSS code sequence ID} \times 3 + \text{PSS code sequence ID}$. Here, the SSS code sequence range is from 0 to 167 and the PSS code sequence range is only from 0 to 2, respectively. Note, the PSS sequence ID uniquely determines the distribution pattern of the cell reference signal (CRS).

Generally, there are three categories of interference regarding PCI arrangement. Firstly, it calls PCI conflict which means the PCI of UE-camped cell is same as the one of adjacent cells deployed by the identical frequency. Secondly, the PCI confusion is determined by the identical PCI arrangements among adjacent cells assuming same frequency networking as well. Finally, it is so-called PCI mod3 interference which is actually defined by identical CRS pattern among adjacent cells of same frequency [6]. The previous two cases would cause call drop, access failure and/or handover failure. However, those would be relatively easier tackled with in the practical network optimization due to longer multiplexing distance (504) compared with the one of PCI mod3 (3). The PCI mod3 interference between adjacent cells is universal in the practical LTE network due to low multiplexing distance. Therefore, the fundamental configuration principle is to maximally diversify the CRS pattern among adjacent cells which also means maximizing the difference of PSS sequence code for adjacent cells.

3 PCI Optimization Model Setup

While the number of one cell's neighbor cells is greater than 2, the mod3 interference is hardly inevitable and the PCI mod3 interference has become the most common interference in the network. Therefore, it is necessary to optimize the PCI configuration of the practical network to minimize the mod3 interference caused by the unreasonable PCI configuration.

From the causes of the mod3 interference, the possibility of mod3 interference is much larger when there are many intra-frequency neighbor cells with similar signal strength in the service cell and PCI mod 3 results of the service cell and neighbor cells are the same. The MR [5] data and DT data in the practical network include the frequency, PCI and RSRP information of the serving cell and all neighbor cells. Through the analysis of the MR and DT test data and combining the resources engineering parameters and the handover data of network management, we create a neighbor cell-level interference matrix, and build mod3 interference calculation model in a certain optimization area. Finally, we found a PCI configuration with the minimum interference in the field optimization field.

If there are N cells in the optimization field, the source cell is i, and the target cell is j, and the interference value between the two cells is equal to $PW_{ij} \cdot PMOD_{ij}$. Among them,

$$PW_{ij} = w_1 \cdot \left(\frac{1}{R_{ij}} + \frac{CR_{ij}}{CR_{max} - CR_{avg}} \right) + w_2 \cdot \left(\frac{1}{D_{ij}} + \frac{CD_{ij}}{CD_{max} - CD_{avg}} \right) + w_3 \cdot \frac{CS_{ij}}{CS_{max} - CS_{avg}} \tag{1}$$

The three parts of formula consider the influence of correlation coefficient about MR data, DT data and handover data respectively. The w which reflects the weight of different data source can be adaptive adjusted according to the optimization demand. Taking MR data as an example, the correlation coefficient between the source cell i and the target cell j in the MR is the mean value of the absolute value of the RSRP difference of the neighborhood pair, the calculation of R_{ij} is as follows:

$$R_{ij} = \sum_{n=1}^{CR_{ij}} \left| RSRP_i^n - RSRP_j^n \right| / CR_{ij} \tag{2}$$

Among them, $RSRP_i^n$ represents the signal strength RSRP value of cell i in the nth sample point in the MR data. CR_{ij} is the number of RSRP sampling points containing the target cell j in the MR data of the source cell i. CR_{max} represents the maximum number of sampling points of neighbor cell measured by source cell in MR, and CR_{avg} represents the mean number of sample of neighbor cell measured by source cell in MR.

$$PMOD_{ij} = \begin{cases} 1, & \text{(If there is a mod 3 interference between cell i and cell j)} \\ 0, & \text{(If there is no mod 3 interference between cell i and cell j)} \end{cases}, PMOD_{ij}$$

represents the mod 3 interference between cell i and cell j.

Then the interference matrix I of the entire optimization region can be expressed as:

$$I = \begin{bmatrix} PW_{11} & PW_{12} & \dots & PW_{1N} \\ PW_{21} & PW_{22} & \dots & PW_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ PW_{N1} & PW_{N2} & \dots & PW_{NN} \end{bmatrix} \cdot \begin{bmatrix} PMOD_{11} & PMOD_{12} & \dots & PMOD_{1N} \\ PMOD_{21} & PMOD_{22} & \dots & PMOD_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ PMOD_{N1} & PMOD_{N2} & \dots & PMOD_{NN} \end{bmatrix} \tag{3}$$

In order to minimize total interference of the optimal field, the overall function can be expressed as: $F = \sum_{i=0}^N \sum_{j=0}^N I_{ij}$, the optimization variable is the value of the PSS of the cell in the optimized field.

4 PCI Optimization Algorithm Design

From objective function, the computational speed of the optimal solution is limited by the number of cells in the optimized region. In order to accelerate the convergence of the algorithm, we introduce the genetic algorithm to output the PCI configuration of the optimization area. The detailed steps are as follows.

4.1 Determine the Initial Value

The heuristic linear search determines the first row of the population matrix in the genetic algorithm.

4.2 Elite Algorithm Filters the Optimal Solution

The best chromosomes are retained in each generation of the population matrix.

4.3 Crossover, Mutation

The two critical steps in the genetic algorithm are the mating and mutation. The mating method uses the full mating of the algebra. That is, each gene of the parental chromosome of the selected parent carries out sufficient information to interact with each other to produce the para gene of the chromosome. And the mating probability can be adjusted. The mutation probability determines that the search space can be adjusted appropriately so as to avoid the search process remaining in the local space. The probability of mutation is too small to ensure that the local search space is searched for the optimal solution, and the probability of mutation is too large may lead to excessive random search, waste efficiency. So the general experience value is set to 10%–20%. In the algorithm, the cost function is evaluated by machine learning. Once the cost function is retained in the local space, the search space is improved by adjusting the mutation probability, and the search efficiency is improved.

4.4 Heuristic Search

For large area where PCI needs to be optimized, the search algorithm for sparse matrices is easy to consume a great deal of time, and the search results are poor. Therefore, in the process of genetic algorithm, the search method of heuristics is introduced appropriately, which can accelerate the search efficiency and improve the search precision. At the same time, it can adaptively adjust the step size of the heuristic search and genetic algorithm, so that the intelligent fusion of two search methods can accelerate the search of the global optimal solution.

4.5 Convergence

The algorithm can adopt two convergence strategies. One is to set the iteration threshold, when the iteration threshold is terminated, the search is over and the general iteration threshold is set to 1/10 of the chromosomes number in the search matrix. Another iterative convergence can be terminated automatically according to the search performance. Once the cost function is consistent over a long time, it is assumed that the global optimal solution is found and the search is stopped.

4.6 Parallel Computing to Improve Operational Efficiency

The performance of the algorithm is based on the design of the software and the hardware support. In the search process, the method of parallel computing is adopted to further improve the calculation efficiency of the fixed population matrix.

In summary, we propose a genetic algorithm based on the PCI optimization algorithm, the specific steps are as follows:

Step 1: Select the optimization field and set the protection zone, enter the engineering parameters of the PCI optimization field, or select the optimization field by delineation of the map.

Step 2: Collect the resource engineering parameters, MR data, handover data, and DT data of the base stations in the optimization field and protection zone.

Step 3: Generate the neighbor cell-level interference matrix based on the massive MR data, DT data and handover data.

Step 4: Use genetic algorithm iteratively output the optimization region's PCI configuration.

Step 5: Output the PCI optimized configuration strategy for the optimization field and the change of Cost of the integrated interference value.

5 PCI Optimization Effect Verification

In order to verify the application effect of PCI optimization algorithm proposed in this paper, we optimize the optimization field of 591 cells in the current network. The following table shows the data of the practical network data before and after using the PCI optimization scheme proposed in this paper, and compares it with the optimization effects of the other three existing PCI tools. From the results, after the PCI optimization, the optimization field of 591 cells KPI indicators run smoothly. Handover success rate has been improved and the mod3 interference has been decreased, the optimization effect is good (Table 1).

In terms of system operation efficiency, the efficiency of this algorithm is greatly improved due to the combination of genetic algorithm and parallel computation. Using the PCI optimization method based on genetic algorithm proposed in this paper, it takes about 15 min to optimize 1000 cells in practical network, and the industry's existing PCI optimization tool for 1000 cells PCI optimization will take about 6 h. The substantial increase of optimal efficiency can effectively meet the needs of the actual optimization.

Table 1. Comparison of before and after PCI optimization.

Terms	Handover success rate	LTE Coverage (RSRP > -110 and SINR >= -3)	Average RSRP	Average SINR	SINR > 0 rate (%)	Continuous SINR difference mileage ratio	FTP average download speed (with dropped) (Mbps)	LTE download speed is greater than or equal to 5 M ratio (%)
Before optimization (mean)	97.01%	97.11%	-75.44	15.25	95.02%	3.05%	30.33	98.09%
Other existing scheme	97.71%	97.80%	-75.3	15.24	95.77%	1.22%	31.22	98.16%
This paper PCI optimization scheme	98.27%	97.87%	-74.19	16.28	96.34%	0.97%	34.58	98.72%

6 Conclusions

Our proposed scheme constructed a comprehensive neighbor cell-level interference matrix based on the weight of the adjustable data source, using MR data, DT data and handover data, respectively. At the same time, the PSS sequence code allocation is calculated by using the advanced intelligent genetic algorithm aiming to mitigate interference at optimum. Compared with the existing PCI planning schemes, our scheme have demonstrated significant optimization performance improvement as well as dramatic reduction on the computational complexity. Also, our scheme radically enhanced optimization efficiency by subtle and customized compute programming realization which could be further commercialized.

References

1. Song, W., Mei, S., Zhang, L.: Analysis of the actual influence of mode three interference in TD-LTE system
2. Ruan, M.: Relation matrix establishment system and method and PCI intelligent optimization system and method. CN 103906078 A (2014)
3. Hou, Y., Sui, Y.: PCI optimization method based on the cell correlation. Telecom Eng. Technics Stand. (2013)
4. GPPTS 36.214, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements
5. China Mobile: TD-LTE digital cell mobile communications network OMC-R measurement report technical specification
6. Chrost, L., Grochla, K.: Conservative graph coloring: a robust method for automatic PCI assignment in LTE. In: Computer Networks. Springer, Heidelberg (2013)
7. Telecommunication engineering technology and standardization, August 2013

Data Mining for Base Station Evaluation in LTE Cellular Systems

Lexi Xu¹(✉), Xueqing Zhao², Yanli Yu³, Yuting Luan⁴,
Xinzhou Cheng¹, Jie Gao¹, Jian Guan¹, and Kun Chao¹

¹ China Unicom Network Technology Research Institute, Beijing 100048, China
xulx29@chinaunicom.cn

² University of California, Davis, CA 95616, USA
xqz Zhao@ucdavis.edu

³ China Railway Eryuan Engineering Group, Chengdu 610031, China

⁴ China Railway Engineering Consulting Group, Beijing 100055, China

Abstract. Effective base station (BS) evaluation can assist telecom operators to find problematic cells and optimize the system performance. This paper proposes a data mining based joint BS evaluation (JBSE) algorithm in LTE cellular systems. Initially, the JBSE algorithm considers four key factors, including the cell energy consumption, the cell revenue, the cells distribution induced interference, the high BS induced cross-boundary coverage. Then, the expert judgement matrix is employed to rank the level of each factor. Finally, the JBSE algorithm evaluates each cell comprehensively. The JBSE algorithm is used in the LTE systems evaluation of a city in China. It can find problematic cells effectively.

Keywords: LTE · Base station · Energy consumption · Cell revenue

1 Introduction

Due to the high data rate and low latency, LTE cellular systems are widely deployed worldwide [1, 2]. The operation of LTE systems includes the planning stage and the optimization stage [3, 4]. Effective base station (BS) evaluation can assist telecom operators to find problematic cells and optimize systems, thus reducing the operational expenditure, as well as improving the system performance and user perception [5, 6].

Many existing works consider sole factor for BS evaluation. In [7], the stochastic geometric model is used to optimize the BS distribution, and the objective is to minimize the outage ratio. Authors of [8] design the relay base station deployment and optimization scheme, and the aim is to reach the balanced traffic distribution among cells.

In this paper, we design a data mining based joint BS evaluation (JBSE) algorithm in LTE cellular systems. Firstly, the JBSE algorithm considers a series of factors, including the cell energy consumption factor, the revenue factor, the dense cells distribution induced interference factor, the high BS induced cross-boundary coverage factor. Secondly, the JBSE algorithm utilizes the expert judgement matrix to rank the level of each factor. Thirdly, the JBSE algorithm evaluates each LTE cell synthetically.

2 Four Factors of JBSE Algorithm

2.1 Cell Energy Consumption

It is generally known that telecom operators spend a huge amount of maintenance cost for the cell energy consumption [9, 10]. The major sources of energy consumption include RF, BS equipment, refrigeration equipment. This paper denotes the energy consumption of $Cell_j$ RF as E_j^{RF} . The energy consumption of $Cell_j$ BS equipment is denoted as E_j^{BSeqp} . The energy consumption of $Cell_j$ refrigeration equipment is denoted as E_j^{REFeqp} . Hence, the energy consumption E_j of $Cell_j$ can be generally calculated as (1).

$$E_j = E_j^{RF} + E_j^{BSeqp} + E_j^{REFeqp} \quad (1)$$

2.2 Cell Revenue

From the perspective of telecom operators, both voice call and data service bring the revenue and profit. We should prefer to provide first-class maintenance and good QoS to the high revenue cell. Therefore, the cell revenue is a key factor for BS evaluation. In this paper, the $Cell_j$ revenue of voice call is denoted as Rev_j^{call} . The $Cell_j$ revenue of data service is denoted as Rev_j^{data} . Hence, the total revenue of $Cell_j$ is denoted as Rev_j , using (2a) and (2b).

$$Rev_j = Rev_j^{call} + Rev_j^{data} \quad (2a)$$

$$\Rightarrow Rev_j = Num_j^{call} \times Price_{unit}^{call} + Num_j^{data} \times Price_{unit}^{data} \quad (2b)$$

where Num_j^{call} is the total number of voice call in $Cell_j$, and the unit of Num_j^{call} is *minute*. $Price_{unit}^{call}$ is the unit price of voice call per *minute*. Therefore, $Rev_j^{call} = Num_j^{call} \times Price_{unit}^{call}$. Similarly, Num_j^{data} is the total flow of data service, and the unit of Num_j^{data} is *MB*. $Price_{unit}^{data}$ is the unit price of data service per *MB*. Hence, $Rev_j^{data} = Num_j^{data} \times Price_{unit}^{data}$.

2.3 Cells Distribution Induced Interference

In LTE systems, each cell employs hard handover [11, 12]. Therefore, user's signal received from adjacent cells becomes the inter-cell interference [13–15]. Hence, the cell density should be in a suitable range. Figure 1 illustrates the analysis method of the cell density, as well as the main coverage area. Specifically, the $Cell_j$'s main coverage area is under ± 47.5 degrees of $Cell_j$'s main lobe direction within R_{dens} meters. (e.g., R_{dens} can be set as 500–900 m in downtown area). Correspondingly, this paper defines the number of adjacent cells in $Cell_j$'s main coverage area as $Cell_j$ density.

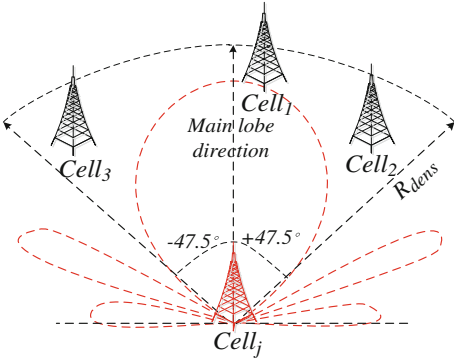


Fig. 1. $Cell_j$'s main coverage area illustration

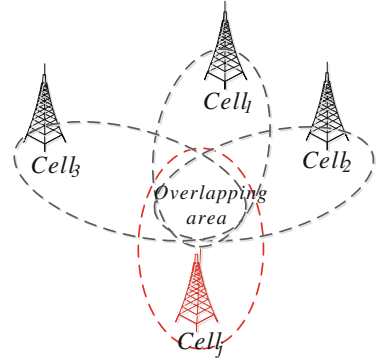


Fig. 2. Overlapping area illustration

In order to comprehensively evaluate BS deployment, we consider the interference induced by dense cells distribution. As shown in Fig. 2, dense adjacent cells will result in the overlapping area. Users in the overlapping area will receive strong signal power from adjacent cells and suffer severe inter-cell interference. Since the inter-cell interference is time-varying and difficult to calculate, this paper employs $Cell_j$ overlapping ratio to reflect the inter-cell interference induced by dense cells distribution. The process to calculate $Cell_j$ overlapping ratio $O_j^{overlap}$ includes four steps:

Step1: Measurement report (MR) of each user is collected by telecom operators [16]. This paper assumes $User_k$ has M measurement reports (MRs).

Step2: In $MR_{k,m}$ ($m \in \{1, 2, \dots, M\}$), for $User_k$ ($k \in \{1, 2, \dots, K\}$) served by $Cell_j$, the reference signal received power from serving $Cell_j$ is denoted as $RSRP_{j,k,m}$. $User_k$'s signal power received from adjacent $Cell_i$ ($i \in \{1, 2, \dots, I\}$) is denoted as $RSRP_{i,k,m}$.

Step3: For $User_k$, the total number of adjacent $Cell_i$ ($i \in \{1, 2, \dots, I\}$), which meets $RSRP_{i,k,m} - RSRP_{j,k,m} > -3dB$, is calculated as the overlapping value $OV_{k,m}$ of this $User_k$'s $MR_{k,m}$. Namely, $OV_{k,m} = Num\ of\ (RSRP_{i,k,m} - RSRP_{j,k,m} > -3dB, (i \in \{1, 2, \dots, I\}))$.

Step4: $Cell_j$ overlapping ratio $R_j^{overlap}$ equals the ratio of all MRs' total overlapping values to the total number of all $MR_{k,m}$, as shown in (3a) and (3b):

$$R_j^{overlap} = \frac{\sum_{k=1}^K \sum_{m=1}^M OV_{k,m}}{\sum_{k=1}^K \sum_{m=1}^M MR_{k,m}} \tag{3a}$$

$$\Rightarrow R_j^{overlap} = \frac{\sum_{k=1}^K \sum_{m=1}^M Num\ of\ (RSRP_{i,k,m} - RSRP_{j,k,m} > -3dB, (i \in \{1, 2, \dots, I\}))}{\sum_{k=1}^K \sum_{m=1}^M MR_{k,m}} \tag{3b}$$

2.4 High Cell Induced Cross-Boundary Coverage

Telecom operator should equip BS antenna with appropriate height. When the BS antenna is equipped on the skyscraper roof, the signal may transmit extremely long distance, thus resulting in the cross-boundary coverage, as shown in Fig. 3. Since the signal under long distance transmission is unstable and time-varying, the cross-boundary coverage impacts the SINR and degrades the network performance.

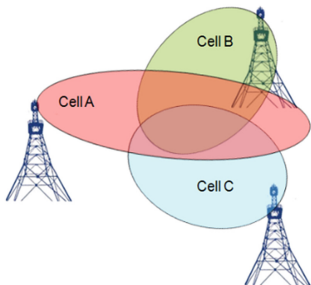


Fig. 3. Diagram of cross-boundary coverage

Table 1. MR.Tadv distance description

MR.Tadv	Distance (meter)
<i>MR.Tadv.00</i>	0–78.24
<i>MR.Tadv.01</i>	78.24–156.28
...	
<i>MR.Tadv.11</i>	860–938
<i>MR.Tadv.12</i>	938–1095
...	
<i>MR.Tadv.17</i>	1330–1408
...	
<i>MR.Tadv.43</i>	15022–20029
<i>MR.Tadv.44</i>	20029 to $+\infty$

According to 3GPP, MR (measurement report) contains Time-advance, denoted as *MR.Tadv*. *MR.Tadv* reflects the distance between user and serving cell, as shown in Table 1 [17]. For example, *MR.Tadv12* reflects the distance between $User_k$ and serving $Cell_j$ is in the range from 938 m to 1095 m. We employ two sub-factors for the cross-boundary coverage. The first sub-factor is ratio of User-to- $Cell_j$ distance larger than 938 m, denoted by $R_j^{dis > 938}$, as shown in (4). The second is ratio of User-to- $Cell_j$ distance larger than 1408 m, denoted by $R_j^{dis > 1408}$, as shown in (5).

$$R_j^{dis > 938} = \frac{\sum_{k=1}^K \sum_{m=1}^M MR.Tadv12_{j,k,m} + MR.Tadv13_{j,k,m} + \dots + MR.Tadv44_{j,k,m}}{\sum_{k=1}^K \sum_{m=1}^M MR.Tadv00_{j,k,m} + MR.Tadv01_{j,k,m} + \dots + MR.Tadv44_{j,k,m}} \quad (4)$$

$$R_j^{dis > 1048} = \frac{\sum_{k=1}^K \sum_{m=1}^M MR.Tadv18_{j,k,m} + MR.Tadv19_{j,k,m} + \dots + MR.Tadv44_{j,k,m}}{\sum_{k=1}^K \sum_{m=1}^M MR.Tadv00_{j,k,m} + MR.Tadv01_{j,k,m} + \dots + MR.Tadv44_{j,k,m}} \quad (5)$$

where $MR.Tadv_{j,k,m}$ is the Time-advance of serving $Cell_j$ reported by $User_k$ ($k \in \{1, 2, \dots, K\}$) in $MR_{k,m}$ ($m \in \{1, 2, \dots, M\}$).

3 Joint Evaluation

3.1 Design Expert Judgment Matrix

In Sect. 2, four factors are considered to evaluate each LTE cell. In order to effectively determine the threshold of each criterion, Sect. 3.1 constructs the *expert judgment matrix* B_{judge} , as shown in (6). B_{judge} is with the dimension of 4×2 . Each horizontal vector is for a specific factor.

$$B_{judge} = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \\ B_{3,1} & B_{3,2} \\ B_{4,1} & B_{4,2} \end{pmatrix} = \begin{pmatrix} E_{low} & E_{high} \\ Rev_{low} & Rev_{high} \\ R_{low}^{overlap} & R_{high}^{overlap} \\ R_{high}^{dis > 938} & R_{high}^{dis > 1048} \end{pmatrix} \quad (6)$$

Specifically, $B_{1,1}$ and $B_{1,2}$ are for the cell energy consumption factor. This paper sets $B_{1,1} = E_{low}$ and $B_{1,2} = E_{high}$. E_{low} denotes the threshold of low energy consumption. E_{high} denotes the threshold of high energy consumption.

$B_{2,1}$ and $B_{2,2}$ are for the cell revenue factor. We set $B_{2,1} = Rev_{low}$ and $B_{2,2} = Rev_{high}$. Rev_{low} and Rev_{high} are the low revenue threshold and high revenue threshold.

$B_{3,1}$ and $B_{3,2}$ are for the factor of cells distribution induced interference. JBSE algorithm sets $B_{3,1} = R_{low}^{overlap}$ and $B_{3,2} = R_{high}^{overlap}$. $R_{low}^{overlap}$ and $R_{high}^{overlap}$ are the threshold of low overlapping ratio and threshold of high overlapping ratio, respectively [16].

$B_{4,1}$ and $B_{4,2}$ are for the factor of high cell induced cross-boundary coverage. JBSE algorithm sets $B_{4,1} = R_{high}^{dis > 938}$ and $B_{4,2} = R_{high}^{dis > 1048}$. $R_{high}^{dis > 938}$ is the threshold of high ratio of User-to-Cell_j distance larger than 938 m. $R_{high}^{dis > 1048}$ is the threshold of high ratio of User-to-Cell_j distance larger than 1408 m.

3.2 Cell Evaluation

On the basis of *expert judgment matrix* B_{judge} , the JBSE algorithm evaluates each cell in LTE systems. As illustrated in Table 2, *Cell_j energy consumption*, which is denoted as E_j in (1), can be divided into three types, according to $B_{1,1}$, $B_{1,2}$, E_{low} and E_{high} .

Table 2. Judgement criterion of cell energy consumption

Cell _j energy consumption	Cell _j energy consumption level
$E_j \leq E_{low}$	<i>Low energy consumption</i>
$E_{low} < E_j \leq E_{high}$	<i>Medium energy consumption</i>
$E_{high} < E_j$	<i>High energy consumption</i>

As illustrated in Table 3, $Cell_j$ revenue (denoted as Rev_j in (2a) and (2b)) consists of three types, on the basis of $B_{2,1}$, $B_{2,2}$, Rev_{low} and Rev_{high} .

Table 3. Judgement criterion of cell revenue

Cell _j revenue	Cell _j revenue level
$Rev_j \leq Rev_{low}$	Low revenue
$Rev_{low} < Rev_j \leq Rev_{high}$	Medium revenue
$Rev_{high} < Rev_j$	High revenue

Similarly, based on $B_{3,1}$, $B_{3,2}$, $R_{low}^{overlap}$ and $R_{high}^{overlap}$, $Cell_j$ overlapping ratio (calculated as $R_j^{overlap}$ in (3a) and (3b)) can be categorized into three types, as shown in Table 4. In Table 4, *high overlapping* level reflects that users in $Cell_j$ has high probability to suffer severe inter-cell interference from adjacent cells.

Table 4. Judgement criterion of cell overlapping

Cell _j overlapping ratio	Cell _j overlapping level
$R_j^{overlap} \leq R_{low}^{overlap}$	Low overlapping
$R_{low}^{overlap} < R_j^{overlap} \leq R_{high}^{overlap}$	Medium overlapping
$R_{high}^{overlap} < R_j^{overlap}$	High overlapping

According to $B_{4,1}$, $B_{4,2}$, $R_{high}^{dis > 938}$ and $R_{high}^{dis > 1048}$, the cross-boundary coverage indicators, including $R_j^{dis > 938}$ in (4) and $R_j^{dis > 1048}$ in (5), are employed to evaluate $Cell_j$ cross-boundary coverage level.

$Cell_j$ cross – boundary coverage level

$$= \begin{cases} \text{Severe cross – boundary coverage} & \text{if } 30\% < R_j^{dis > 938} \text{ and } 10\% < R_j^{dis > 1048} \\ \text{Medium cross – boundary coverage} & \text{if } 30\% < R_j^{dis > 938} \text{ or } 10\% < R_j^{dis > 1048} \\ \text{Good coverage} & \text{if } 30\% \geq R_j^{dis > 938} \text{ and } 10\% \geq R_j^{dis > 1048} \end{cases} \tag{7}$$

From (7), $Cell_j$ will be ranked as *severe cross-boundary coverage* when both $30\% < R_{high}^{dis > 938}$ and $10\% < R_{high}^{dis > 1048}$ are meet. $Cell_j$ will be ranked as *medium cross-boundary coverage* when either $30\% < R_{high}^{dis > 938}$ or $10\% < R_{high}^{dis > 1048}$ is meet. Otherwise, $Cell_j$ will be ranked as *good coverage*.

After reaching above four levels, this paper takes the joint evaluation and labels each cell, as shown in Table 5.

Table 5. Joint evaluation of cell

Step1: If Cell_j energy consumption level is high, and Cell_j revenue level is low, Cell_j's energy label will be ranked as 'Energy optimization'
 * 'Energy optimization' indicates Cell_j's high energy consumption is not generated by large voice and data services. Telecom operators should check and optimize the BS equipment, refrigeration equipment, in this way to reduce the huge consumption

Step2: If Cell_j revenue level is high, Cell_j's revenue label will be ranked as 'High-value maintenance'
 * 'High-value maintenance' indicates Cell_j generates both large voice and data services, thus bringing huge revenue and profits. Operators should take preference to maintain this cell.

Step3: If Cell_j overlapping ratio level is high, Cell_j's interference label will be ranked as 'Structure optimization'
 * 'Structure optimization' indicates Cell_j is surrounded by dense adjacent cells, thus network structure optimization is needed. Typical methods include reducing the cell density, RF optimization of the adjacent cell which generates the heaviest interference.

Step4: If Cell_j cross-boundary coverage level is either severe or medium, Cell_j's cross-boundary label will be ranked as 'Coverage control'
 * 'Coverage control' indicates Cell_j has a large number of users under long-distance signal transmission. Typical methods include reducing antenna height, reducing transmission, antenna tilt optimization

4 Application and Analysis of JBSE Algorithm

In order to evaluate the JBSE algorithm, this paper employs the JBSE algorithm to the LTE systems of a city in China. Figure 4 illustrates the evaluation results. From Fig. 4, 311 cells are labelled as *Energy optimization*, 637 cells as labelled as *High-value maintenance*, 476 cells are labelled as *Structure optimization*, 213 cells are labelled as

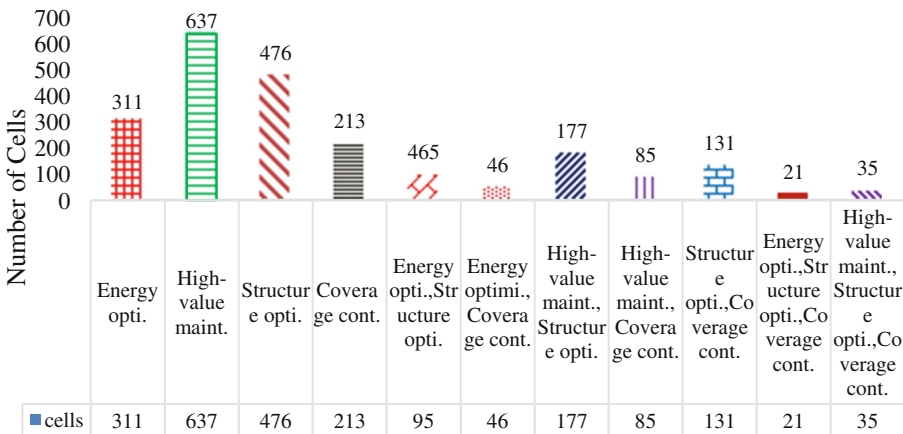


Fig. 4. Evaluation results of a city district LTE systems

Coverage control. The network optimization department can employ relevant methods to address these cells, as detailed introduced in Table 5.

Furthermore, a series of cells are double labelled. For example, 95 cells are double labelled as *Energy optimization & Structure optimization*. 46 cells are double labelled as *Energy optimization & Coverage control*. Telecom operators should prefer to tackle these cells than single-label cells. In addition, 21 cells are triple labelled as *Energy optimization & Structure optimization & Coverage control*. 35 cells are triple labelled as *High-value maintenance & Structure optimization & Coverage control*. These triple labelled cells have the highest priority of optimization and maintenance.

Above mentioned 2227 cells are checked by telecom operator. 1537 cells are confirmed with some problems or being high revenue cells. Therefore, the accuracy of the JBSE algorithm reaches 69%.

5 Conclusions

This paper studies base station (BS) evaluation and designs a data mining based joint BS evaluation (JBSE) algorithm in LTE cellular systems. JBSE algorithm considers the energy consumption, the revenue, the interference, the cross-boundary coverage. Based on above four factors, each LTE cell is evaluated and labelled comprehensively. We apply the JBSE algorithm to the LTE systems in a city in China, thus assisting the telecom operator to seek and optimize the problems of cells.

References

1. Xu, L., Cheng, X., Chen, Y., Chao, K., Liu, D., Xing, H.: Self-optimised coordinated traffic shifting scheme for LTE cellular systems. In: 1st EAI International Conference on Self-Organizing Networks, pp. 67–75. Springer Press, Beijing (2015)
2. Cao, Y., Wang, N., Sun, Z., Cruickshank, H.: A reliable and efficient encounter-based routing framework for delay/disruption tolerant networks. *IEEE Sens. J.* **15**(7), 4004–4018 (2015)
3. Xu, L., Luan, Y., Cheng, X., Fan, Y., Zhang, H., Wang, W., He, A.: Telecom big data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)
4. Cao, Y., Sun, Z., Wang, N., Riaz, M., Cruickshank, H., Liu, X.: Geographic-based spray-and-relay (GSaR): an efficient routing scheme for DTNs. *IEEE Trans. Veh. Technol.* **64**(4), 1548–1564 (2015)
5. Xu, L., Chen, Y., Gao, Y., Cuthbert, L.: A self-optimizing load balancing scheme for fixed relay cellular networks. In: 2th IET International Conference on Communication Technology and Application, pp. 306–311. IET Press, Beijing (2011)
6. Cao, Y., Wang, T., Kaiwartya, O., et al.: An EV charging management system concerning drivers' trip duration and mobility uncertainty. *IEEE Trans. Syst. Man Cybern. Syst.* **99**, 1–12 (2016)
7. Andrews, J.G., Baccelli, F., Ganti, R.K.: A tractable approach to coverage and rate in cellular network. *IEEE Trans. Commun.* **59**(11), 3122–3134 (2011)
8. Jiang, P., Bigham, J., Wu, J.: Self-organizing relay stations in relay based cellular networks. *Comput. Commun.* **31**(13), 2937–2945 (2008)

9. Xu, L., Chen, Y., Chai, K.K., Liu, D., Yang, S., Schormans, J.: User relay assisted traffic shifting in LTE-advanced systems. In: 77th IEEE Vehicular Technology Conference, pp. 1–6. IEEE Press, Dresden (2013)
10. Cao, Y., Sun, Z., et al.: Routing in delay/disruption tolerant networks: a taxonomy, survey and challenges. *IEEE Commun. Surv. Tutorials* **5**(12), 654–677 (2013)
11. Xu, L., Chen, Y.: Priority-based resource allocation to guarantee handover and mitigate interference for OFDMA systems. In: 20th IEEE International Symposium on Personal Indoor and Mobile Radio Communications, pp. 783–787. IEEE Press, Tokyo (2009)
12. Xu, L., Cheng, X., Liu, Y., et al.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: 16th IEEE International Conference on Communication Technology, pp. 806–812. IEEE Press, Hangzhou (2015)
13. Wang, W., Xu, L., Zhang, Y., Zhong, J.: A novel cell-level resource allocation scheme for OFDMA system. In: 1st International Conference on Communications and Mobile Computing, pp. 287–292. IEEE Press, Kunming (2009)
14. Fan, X., Lu, J., et al.: Analysis of key wireless factors affecting on VoLTE user-perceived performance based on measured data. *Mobile Commun.* **41**(14), 5–12 (2017)
15. Xu, L., Luan, Y., Cheng, X., et al.: WCDMA data based LTE site selection scheme in LTE deployment. In: 1st International Conference on Signal and Information Processing, Networking and Computers, pp. 249–260. CRC Press, Taylor & Francis Group, Beijing (2015)
16. Liu, Y., Xu, L., Chen, Y., et al.: A novel power control mechanism based on interference estimation in LTE cellular networks. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 397–401. IEEE Press, Qingdao (2016)
17. Han, Z., Kong, L., Chen, G., Li, F.: *LTE FDD Technology Principle and Network Planning*, 1st edn. China Post and Telecommunications Press, Beijing (2012)

Blind Video Quality Assessment Based on Human Visual Speed Perception and Nature Scene Statistic

Shiyu Zhou^{1(✉)}, Xiuyan Xia², Meng Ran¹, Luhan Wang²,
and Chen Cheng¹

¹ China Unicom Network Technology Research Institute, Beijing 10048, China
zhoushy@dimpt.com

² Beijing University of Posts and Telecommunications, Beijing 10048, China

Abstract. In this paper, we incorporate human visual speed perception model into a NSS based VQA method for video sequences transmitted via wireless network. The human visual speed perception contains two parts: one is motion information, which is calculated by using the prior probability distribution of the relative motion in each frame of the video; the other one is the perception noise, derived from the background motion. The weighting factors are defined as perceptual information that minus perception noise from motion information. We extract both spatial and temporal statistical features in videos (NVS-S and NVS-T), and measure their deviations from pristine statistical features. Consequently, the deviations can be synthesized with perceptual information based weighting coefficients to get the video quality score. The proposed blind VQA model is trained and tested in the LIVE database and EPFL-PoliMI database. The experimental results indicate that our model outperforms other blind VQAs.

Keywords: Blind video quality assessment · Motion information
Natural video statistics · Motion residual

1 Introduction

With the rapid proliferation of internet, video service has become part of people's life gradually. But due to lossy compression, packet loss, video qualities that customers perceived may be degraded. For this reason, video quality assessment (VQA) plays a more important role in video processing applications. Video services especially mobile videos being transmitted via wireless channel, and the demands of the real-time and encryption original video both make the original videos unavailable. Thus approach used without any reference to original videos, is deemed as the most promising way for assessing network video quality, which is known as NR VQA.

MOVIE index proposed by Bovik [1] uses the outputs of a Gabor filter family to calculate optical flow and then estimates motion. Another widely used method is based on structural information detection—3D structure tensor [2]. Combined with the structural similarity (SSIM), Z. Wang proposed a weighting model in [3]. However, previous researches using weighting factors are basically full or reduced reference VQA,

for that weighting factor should be given to a certain pixel or area and hence it is hard for no reference VQA since direct information in a pixel or small area is usually unavailable. For the reason that, to no reference VQA especially based on NSS, we can only gain some rough statistical information. Therefore, it is necessary to apply the weighted operation on the premise of frame segment.

In [4], Z. Wang proposed a human visual speed perception model to obtain motion information, which performed well. In this paper we incorporate this model into spatio-temporal NSS VQA algorithm, in which NSS features are weighted using this perceptual information model. And in the paper, we have developed a framework that utilizes a perceptual information content model based on motion vectors to realize video frame segment.

The remainder of the paper is organized as follows. We introduce the visual speed perception method as the weighting factors in Sect. 2. In Sect. 3, we extract spatial features from pixels information and we analyze the DCT coefficients derived from motion residual image to obtain temporal feature. In Sect. 4, we adopt frame quality using the weighted NSS features and then gain the final video score via temporal pooling scheme. Section 5 will describe the experiment and performance analysis in detail. Finally we give our conclusion in Sect. 6.

2 Spatio-Temporal Weighting Factors

We propose a no reference video quality assessment consider not only the statistical change of spatio-temporal features caused by external distortions, but also the following two aspects: biased judgements visual system gives to different degree motional stimulas and internal distortions generated by HVS’s own error-prone characteristic as a information communication channel. On the basis of previous theory, to the first aspect, a prior probability is used to describe the information content [5] and to the second, a likelihood function is established to simulate the HVS channel [5], as shown in Fig. 1.

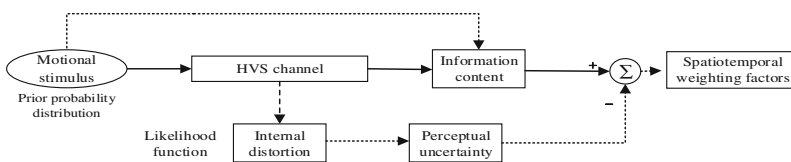


Fig. 1. Framework of spatio-temporal weighting factors calculation

In the video frame, motion information is usually represented as a two-dimensional field of motion vectors, where each spatial location (x, y) is associated with a motion vector $\vec{v}(x, y)$. In our method, adaptive-rood-pattern-search (ARPS) [6] has been utilized for motion estimation. For a video frame, three types of motion fields have been taken into account, which are absolute motion, background motion and relative motion. And they are represented by $\vec{V}_a, \vec{V}_b, \vec{V}_r$, respectively. Then, the length of the motion vector, denoted as $v = \|\vec{v}\|_2$, measures the speed of motion. The above is the preparation of computing these two indexes.

Statistically, an object in high speed movement relative to the background would cause a strong strike to our eyes. HVS would certainly pay more attention on those moving part. Therefore, it's intuitively believed that object motion always has a strong association with human visual attention. And it can provide a basis for predicting visual fixations. Here, we adopt a quantitative measurement of motion information content using the speed of motion. Previous work has shown that the distribution for the speed prior can be well fitted with:

$$p(v_r) = \frac{\varphi}{v_r^\partial} \quad (1)$$

And based on computing self-information of the information content, the information content contained in motion can be estimated:

$$I = -\log p(v_r) = \partial \log v_r + \beta \quad (2)$$

where φ , ∂ are two positive constants and $\beta = -\log \varphi$.

Since the perception of HVS can be regarded as an information communication channel (distortion channel), we could quantify the intuitive perceptual uncertainty via measuring these distortions. That is to say, given the likelihood function of internal distortion, the HVS perceptual uncertainty can be easily determined. According to [5], for a given background speed, the likelihood function is:

$$p(m|v_b) = \frac{1}{\sqrt{2\pi}\sigma m} \exp\left[-\frac{(\log m - \log v_b)^2}{2\sigma^2}\right] \quad (3)$$

and then perceptual uncertainty can be quantified by computing the entropy of the likelihood function:

$$\begin{aligned} U &= - \int_{-\infty}^{+\infty} p(m|v_b) \log p(m|v_b) dm \\ &= \log v_b - \gamma \log c + \delta \end{aligned} \quad (4)$$

Where m and v_b are the uncertainty measurement and the speed of background motion, respectively. σ is inversely dependent on the contrast c , modeled as $\sigma = \frac{\lambda}{c^\gamma}$ and $\delta = \frac{1}{2} + \frac{1}{2} \log(2\pi\sigma^2) + \log \lambda$ is a constant.

As we can see in Fig. 2, to every spatial location in each video frame, the information content and perceptual uncertainty can be calculated. On the basis of previous hypothesis on HVS that the importance of a visual stimulate should increase with information content but decrease with the perceptual uncertainty, authors in [4] defined the weighting factor in every pixel as:

$$w = I - U \quad (5)$$

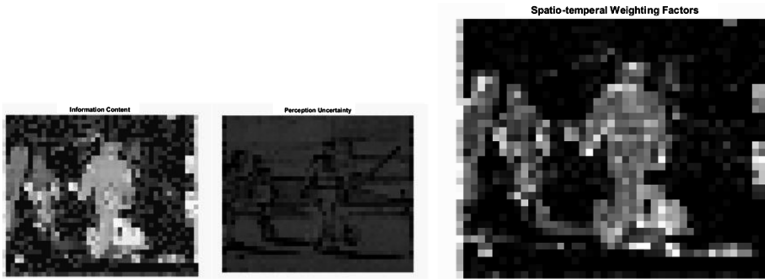


Fig. 2. Illustration of information content, perceptual uncertainty and the final weighting factors calculated from the motion vector field

In other words, every pixel has its own weight. Location with higher weight means more attention paid by HVS and consequence should be assigned more proportion in estimating quality score of the whole frame.

3 Spatio-Treporal Feature Abstraction

Here we divide each video frame into $n \times n$ areas and for per area, we abstract features from both spatial and temporal domain. To remove redundancy and capture the underlying features, we adopt a popular local non-linear transformation by removing mean displacements from pixels values and normalizing the local variance, which results in a more Gaussian-like histogram [7]. This operation was also used in many other blind VQAs, such as [8, 9]. For a given frame I_n in video sequences, $\Phi_n(i, j)$ is the pixel value locate at (i, j) , and the normalized pixel will be:

$$\psi_n(i, j) = \frac{\Phi_n(i, j) - \mu_n(i, j)}{\sigma_n(i, j) + C} \tag{6}$$

where $i \in \{1, 2, \dots, H\}$, $j \in \{1, 2, \dots, W\}$ are spatial indices. H and W represent the image height and width respectively. $C = 1$ is a constant that make sure the denominator of the formula won't tend to zero. $\mu_n(i, j)$ and $\sigma_n(i, j)$ represent the mean and variance of $N * N$ block around the pixel (i, j) .

For each area, a generalized Gaussian distribution (GGD) is adopted, which has been proved to have an excellent performance in previous works [8, 9]. And we can capture a broader spectrum of distorted frame statistics by using the GGD with zero mean:

$$f(x; \alpha, \beta, \gamma) = \alpha \exp(-(\beta|x|)^\gamma) \tag{7}$$

where $\beta = \frac{1}{\sigma} \sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}$, $\alpha = \frac{\beta\gamma}{2\Gamma(1/\gamma)}$, and $\Gamma(\bullet)$ is the gamma function: $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

Accordingly, the shape of the distribution is decided by the above three parameters α, β, γ . The existence of distortions will exactly alter this shape and thus these three parameters will change correspondingly. We use $n \times n$ groups of α, β, γ value as the first three parameters in NVS-S.

While during transmitted over the network, video sequences are often faced with motion distortion. Thus, temporal distortion can be regarded as motion distortion when referred to network video sequences. Here we use the motion residual to obtain the temporal features.

Similarly, we do the following estimation and transformation operations on $n \times n$ areas separately. Here we use the absolute motion vector field \vec{V}_a . Each spatial location will have a motion vector $\vec{V}_a(x,y)$, which represents the corresponding motion direction and motion strength between two adjacent frames. According to the n th frame and the motion vectors, we can get the predicted $(n+1)$ th frame:

$$I_{n+1}^p(i,j) = I_n(i + V_{a(x)}, j + V_{a(y)}) \quad (10)$$

So the motion residual image is:

$$I_{residual} = I_{n+1} - I_{n+1}^p \quad (11)$$

In pristine video, except for the zero coefficients, motion residual image only has a few small coefficient, and temporal distortion will introduce much energy to the motion residual image. In order to capture the characteristic, we get the DCT coefficients by performing a 16×16 DCT transform to the motion residual images. Likewise, we use a GGD to capture the distribution of the distorted image and also use the three distribution shape parameters α, β, γ as a representation of temporal NSS feature. Finally, we capture $n \times n$ groups of α, β, γ values in all to represent motion residual DCT coefficient distribution for the whole frame.

4 Pooling Scheme

We adopt Mahalanobis distance to calculate the difference between the characteristic parameters abstracted from the distorted and natural video frames. Contrast to Euclidean distance, Mahalanobis distance is scale-invariant and is more effective to exhibit the distance between different feature vectors even though the vector elements have different scope, which is expressed as:

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{(v_1 - v_2)^T \frac{(\Sigma_1 + \Sigma_2)^{-1}}{2} (v_1 - v_2)} \quad (14)$$

where v_1, v_2 are the mean vectors and Σ_1, Σ_2 are the covariance matrices of parameters abstracted from natural and distorted video frame, respectively.

Since the spatio-temporal weighting factors are counted via pixels, for area k , we

calculate the average value $w_{ave}(k) = \frac{\sum_{i=1}^{H_k} \sum_{j=1}^{W_k} w(i,j)}{H_k \times W_k}$, where H_k , W_k are the height and the width of the area k respectively, to get the corresponding weighting value for $D(k)$. Finally, quality score for a single frame can be represented as:

$$Q = \frac{\sum_{k=1}^{n \times n} w_{ave}(k) \bullet D(k)}{\sum_{k=1}^{n \times n} w_{ave}(k)} \quad (15)$$

Given the frame quality, we adopt Minkowski summation to integrate these video frames to yield the overall quality score.

5 Experiment and Performance

We used LIVE-database and the EPFL-PoliMI database to show that the performance of our algorithm, and we used multiple train-test groups to increase the reliability of experiment. In each group, the two databases were subdivided into training and test sets, and 80% of the videos of the EPFL-PoliMI database were used for training while the remaining 20% plus the whole videos in the LIVE-database for testing.

When calculating the Mahalanonis distance between the NVS features of the natural video frame and the distorted one, a key step is to extract the NVS parameters from the natural video. However, in fact, there doesn't exist a pristine video that is completely undistorted, thus we choose 12 reference videos in the EPFL-PoliMI database, and use the mean of their NVS features as the nature NVS features, thereby obtain the covariance matrices.

In our implementation, a 38 pixels/degree of viewing distance is assumed. And through test for many timed, we choose the following parameters: $\vartheta = 0.2$, $\beta = 0.09$, $\gamma = 2.5$, $\delta = 2.25$, For the choice of the segment size n , we tried 1×1 (without weighing factors), 2×2 , 3×3 , and 4×4 and the results show that the performance reach the best level at 3×3 , which is in accordance with our expectations. This is because if the size n is too large, the accuracy to abstract NSS features from a relative small area will decline and hence the final quality will deviate from the nomal even if completed with weighting scheme, while on the contrast, if n is too small, weighting scheme cannot show it's effect to the greatest extent. Therefore, we reach a compromise and set $n = 3$.

In order to analyze the overall performance of our method, we intend to demonstrate the contributions in two aspects. First, we compare the different performances when the abstracted features are used alone or combined with others to prove the contribution of each individual features to the overall prediction performance, as shows in Table 1. Second, We have also compared our algorithm with the classic algorithm NIQE [9, 11], Video BLIINDS [10, 12], and our own NSS method proposed without weighting scheme in Table 2.

Table 1. Spearman rank order correlation coefficients obtained when using different features

Feature	Spatial feature	Temporal feature	Combined all features
SROCC	0.563	0.600	0.702

Table 2. Performance compared with SSIM, NIQE, BRISQUE, Video BLIINDS

Algorithm	Spearman CC	Pearson CC
NIQE	0.552	0.433
Video BLIINDS	0.730	0.881
Proposed(without weighting scheme)	0.754	0.893
Proposed(weighting scheme)	0.788	0.906

6 Conclusion

In this paper, We have proposed an NR-VQA approach based on natural video statistics features extraction and weighting scheme related to HVS in both spatial and temporal domains. The new method choose information content and perceptual uncertainty these two indexes as the key weighting factors, which consider the impact both video content itself and HVS's characteristic have on human subjective perception. We also extract NVS features in total: normalized pixel distribution coefficient and residual DCT coefficient [13, 14]. To pooling scheme, we integrate the weighting coefficients into feature parameters in the corresponding area to obtain the quality for per frame and adopt hysteresis effect based temporal pooling. Having been trained and tested using two databases: LIVE and EPFL-PoliMI, the experimental result outperforms the state-of-art NR algorithms.

References

1. Seshadrinathan, K., Bovik, A.C.: Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans. Image Process.* **19**(2), 335–350 (2010)
2. Wang, Y., Jiang, T., Ma, S., Gao, W.: Novel spatio-temporal structural information based video quality metric. *IEEE Trans. Circuits Syst. Video Technol.* **22**(7), 989–998 (2012)
3. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment using structural distortion measurement. In: *Proceedings of the 2002 International Conference on Image Processing*, vol.3, pp. III-65–III-68 (2002)
4. Wang, Z., Li, Q.: Video quality assessment using a statistical model of human visual speed perception. *JOSA A* **24**(12), B61–B69 (2007)
5. Stocker, A.A., Simoncelli, E.P.: Noise characteristics and prior expectations in human visual speed perception. *Nature Neurosci.* **9**, 578–585 (2006)
6. Nie, Y., Ma, K.-K.: Adaptive rood pattern search for fast block-matching motion estimation. *IEEE Trans. Image Process.* **11**(12), 1442–1449 (2002)
7. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. In: *Sensory Communication*, pp. 217–234 (1961)

8. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **21**(12), 4695–4708 (2012)
9. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. *IEEE Signal Process. Lett.* **20**, 209–212 (2013)
10. Saad, M.A., Bovik, A.C., Charrier, C.: Blind prediction of natural video quality. *IEEE Trans. Image Process.* **23**, 1352–1365 (2014)
11. Ge, C., Wang, N., Skillman, S.: QoE driven dash video caching and adaptation at 5G mobile edge. In: *ACM ICN IC5G workshop, Kyoto* (2017)
12. Zhang, X., Wang, N., Cao, Y.: A stochastic analytical modelling framework on ISP-P2P collaborations in multi-domain environments. *IEEE Syst. J.* **PP**(99), 1 (2017)
13. Xu, L., Luan, Y., Cheng, X., Fan, Y., Zhang, H., Wang, W., He, A.: Telecom big data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)
14. Xu, L., Cheng, X., Liu, Y., Chen, W., Luan, Y., Chao, K., Yuan, M., Xu, B.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: *16th IEEE International Conference on Communication Technology*, pp. 806–812. IEEE press, Hangzhou (2015)

A Downlink Coverage Self-optimizing Algorithm for LTE Cellular Networks Based on Big Data Analytics

Jie Gao^(✉), Xinzhou Cheng, Lexi Xu, Lijuan Cao, and Chen Cheng

Department of Network Optimization and Management,
China Unicom Network Technology Research Institute,
Beijing 100048, People's Republic of China
{gaojie49, chengxz11, xulx29, caolj68,
chengc40}@chinaunicom.cn

Abstract. Inappropriate parameters and antenna problems will result in abnormal coverage performance of LTE networks. In order to deal with the problems mentioned above and improve the coverage performance, this paper proposes a downlink coverage Self-optimizing algorithm (DCSA) on the strength of big data analytics. The proposed algorithm obtains and analyzes the data, which records the performance situation of existing wireless networks to locate the cells with abnormal coverage performance. Then a Self-optimizing method is proposed to improve the coverage by adjusting the parameters. In the last part, the analysis results will display that the coverage optimization algorithm is high-efficiency and low-cost for telecom operators.

Keywords: Big data · Downlink coverage · Data mining · Self-optimizing

1 Introduction

Networks operation becomes complex gradually with the rapid development of smart phones. In addition, the telecom industry becomes competitive extremely. Therefore, the telecom operators have to provide wireless networks with high quality performance and good awareness of service. Meanwhile, network operators expend a huge amount of operating expense [1] to take the maintenance and optimization in daily work, as well as keep the stability of network.

With the rapid growth of users' requirements, the services become diversity, including voice call, video traffic etc. Network operators will collect the network performance measurements of wireless network by the help of operation support system (OSS) in huge amount. These real-time data could be utilized to optimize the performance of wireless network. To achieve that, big data technology can realize automatic and high efficiency solutions to optimize the wireless downlink coverage performance by process the OSS data [2]. In this paper, a downlink coverage Self-optimizing algorithm (DCSA) is proposed on the strength of big data analytics. The objective of the algorithm is to optimize and improve the coverage performance for LTE cellular networks.

The latest researches on the topic of coverage optimization usually summarize the problems into three aspects, such as cross-boundary coverage, coverage blind spots, and weak coverage. The coverage blind spots and cross-boundary coverage problems usually be caused by inappropriate settings of transmission power or downtilt of transmitting terminal in eNodeB. In [3], the transmission power adjustment is designed for the coverage optimization. However, the transmission power has the adjustment range, and the maximum transmission power in LTE network is 46 dBm. When the transmission power reaches adjustment limitation, the coverage problem cannot be solved effectively just by adjusting the transmission power. In addition, antenna downtilt adjustment is also a method to control the cell coverage.

This paper is organized as follows: Sect. 2 declares the related works of coverage optimization. Section 3 describes the process of DCSA. Section 4 displays the analysis results of the proposed effective algorithm by analyzing the OSS data. The last section shows the conclusions of this paper.

2 Related Work

In order to solve the coverage problems and improve the performance, many approaches have been proposed and used in earlier times. The strategy which is widely used is the drive test, but it will take amounts of resources during the process. Also, there are several automated strategies. Such as the authors in [4] and [5] that propose optimization schemes without human intervene, but these strategies don't have the ability to adapt to the complex and variable structure of networks.

The typical methods of data mining in big data technology include clustering method, rough sets, genetic algorithm, decision tree, neural network and so on. The clustering method is a widely used scheme, which divides the data into several clusters based on the similarity of data.

In this paper, the proposed DCSA is based on the k-means technology [6] which can enhance the efficiency of the algorithm. The whole optimization procedure is accomplished without manual intervene.

3 A Self-optimizing Scheme Based on K-means

K-means technology is utilized in DCSA. Figure 1 shows the procedure of DCSA which can be divided into four steps: data collection, intermediate calculation, K-means analysis and coverage optimization.

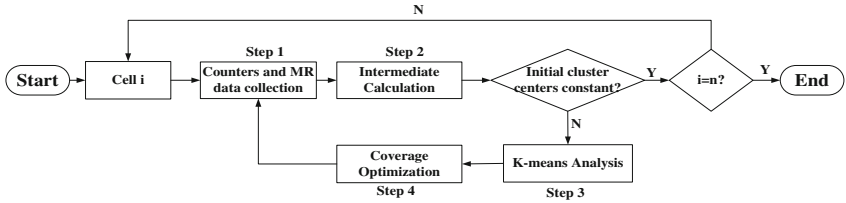


Fig. 1. The flowchart of DCSA

3.1 Step 1: OSS Data Collection

In the process of step 1, the OSS collector will gather measurement report (MR) and engineering parameter periodically. Tables 1, 2, 3 and 4 show the data used, i.e., engineering parameters, measurement report (MR), traffic data and performance data. The engineering parameters include some static network settings, such as positional information which is described by longitude and latitude. The MR includes downlink channel information, such as RSRP of serving cells and monitoring cells respectively. The traffic counters record abnormal events happened to users during the signaling and downloading procedure. The handover counters record the handover relationship between a pair of cellular cells, and include the times of successful and unsuccessful handover events between the two cells.

Table 1. Engineering parameters

eNodeB ID	Cell ID	Longitude	Latitude	N-Cell1 ID	N-Cell2 ID	N-Cell3 ID
10125	101251	23.41	116.58	101241	101242	101243

Table 2. Measurement report

eNodeB ID	Cell ID	N-eNodeB ID	N-Cell ID	RSRP	N-RSRP
10125	101251	10124	101241	-85 dBm	-95 dBm

Table 3. Traffic data

eNodeB ID	Cell ID	Time	Abnormal RAB release	Normal RAB release
10125	101251	8:00:00–8:30:00	59	57

Table 4. Performance data

eNodeB ID	Cell ID	N-eNodeB ID	N-Cell ID	Handover request	Handover success
10125	101251	10124	101241	51	49

3.2 Step 2: Intermediate Data Calculation

In this step, several intermediate results need to be ready for the k-means analysis in the next. As mentioned, the downlink coverage problems appear as two aspects: cross-boundary coverage and low coverage validity.

A. Distance Between Two Cells (DBTC)

Before optimization, the distance between each cell is needed. In this algorithm, we make the static distance as the distance between two cells. It's well known that the spherical distance formula is usually to be used to calculate the distance between two things on the earth surface. The formula is described as followed:

$$\text{DBTC} = 1000 * 6371.004 * \arccos(\cos(\pi/2 - \text{lat}_S * \pi/180) * \cos(\pi/2 - \text{lat}_N * \pi/180) + \cos(\text{lon}_S * \pi/180 - \text{lon}_N * \pi/180) * \sin(\pi/2 - \text{lat}_S * \pi/180) * \sin(\pi/2 - \text{lat}_N * \pi/180)) \quad (2)$$

where lat_S , lon_S , lat_N and lon_N represent the latitude and longitude of serving cell and neighbor cell respectively.

B. Downlink Coverage Validity (DCV)

In order to estimate the coverage performance of the serving cell for a user, we define an index downlink coverage validity to describe the coverage dominance in the area of responsibility. Then we can evaluate and order the influence degree from different neighboring cells by this consequence. Before calculation, we define a metric closeness ($C_{ij} = \{0, 1\}$) as follows, to determine whether a signal from serving cell is an effective coverage point.

$$C_{ijn} = \begin{cases} 1, & \text{RSRP}_{in} - \text{RSRP}_{jn} \geq D_{th} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where RSRP_{in} and RSRP_{jn} are the received power of serving cell and neighboring cell respectively, which will be recorded in MR data as presented in Table 3. D_{th} is the threshold for judging whether the interference from neighbor cell is strong enough to affect the downlink channel quality of the serving cell. Usually, D_{th} can be set as 3 dB. $C_{ijn} = 1$ means the serving cell has strong coverage dominance in the measurement area, and $C_{ijn} = 0$ means the interference from neighboring cell will have influence on the downlink system capacity of serving cell.

After calculation of C_{ijn} for each cell of the cellular networks. DCV can be calculated by the following formula based on the last step.

$$\text{DCV}_{ij} = \frac{\sum_{n=1}^{\text{count}(\{MR_{ij}\})} C_{ijn}}{\text{count}(\{MR_{ij}\})} \quad (4)$$

where $\{MR_{ij}\}$ is the set of measurement report which 'Cell ID' is i and 'N-Cell ID' is j . Function $\text{count}()$ will return the total number of a set. DCV_{ij} represents the degree of interference from neighboring cell j that cell i bears. $\text{DCV}_{ij} = 0$ means the downlink

channel quality of cell i is not affected by cell j . We need focus on the cells with low DCV and settle high priority in the optimization step.

After calculation of DCV_{ij} for each cell in pair, a matrix of someone serving cell can be get and described by the following table (Table 5):

Table 5. DCV matrix of serving cell

eNodeB ID	Cell ID	N-eNodeB ID	N-Cell ID	Handover success	Distance	DCV
10125	101251	10124	101241	87	230	98%
10125	101251	10123	101231	64	160	95%
10125	101251	10106	101062	52	65	15%
10125	101251	10108	101083	1	780	10%

where the data is in the reverse order by the index ‘handover success’, and the matrix meets ‘handover success’ $\neq 0$.

C. Call Drop Rate (CDR)

In the DCSA, network performance and service awareness are combined together to represent the quality of the downlink wireless channel. It is well known that call drop is the most effective index to reflect the channel quality which could be higher in the high interference environment. The call drop rate can be described as (5):

$$CDR = \frac{abnor_rab_rel}{abnor_rab_rel + nor_rab_rel} \times 100\% \tag{5}$$

where $abnor_rab_rel$ and nor_rab_rel are the number of normal and abnormal RAB release in a target cell as described in Table 4.

D. Initial Center and K

Initial centers and K is needed in the k-means analysis. Because of that, we assume cell distance and coverage performance as two aspects to duplicate the downlink coverage performance. Then we set the mathematical model of cluster samples as $M_n (L_n, DCV_n)$, where DCV_n can reflect the downlink coverage performance and L_n is the distance between the target cell and its neighboring cell. We define an index P to represent the different efficiency levels of DCSA which will influence the centers for each cluster and set 3 as default value for K . $C1$ is represented the cross-boundary coverage cluster, and $C3$ is represented low coverage validity and $C2$ is represented the normal coverage. Figure 2 shows the distribution of cluster samples after analysis.

$$\begin{cases} C_1 & initialize : \{ \overline{L} \times (1 + P), \overline{DCV} \times (1 - P) \} \\ C_2 & initialize : \{ \overline{L}, \overline{DCV} \times (1 + P) \} \\ C_3 & initialize : \{ \overline{L} \times (1 - P), \overline{DCV} \times (1 - P) \} \end{cases} \tag{6}$$

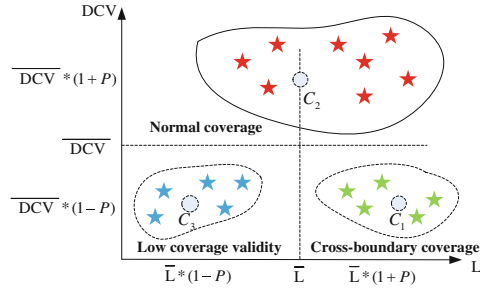


Fig. 2. The structure of k-means analysis

3.3 Step 3: K-means Analysis

In this step, k-means algorithm is used to generate the three cluster sets which represent different performance of downlink coverage. The algorithm is described as follows (Table 6):

$$d(M_n, C_k) = \sqrt{(M_{nx} - C_{kx})^2 + (M_{ny} - C_{ky})^2} \tag{7}$$

Table 6. K-means analysis of DCSA

INPUT: $K=3$, n data samples.

OUTPUT: three cluster sets.

1. Determine centers for the three clusters as (6);
2. Assign each data sample to the nearest cluster set according to the minimum distance principle by Euclidean distance model as (7);
3. Setting the mean of the cluster samples as the new center for each cluster;
4. Set i from 1 to n , circle 2 and 3 until each cluster center converge to constant.

where M_n is a sample, and C_k is the initial center of the k th cluster.

3.4 Step 4: Coverage Optimization

As described above, we set the impact factors of coverage problems as ε and ω , which meet the constraint of $\varepsilon + \omega = 1$. The adaptive downtilt adjustment (ADA) is defined to implement optimization. For each cluster respectively, ADA is described as follows:

$$\begin{cases} \theta_{j(t+1)} = \theta_{jt} \times (1 + \varepsilon) & M_n \in C_1 \\ \theta_{t+1} = \theta_t & M_n \in C_2 \\ \theta_{i(t+1)} = \theta_{it} \times \omega & M_n \in C_3 \end{cases} \tag{8}$$

where M_n is a data sample, $\theta_{i(t+1)}$ and $\theta_{j(t+1)}$ are the downtilts of serving cell and neighboring cell respectively after optimization.

4 Numerical Result

In this paper, a Self-optimizing algorithm is proposed, and the performance is verified by data analysis. The data processing is finished by computer software automatically. As the results in the following, the probability distribution of RSRP, SINR and call drop rate are compared before and after DCSA. The results can obviously prove that the downlink coverage Self-optimizing algorithm really has positive effects on the coverage quality for LTE networks.

The distribution range of RSRP and SINR are illustrated in Fig. 3 respectively, and the call drop rate comparison result is illustrated in Fig. 4.

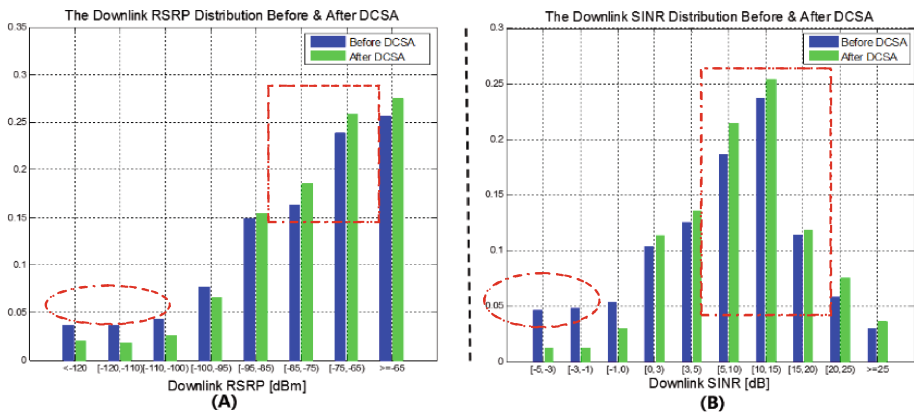


Fig. 3. The comparison of downlink coverage performance before and after DCSA

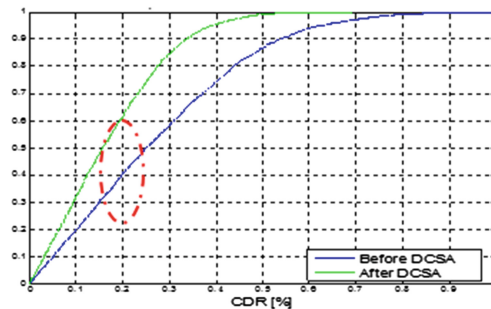


Fig. 4. The CDR distribution before and after DCSA

As illustrated in Fig. 3, the distribution of RSRP (A) and SINR (B) have more than 10% below -100 dBm and 0 dB (circle) and doesn't meet the condition of reference value for coverage performance before DCSA. The downlink wireless signal will be instable in this environment and call drop rate will higher with a high probability.

After DCSA, only 1.16% samples of MR data are below-100 dBm. There are nearly 65% of the MR samples are above -85 dBm (rectangle) and nearly half of the SINR samples are above 10 dB (rectangle). It is obviously that there is a significant improvement of the coverage environment after DCSA.

Beside of that, the service awareness of the users also get an experience improvement as the probability to meet a call drop below 20% increases from 40% to 60% and has 20% improvement. To achieve this result, the proposed DCSA not only takes coverage but also key performance into consideration during the procedure. The numerical results can obviously demonstrate that the DCSA has the ability to deal with the downlink coverage problems in LTE networks.

5 Conclusions

In this paper, in order to deal with the coverage problems, a downlink coverage Self-optimizing algorithm on the strength of big data analytic is proposed. The proposed algorithm can improve the downlink coverage performance by antenna parameter adjustment without human intervention. Also, DCSA considers both coverage performance and service awareness as analytical factors. It is more reasonable and comprehensive than the algorithms [3–5] which consider network performance as the only factor. The numerical results also can easily provide the consequence of DCSA is significant for promotion of coverage performance for LTE networks.

References

1. Xu, L., Luan, Y., Cheng, X.: Telecom big data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Sys. Technol.* **8**, 27–46. IEEE Beijing (2017)
2. Celebi, O.F., Zeydan, E., Kurt, O.F.: On use of big data for enhancing network coverage analysis. In: *International Conference on Telecommunications*, pp. 1–5. IEEE Casablanca (2013)
3. Shu, T., Krunz, M.: Coverage-time optimization for clustered wireless sensor networks: a power-balancing approach. *ACM Trans. Netw.* **18**, 202–215 (2012)
4. Fagen, D., Vicharelli, P.A., Weitzen, J.: Automated wireless coverage optimization with controlled overlap. *Trans. Veh. Technol.* **57**, 2395–2403 (2008)
5. Siomina, I., Varbrand, P., Yuan, D.: Automated optimization of service coverage and base station antenna configuration in UMTS networks. *Wireless Commun.* **13**, 16–25 (2006)
6. Gao, J., Cheng, X., Xu, L.: An interference management algorithm using big data analytics in LTE cellular networks. In: *ISCIT*, pp. 246–251. IEEE Qingdao (2016)
7. Gao, J., Cheng, X., Xu, L.: A coverage Self-optimizing algorithm using big data analytics in WCDMA cellular networks. In: *ICSINC*, pp. 27–46. Taylor, Beijing (2015)
8. Xu, L., Chen, Y., Gao, Y.: A Self-optimizing load balancing scheme for fixed relay cellular networks. In: *ICCTA*, pp. 306–311. IET, Beijing (2011)

Research on Wireless Network Planning of Railway TD-LTE System

Kai Yu^(✉), Yanli Yu, and Jie Xiong

China Railway Eryuan Engineering Group Co.Ltd, Chengdu 610031, China
ekyukai@gmail.com, 1716540752@qq.com,
xiongjie228@hotmail.com

Abstract. Wireless network planning is a key component for TD-LTE system deployment. In this paper, we investigate the wireless network planning method for a railway TD-LTE system. This evaluation is based on an actual TD-LTE system deployment for Shuohuang railway and includes frequency planning, network capacity prediction, wireless network coverage planning and cell planning. The corresponding Reference Signal Received Power (RSRP), Signal to Interference plus Noise Ratio (SINR) and throughput of the Railway TD-LTE System are further evaluated by Atoll network planning software, to validate the rationality for wireless network planning in such scenarios.

Keywords: Railway · TD-LTE · Wireless network planning
Wireless network coverage

1 Introduction

As a 4G mobile communication technology, TD-LTE has the advantages of lower transmission delay, higher transmission rate, larger capacity and optimized network deployment [1]. It has been widely used in public mobile communication, as well as in railway [2], electricity smart grid [3] and other domains such as private communication networks. TD-LTE technology has been employed in Shuohuang heavy haul railway for the first time for functions such as Locomotive Synchronic Operation and Control. The International Union of Rail ways (UIC) and China Railway have advocated railway wireless communication system being evolved from GSM-R to LTE. As a result, more railway private wireless communication networks will adopt LTE.

Network planning is a key component in LTE network deployment. The aim of network planning is to rationally set locations of base stations and LTE network parameters, in order to improve the quality of network service and ensure that the to-be-built network provides specified system capacity and coverage.

Based on conditions of technical characteristics of wireless access network, RF requirements and wireless transmission environment, the primary task of wireless network planning is to optimize locations and parameter settings of base stations and system-wide parameters, so as to meet the requirements of network coverage, capacity and quality [4].

In this paper, the wireless network planning method of railway TD-LTE system is investigated via the Atoll network planning software, based on TD-LTE system

deployment for Shuohuang railway. Due to technical characteristics of TD-LTE wireless network, the network planning of railway TD-LTE system mainly includes procedures of frequency planning, network capacity prediction, wireless network coverage planning such as link budget and site planning, cell planning.

2 Frequency Planning

A TD-LTE system can flexibly utilize frequency according to available frequency resources and network capacity. The total bandwidth for TD-LTE network of Shuohuang railway is 10 MHz, from 1785 to 1795 MHz. Using co-site double layer networks (network A and network B) to construct the network, we divide the 10 MHz into two equal bands of 5 MHz. Networks A and B utilize 1785~1790 MHz (F1) and 1790~1795 MHz (F2), respectively. The coverage areas of Networks A and B are essentially identical (Fig. 1).

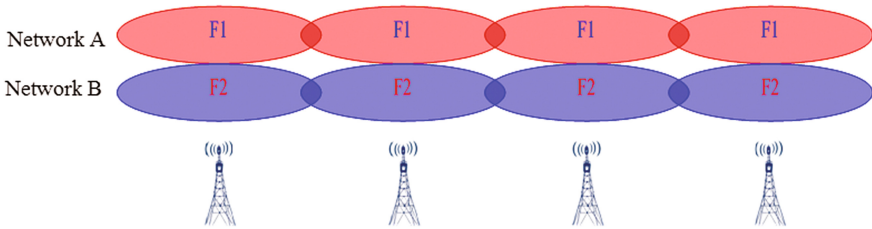


Fig. 1. Frequency planning and filed strength coverage

3 Network Capacity Prediction

The TD-LTE network of Shuohuang railway carries four types of business [5] (Table 1):

Table 1. The business on TD-LTE network of Shuohuang railway

No.	Business	Description	Carrying network
1	Real-time safety data	Locomotive Synchronic Operation and Control, Train tail	Network A + network B (together)
2	Non real-time safety data	Train number check information, Dispatching Command Information	Network A (primary) + network B (backup)
3	Voice	Cab integrated radio communication (CIR)	Network A (primary) + network B (backup)
		voice communication over handheld terminal	Network A (backup) + network B (primary)
4	Image	Video surveillance	Locomotive terminal A: network A Locomotive terminal B: network B

3.1 Cell Capacity Prediction

Cell capacity prediction is based on the following principles:

- (1) Assuming the train headway is 10 min and the train speed is 80 km/h, we set the minimum distance between two trains in the same direction as 13 km. As the radius of cell coverage is 4 km, there are at most two simultaneously passing trains in a base-station's field strength coverage area serving a double-line railway.
- (2) Each train consists of four locomotives and one train tail.

The TD-LTE system of Shuohuang railway is an uplink (UL) capacity limited system. Capacity prediction is based on uplink capacity calculation. Using bandwidth requirements of supported businesses, the calculated capacity of each cell is 1 Mbps.

3.2 Capacity Prediction on the Cell Edge

The capacity prediction on the cell edge is based on the following principles:

- (1) There are at most two locomotives for each heavy haul combined train on a cell edge simultaneously. Therefore, for a double-line railway, there can be at most four locomotives on a cell edge.
- (2) Assume two mobile users with handheld terminals on the cell edge.
- (3) Don't consider video surveillance business on the cell edge.

Based on above principles, the calculated uplink capacity on the cell edge is 450 kbps.

4 Wireless Network Coverage Planning

4.1 Planning Procedure

To design a TD-LTE wireless network, the first step is using link budget analysis to calculate the required number of base stations to cover the specified area after satisfying certain requirements such as uplink and downlink (DL) speeds.

Firstly, the maximum allowable path loss between base station and mobile station is calculated. Afterwards, the average cell radius and coverage are determined by propagation model and terrain classes on the basis of path loss calculation.

The required minimal number of base station to satisfy coverage requirement is calculated using cell radius of each terrain class. We can further determine the sites of base stations.

4.2 Link Budget

The maximum uplink path loss PL_{UL} is:

$$PL_{UL} = P_{out-UE} + G_{a-BS} + G_{a-UE} - L_{f-BS} - M_f - M_I - L_p - L_b - S_{-Bs} \quad (1)$$

where P_{out-UE} is the maximum transmission power of mobile station, L_f_{BS} is feeder loss, Ga_{BS} is antenna gain of base station, Ga_{UE} is antenna gain of mobile station, M_f is shadow fading margin, M_I is interference margin, L_p is building penetration loss, L_b is body loss, and S_{BS} is base station receiver sensitivity.

The maximum downlink path loss PL_{DL} equals:

$$PL_{DL} = P_{out-BT} - L_f_{BS} + Ga_{BS} + Ga_{UE} - M_f - M_I - L_p - L_b - S_{UE} \quad (2)$$

where P_{out-BT} the maximum transmitted power of service channel of base station, and S_{UE} is mobile station receiver sensitivity.

For Shuohuang railway TD-LTE project, we determined network parameters based on our own experience of both the system and the project. P_{out-BT} and P_{out-UE} are 43 dBm and 23 dBm respectively. If the required capacity on the cell edge should reach 450 kbps, the calculated maximal allowable path loss by link budget formula is 125.97 dB.

4.3 Distance Between Base Stations

Open areas

In open areas, the wireless network coverage is supported by antennas. Using the Cost231-hata model, we obtained the distance values between base stations under various coverage situations as Table 2 illustrates.

Table 2. Distance between base stations in open area

Base station antenna height	m	25	30	35	40
Mobile station antenna height	m	4	4	4	4
Path loss	dB	125.9	125.9	125.9	125.9
Distance between base stations					
City	km	1.67	1.25	1.32	1.39
Suburb	km	2.31	2.49	2.66	2.81
Countryside	km	3.68	3.99	4.28	4.55

Tunnels

In tunnels, the wireless network coverage is supported by leaky coaxial cable (LCX). With the maximal allowable path loss of 129.6 dB, we obtained the distance value between base stations in tunnels as Table 3 illustrates.

Table 3. Distance between base stations in tunnels

Path loss	dBm	125.97
Transmission loss of LCX	dB/100 m	3.9
Coupling loss(95%, 2 m)	dB	68
Design margin	dB	10
Additional loss	dB	9
Width factor: 10lg (X/2)	dB	6.02
Length of LCX	m	844
Distance between base stations	km	1.6

Based on the calculation in Table 3 and given the design margin and related device loss, the distance between base stations should be 1.6 km.

4.4 Site Design of Base Station

We calculated cell radius in open areas and tunnels based on the above link budget results. This calculation was then used to guide site design of base station combing with site inspection.

The following graph shows sites of base stations. Red dots indicate base station transmitters. Sectors indicate the sites of transmitting antennas. Sectors in the same color indicate different antennas of a single transmitter.

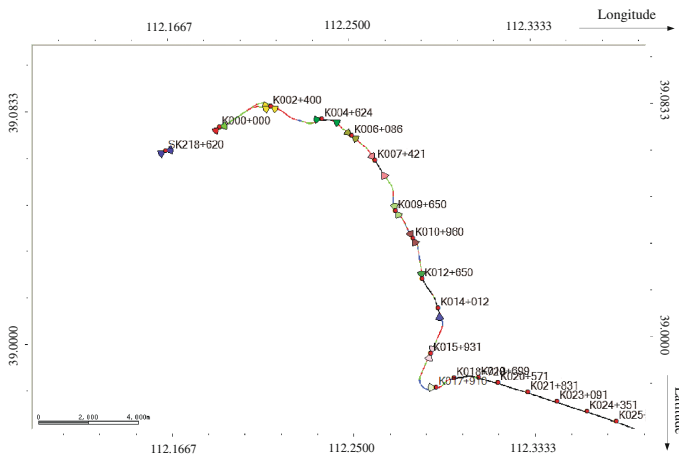


Fig. 2. Sites of transmitting antennas (Color figure online)

5 Cell Planning

5.1 Neighboring Cells Planning

For Shuohuang railway TD-LTE project, the main principles of neighboring cells planning are as follows:

- (1) In the same layer network (primary network or backup network), it is assigned neighboring cell relationship with both 2 cells before and after it.
- (2) The primary and backup cells of the same site are assigned mutual neighboring cell relationship.
- (3) For the primary cell at one site, the backup cells for the previous and latter sites are assigned neighboring cell relationship.
- (4) For the backup cell at one site, the primary cells for the previous and latter sites are assigned neighboring cell relationship.
- (5) It requires a mutual neighboring cell relationship, i.e. cell B is the neighboring cell of cell A, vice versa.

5.2 Physical Cell Identifier (PCI) Planning

PCI of LTE network should distinguish wireless signals from different cells. PCI planning assures absence of identical PCI in related cell coverage. The LTE protocol specifies physical layer CELL ID being divided by two, i.e. cell group ID and intra-group ID. The latest protocol specifies 168 physical layer cell groups, each composed of 3 ID groups, thus making $168 \times 3 = 504$ (0~503) independent Cell IDs.

- (1) Neighboring cells should be assigned different PCI.
- (2) The PCI should be unique among neighboring cells in two layer network.
- (3) The values of inter-cell PCI mod 3 should be different, to reduce the interference between pilot symbols.

6 Network and Performance Simulation

The system performance, such as RSRP, SINR and throughput, of the TD-LTE system of Shuohuang railway is simulated by Atoll network planning software. P_{out-BS} and P_{out-UE} are 43 dBm and 23 dBm respectively. The sites of base stations are set as Fig. 2 and the parameters of some base stations are as follows (Table 4).

Table 4. Parameters of some base stations

No.	Antenna num. of each BS	Sites	Longitude	Latitude	Height	Antenna azimuth (°)	Downtilt (°)
1	2	SK218 + 620	112.16509	39.06856	27	240	4
2						80	6
3	2	K000 + 000	112.19019	39.07696	27	225	6
4						75	6
5	3	K002 + 400	112.21362	39.08398	23	275	4
6						240	6
7						117	6
8	1	K004 + 624	112.23722	39.07931	18	261	2

6.1 Recommended Index of Coverage Performance

Based on link budget and related experience, we recommend the following main index of coverage performance of Shuohuang railway:

- (1) When pilot channel RSRP on the cell edge > -105 dBm, probability > 95%.
- (2) When neighboring cell has no load, SINR > 0 dB, probability > 95%.
- (3) When neighboring cell has no load, UL throughput > 1 Mbps, probability > 95%.

6.2 Performance Simulation

According to the simulation results of DL RSRP as shown in Fig. 3, DL RSRP can meet the coverage requirements, and the coverage probability of DL RSRP > -105 dBm can reach above 99%.

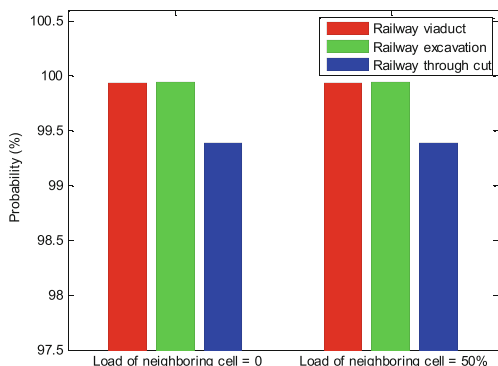


Fig. 3. The coverage probability of DL RSRP > -105dBm

Figure 4 illustrates simulation results of PUSCH SINR. In the intra-frequency network, the interference on the cell edge is obvious when the neighboring cell is loaded. The probability of PUSCH SINR > 0 dB can reach above 98%, fulfilling the index requirement.

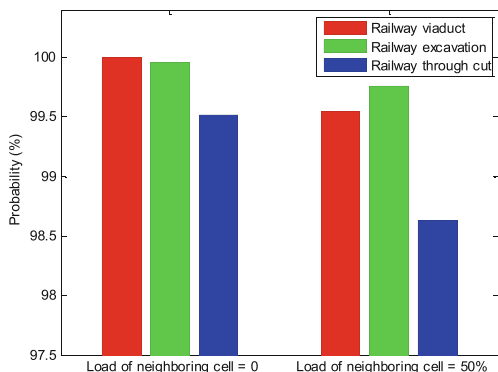


Fig. 4. The coverage probability of PUSCH SINR > 0 dB

Figure 5 shows the simulation results of UL throughput. In the intra-frequency network, UL throughput is affected to a certain degree, especially when the interference on the cell edge is severe due to neighboring cell being loaded. The throughput might fluctuate as well. However, the overall throughput can still reach over 450 kbps, which would ensure the transmission of important businesses such as locomotive synchronic operation and control, even under the most extreme condition.

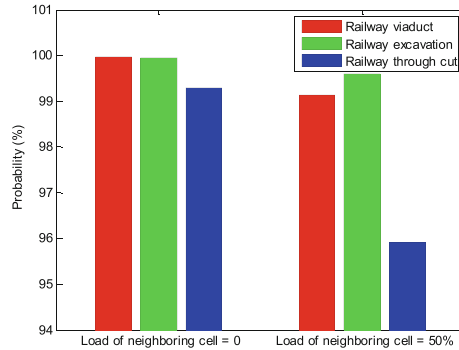


Fig. 5. The coverage probability of UL throughput > 1Mbps

According to all the simulation results, in the intra-frequency network, when the neighboring cells are with no load, all parameters can well satisfy requirements. The probability can reach over 99.30%. When the neighboring cells are loaded 50%, the interference on the cell edge increases, affecting the UL throughput. However, it can still meet the requirements of important businesses. To improve the network performance, we can utilize optimization methods after network deployment.

7 Conclusions

The TD-LTE system of Shuohuang railway has been operational since 2014. By testing this system, we show that its wireless network planning is adequate because the present system can fully meet the requirements for mobile communication of Shuohuang heavy haul railway. This system is the first TD-LTE deployment in the railway domain worldwide. We need to further validate and improve related network planning technologies of railway TD-LTE system in subsequent practical operations. This will enable us maximally utilize the advantages of LTE system and make sure the railway is operating safely and efficiently.

References

1. Wang, Y., Sun, S.: TD-LTE Principles and System Design. Post & Telecom Press, Beijing (2010)
2. Cao, Y.: Feasibility study on the broadband mobile communication system of Shuohuang railway based on TD-LTE technology. *Railw. Signal. Commun.* **4**, 68–72 (2014)
3. Li, W., Chen, B., Wu, Q., Lai, M.: Applied research of TD-LTE power wireless broadband private network. *Telecommun. Electr. Power Syst.* **11**, 82–87 (2012)
4. Zheng, G.: Analysis of TD-LTE Wireless Network Planning and Optimization. *Chin. Internet* **2**, 32–35 (2015)
5. Jie, X., Yu, K.: Application of the TD-LTE technology in Shuohuang heavy haul railway. *Railway Signal. Commun. Eng.* **12**, 20–24 (2015)

Data Mining Based Modeling and Application of Mobile Video Service Awareness

Kun Chao¹(✉), Pengfei Wang², Haina Ye¹, Lexi Xu¹,
Xinzhou Cheng¹, Mingjun Mu¹, and Chen Cheng¹

¹ China Unicom Network Technology Research Institute, Beijing 100048, China
chaokunl@chinaunicom.cn

² Beijing Branch of China Unicom, Beijing 100052, China

Abstract. With the popularity of 4G mobile networks, mobile video service becomes the key service in the 4G era. In order to improve users' awareness and increase the network optimization efficiency, it is important to establish a scientific and accurate model to evaluate the video service from the user's perception. In this paper, we focus on the video streaming traffic and propose a modelling approach to evaluate the video service performance. The essential characteristics of video traffic are taken into account. Based on the hierarchical clustering and the Pearson correlation coefficient method, key factors of video service perception are determined. Furthermore, the threshold values of key factors are obtained through extensive user surveys and simulation tests. The results of the application in the realistic network demonstrate the effectiveness of the proposed model. In addition, results show the proposed model enables the telecom operator to evaluate the video service quality of each user or user group, which helps improve the network optimization efficiency.

Keywords: Video service awareness · Data mining · Clustering · Correlation

1 Introduction

With the popularity of 4G mobile networks and the adjustment of telecom operators' tariff, mobile video service has become the most important business in the 4G era. The video traffic has accounted for more than half of the total traffic in most cities. According to the statistic of [1], nearly 70% of the current network traffic comes from video applications and the proportion will rise continuously in the future. Therefore, in order to provide better service and experience, it is critical to improve the users' perception for telecom operators, especially for mobile video service [2].

The traditional evaluation methods are based on network key performance indicator (KPI), such as signal strength and drop rate [3]. These KPI based methods can analyze the network performance, however, these methods are not aware of the user's perception. In addition, data mining is a powerful technology with great potential to discover the vital information in the data. For example, authors of [4] design a coverage self-optimization algorithm using big data analytics for WCDMA networks. However, due to specific characteristics of video service, it is important to establish a scientific and accurate model to evaluate video service from the user's perception.

In this paper, the characteristics of video service are exploited. Then, based on data mining, a video service awareness model is proposed. The model enables the operator to evaluate the video service quality of each user or user group, which helps improve the network optimization efficiency. Specifically, we consider several video perception related indexes firstly. Based on the hierarchical clustering and the Pearson correlation coefficient method, key factors of video service perception are determined. Secondly, the threshold values of key factors are obtained through extensive user surveys and simulation tests. Thirdly, we apply the proposed model in the realistic network to validate the feasibility and the performance of the proposed model.

The rest of this paper is organized as follows. Section 2 presents the proposed model of video traffic perception based on data mining. The application of the model and the results are shown in Sect. 3, followed by conclusions in Sect. 4.

2 Modeling of Video Traffic Perception Based on Data Mining

2.1 Selection of Evaluation Indicators

Referring to the division of services performance in the national standard TMF.GB917 [5], this paper divides the user perception into three dimensions, including information interaction availability, information interaction timeliness and information interaction stability. Information interaction availability refers to the access success index, which is used to measure the business availability. Information interaction is a delay related index, which measures the real-time characteristics of the service. Information interaction stability refers to metrics that measure whether the service is stable and sustainable, such as video fluency.

According to the video service signaling process and the combination of a large number of call testing, it is found that the user's perception experience is mainly embodied in two aspects. The first aspect is that the video can play normally. The second aspect is whether the video viewing is smooth. Therefore, we can get the evaluation indicators as follows in Table 1.

Table 1. Video perception related indicators

User perception	Evaluation indexes
Information interaction availability	Video playing success rate
Information interaction timeliness	First-play waiting delay
	Download speed
Information interaction stability	Average stalling internal
	Average stalling duration
	Proportion of stalling duration

The definition of each evaluation index is explained as follows. *Video playing success rate* is the percentage of successful video playing by users in total number of

video playing. *First-play waiting delay* is the length of time a user initiated video playback to a successful playback. *Download speed* is the download rate of video. *Average stalling internal* indicates the average duration between the two stalling in the playing process in unit time. *Average stalling duration* is the average duration of all the stalling. *Proportion of stalling duration* is the proportion of all the stalling duration in the total time of the video.

2.2 Determination of Key Evaluation Indicators

These indicators mentioned above are based on the service characteristics obtained from the user's perspective, and no screening index through the data, there may be redundancy. Furthermore, we have abundant data in the mobile network. Hence, we can find the key impacts on the video perception via data mining technology.

The system clustering method is filtering index based on the historical data, and pays attention to the correlation between the indexes. It considers that the index with strong correlation is redundant, and can be merged or deleted [6]. So the system clustering method is a suitable method to choose the key index from different kinds of evaluation index, and the following is the introduction of the steps of the method [7].

According to the m objects to be investigated, the relevant n indexes are selected, and the sample data matrix B is obtained by (1):

$$B = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

where x_{ij} stands for the j index in i -th clustering object.

Step1: Preprocess Data

Usually, the dimension and the order of magnitude of the index variables vary greatly. In order to put these data together for comparison, transformation are often needed. The correlation coefficient method [8] is used to measure the similarity among the indexes. The main process is as follows. Firstly, according to the existing data, calculate the correlation coefficient of every two indicators selected. Then, based on the results of the correlation coefficient, a correlation coefficient matrix R can be made, which can be used to measure the similarity between the degrees of each index. There are several methods to preprocess the data to get the correlation coefficient matrix R , such as sum standardization, standard deviation standardization, the maximum value normalization and range normalization. This paper applies the range normalization to organize the data of the matrix, as shown in (2). In (2), x_{ij} stands for the j index in i -th clustering object, x'_{ij} stands for the standardization result of x_{ij} .

$$x'_{ij} = \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}} \tag{2}$$

Among the new data using this standardization, the maximum value of each element x'_{ij} is 1, the minimum value is 0, and the remaining values are between 0 and 1.

Step2: Determine Correlation between Evaluation Indicators

Pearson correlation coefficient is a widely used metric to measure correlations. The correlation coefficient is between -1 and 1. The formula is as (3):

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}} \tag{3}$$

where N stands for the number of variables, and $i \in [1, N]$. Large absolute value of correlation coefficient r indicates the stronger correlation [9]. On the contrary, when the correlation coefficient r is close to 0, it indicates weak correlation. We can determine the degree of correlation between variables by the range of values in Table 2:

Table 2. Pearson correlation coefficient and the correlation degree

Correlation coefficient	Correlation degree
0.8–1.0	Extremely strong correlation
0.6–0.8	Strong correlation
0.4–0.6	Medium correlation
0.2–0.4	Weak correlation
0.0–0.2	Extremely weak correlation or uncorrelated

In addition, the correlation indicator can also be converted to the distance indicator d , and the distance represents the difference between metric. The value of d is smaller, the correlation of the two evaluation indexes is stronger, and then an index can be used to replace another index. On the other hand, the greater the distance is, the less the similarity between the indexes, and the two indexes are irreplaceable.

Step3: Draw Clustering Hierarchical Diagram

We need to find the shortest distance $d_{pq} = \min\{d_{ij}\}$ in the previous $m \times m$ correlation coefficient matrix or in the non-diagonal elements of the distance matrix. Then, we combine the classification objects G_p and G_q into a new class G_r . Therefore, we can compute the distance between the new class and the original various classes using (4):

$$d_{rk} = \min\{d_{rk}, d_{qk}\} (k \neq p, q) \tag{4}$$

Thus a new $m-1$ order distance matrix can be obtained. Then select the closest d_{ij} from this new distance matrix, and merge G_i and G_j into the new class.

We get the correlation coefficient matrix using the range standardization formula, and then obtain the symmetric matrix R by the use of the Pearson correlation:

$$R = \begin{bmatrix} 1 & 0.235 & 0.097 & 0.021 & 0.025 & 0.006 \\ 0.235 & 1 & 0.063 & 0.051 & 0.002 & 0.015 \\ 0.097 & 0.063 & 1 & 0.834 & 0.039 & 0.001 \\ 0.021 & 0.051 & 0.834 & 1 & 0.011 & 0.076 \\ 0.025 & 0.002 & 0.039 & 0.011 & 1 & 0.618 \\ 0.006 & 0.015 & 0.001 & 0.076 & 0.618 & 1 \end{bmatrix} \quad (5)$$

The Pearson correlation coefficient method is used to cluster analysis, and the shortest distance clustering hierarchical diagram is drawn by combining the above clustering process, as shown in Fig. 1. The coefficients obtained by clustering are listed in Table 3.

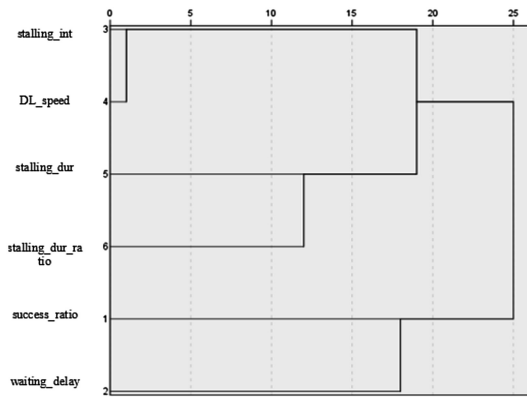


Fig. 1. Shortest distance clustering hierarchical graph

Table 3. System clustering parameters

Stage	Cluster combined		Coefficients	Stage cluster first appears		Next stage
	Cluster1	Cluster2		Cluster1	Cluster2	
1	3	4	0.834	0	0	4
2	5	6	0.618	0	0	4
3	1	2	0.235	0	0	5
4	3	5	0.130	1	2	5
5	1	3	0.097	3	4	6

Step4: Determine Key Evaluation Indicators

According to the clustering results, if there is a strong correlation between the two indexes, we can use an index to replace the other index or merge the two indexes. Thus, the number of the final evaluation indexes will be reduced. However, if the correlation coefficients among all indexes are small, we do not need to exclude the indexes.

Figure 1 and Table 3 show that the indexes download speed and average stalling internal have extremely strong correlation, and the indexes average stalling duration and proportion of stalling duration have strong correlation. Therefore, we can simplify the six indexes into the four indexes which are video playing success rate, first-play waiting delay, the average stalling internal and the average stalling duration.

2.3 Evaluation of Video Service Perception

Based on above result of the key indexes, through a questionnaire survey on a large number of users and simulation, we confirm the threshold of each key evaluation index. By the threshold, we can classify the grade of each key evaluation index as shown in Fig. 2. Then, the video service perception can be evaluated.

User perception	Information interaction availability	Information interaction timeliness	Information interaction stability	
Key evaluation indicator	Video playing success rate	First-play waiting delay	Average stalling internal	Average stalling duration
Threshold and grade of each indicator	A: ≥ 90% B: ≥ 85% C: <85%	A: ≤ 10s B: ≤ 30s C: >30s	A: ≥ 60s B: ≥ 30s C: <30s	A: ≤ 15s B: ≤ 20s C: >20s

Fig. 2. The grade classification of each key evaluation index

Based on above grade classification result of each key evaluation index, together with the actual experience of using video service for users, we can reach the degree of the user experience. The details are shown in the Table 4. Usually, the customers are more stalling sensitive than the accessing rate and delay. After all, it is angrier that the

Table 4. Evaluation degree division of video service perception

Degree of user experience	User experience	The indexes “a b c d”
Excellent	Watch smoothing	AAAA
Good	Totally good, but some minor flaws	AAAB, AABA, ABAA, BAAA
Medium	Experience is already impacted	BAAB, BABA, ABAB, ABBA, AABB, BBAA, CAAA, ACAA
Bad	Bad experience	The remaining possibility

watching is interrupted when you are interesting in it. In order to simplify the subsequent express, we identify the indexes *a*, *b*, *c*, *d* for video playing success rate, first-play waiting delay, average stalling internal and average stalling duration. Moreover, A means “good”, B represents “medium”, and C explains “bad”.

3 Application of Video Service Perception Model

Based on the proposed video service perception model, we can get the perception of each user or certain user group. Meanwhile, we can obtain the video service quality of each mobile station cell or some area. One of the applications of the proposed video service perception model is as follows: we used this model to evaluate the realistic mobile network of China Unicom in an eastern city in China. The city has more than 6 million billing users, and each of them has a video perception result, such as excellent, good, medium or bad. Table 5 shows the example of the video awareness for each user.

Table 5. Video awareness for each user

IMSI	a	b(ms)	c(ms)	d(s)	The rating				User experience
					a	b	c	d	
20205XXXXXX2050	100%	162.4	26942.99	10	A	A	B	B	Good
20404XXXXXX2304	50%	5195	26129	6	B	A	B	A	Medium
20404XXXXXX6332	100%	638	10661.88	12	A	A	C	B	Bad
20404XXXXXX0534	100%	344	0	0	A	A	A	A	Excellent
20801XXXXXX6798	100%	140	0	0	A	A	A	A	Excellent

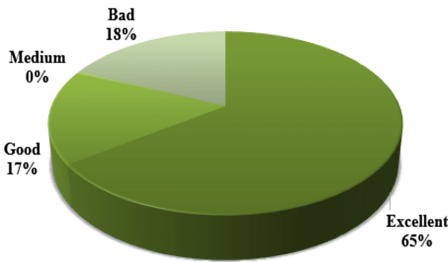


Fig. 3. Perception of users in one city

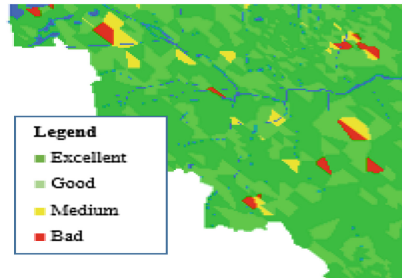


Fig. 4. Perception of areas in one city

On one hand, Fig. 3 shows the video service perception of all the users in the city. The results are 65% of the users using video service in the city have excellent experience, 17% have good experience, 18% have bad experience and nearly 0% have medium experience. On the other hand, Fig. 4 presents the video service quality of all the mobile station cells of China Unicom in the city. The dark green means the excellent quality of the video service of the cell, the green means the good quality, the

yellow means the medium quality and the red means the bad quality. Hence, the red areas are extremely needed to be focused on and optimized.

By means of the result of the proposed model, we can find the problem cells in the realistic network. The results verify 70% of the problems found in the eastern city in China by the proposed model have been certified existing in the realistic network. Therefore, the proposed model provides an effective and efficient way to evaluate the video service perception and locate the related problems in the realistic network.

4 Conclusion

This paper introduces an analytical approach to evaluate the video service from users' perception based on data mining technology. The approach takes into account the essential characteristics of video traffic, such as video playing success rate, first-play waiting delay, download speed, average stalling internal, average stalling duration, and the proportion of stalling duration. Based on the hierarchical clustering and Pearson correlation coefficient method, the correlation among the factors are analyzed in order to determine the key factors of video service perception. Further, the threshold values are obtained through extensive user surveys and simulation tests. Results of the application in the realistic network demonstrate the effectiveness of the proposed model. In addition, results also show the proposed model can be used to evaluate the video traffic quality of each user or user group, which presents users' awareness of the video service directly. Furtherly, telecom operators can use the analyzing results to optimize the network effectively.

References

1. Analyzer: video is the future highland for telecom operators. <http://tech.sina.com.cn/t/2016-07-06/doc-ifyxtrwtu9993330.shtml>. Accessed 10 Jul 2017
2. Chao, K., Wang, P., Xu, L., et al.: A novel big data based telecom user value evaluation method. In: 1st International Congress on Signal and Information Processing Networking and Computers, Beijing, pp. 375–384 (2015)
3. Zhang, X., Liao, Z., Cao, L., et al.: Research on perceptual evaluation of end-to-end mobile users based on mobile internet. *Mod. Sci. Technol. Commun.* **2012**(11), 73–75 (2012)
4. Gao, J., Cheng, X., Xu, L., et al.: A coverage of self-optimization algorithm using big data analytics in WCDMA cellular networks. In: 1st International Congress on Signal and Information Processing Networking and Computers, Beijing, pp. 289–297 (2015)
5. Du, Y., Bao, H.: A QoE based evaluation of service quality on WAP in wireless network. *J. Beijing Union Univ.* **24**(2), 78–81 (2010)
6. Tang, Y., Sun, F., Sun, Z.: Neuro-fuzzy system modeling based on automatic fuzzy clustering. *J. Control Theor. Appl.* **2**, 121–130 (2005)
7. Li, M., Jiao, L., Lin, B., et al.: Polyclonal clustering algorithm and its convergence. *J. China Univ. Posts Telecommun.* **15**(3), 110–117 (2008)
8. Li, D.: Measurement of geo economic relations in Shanghai. *Stat. Obs.* **18**, 83–84 (2007)
9. Zhang, J., Gao, R., Hu, J., et al.: Application comparison of grey relational grade and pearson correlation coefficient. *J. Chifeng Univ. (Natural Science Edition)* **21**, 1–2 (2014)

The Research of Virtual Drive Test Based on MR and CDR

Zhiqiang Lv^(✉), Saibin Yao, Ling Li, Yongjia Qi, and Jialong Liang

Shanghai Branch of China Unicom, Shanghai 200000, China
lvzjl6@chinaunicom.cn

Abstract. Drive Test (DT) is a very important approach to measure the quality of the wireless networks. But with the rapid development of wireless networks, the network structure comes to more and more complicated, such as massive network, different systems and multiple manufacturers. The complexity of network leads to a mass expenditure in operation and maintenance which including DT. In this paper, we study a new method, called Virtual Drive Test (VDT) by using Measurement Report (MR) and Call Detail Record (CDR), to measure the quality of the wireless networks in more efficient and less costly ways.

Keywords: Drive test · Wireless networks · Virtual drive test
MR · CDR

1 Background

Wireless networks face many challenges, such as mass networks, different systems, multiple equipment manufacturers. The complexity of network operation and maintenance, technical requirements and maintenance costs increased significantly. MR (MDT) measurements can quickly capture the network coverage of different operators, effectively assessing the competitor's network quality. This method is more efficient and less costly than road tests.

2 The Principle of Virtual Road Test

2.1 Indoor Site Location

By obtaining the site type from the engineering parameter, the indoor site position can check whether the measured cell is an indoor cell, and use the weighted Centroid Geometry to position the measurement results which serving cell is an indoor cell indoor unit of the serving cell in the adjacent list. Based on the fact that the indoor cell's coverage should be smaller than the Macro cell, so when the indoor weight increase, the positioning mean error can be reduced to a reasonable range.

$$M_{WC}(x, y) = \frac{\sum_{i=1}^k w_i \cdot BS_i(x, y)}{\sum_{i=1}^k w_i} \tag{1}$$

$M_{WC}(x, y)$ represents the location generated by Centroid Geometry location, w_i is the weight, and $BS_i(x, y)$ is the cell position measured by the measurement result (Fig. 1).

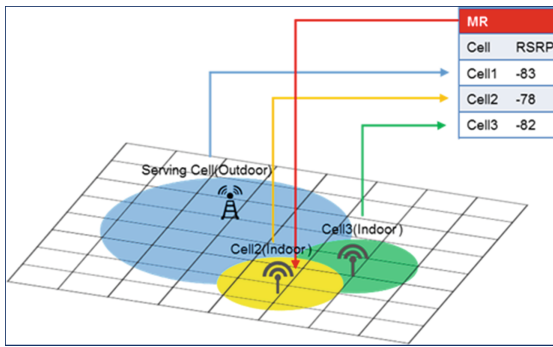


Fig. 1. The introduction of MWC

2.2 Feature Pattern Matching

Matching the former results with the existing feature map library, the measurement result can be replaced by using the closest “feature pattern”.

The key points of its are:

- (1) The high-precision library should be based on existing networks;
- (2) The evaluation model of the feature should be based on adjacent cell strength and other measurement information.
- (3) Substituting the time delay and antenna direction angle into the range of the matching candidate grid group.
- (4) By iterative operations, the best grid with the highest score can be found from the candidate grid.

$$M = \sum_{n=1}^N k_n * C_n * [C_{MatchConstant} - abs(Rx_n - rx_n)] \tag{2}$$

Where M is the score of the matching result, k_n is the weight of the matching cell, C_n is the common damping coefficient, $C_{MatchConstant}$ is a constant match, it adjust the

relationship between adjacent cells and level matching, R_{x_n} is the signal strength of cell n in the measurement result, R_{x_n} is the signal strength of cell n in the grid.

Each grid can generated a score of M, and the greater the M score, the higher the match, and the grid with the largest M score is the located point (Fig. 2).

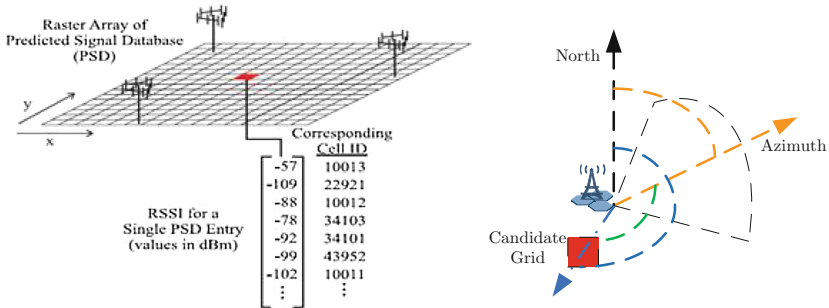


Fig. 2. The introduction of M

2.3 Map Matching and Position Calibrating Algorithm

The map matching and Position Calibrating algorithm is based on the analysis of user behavior characteristics and combines the electronic map of the geographical information for further calibration, including:

- (1) Building Matching Method: Analyze user behavior characteristics to identify indoor measurement results, and combine the corresponding location of the building on the electronic map, if measurement report outside the building, it is calibrated to the building b by a grid indexing algorithm based on Euclidean distance.
- (2) Road Matching Method: For outdoor measurements that meet the requirements for high-speed and long-distance calls, projection and calibration can be done by examining the most similar path through the measurement path in the electronic map.
- (3) High speed: According to the user characteristics to identify the outdoor call, more than 2/3 of the users keep moving in more than 30 km/h.
- (4) Long call: Identify the interval time of measurement less than 42 s, the duration of the call last for more than 2 min, and the index of MRCnt over 30 (Fig. 3).

The best path selection formula is:

$$R_{matching} = \text{MAX} \left\{ \sum_{i=1}^{P_1} W_{R_i}, \sum_{i=1}^{P_2} W_{R_i}, \dots, \sum_{i=1}^{P_q} W_{R_i} \right\} \quad (3)$$

W_{R_i} is the weight of road R_i , $R_{matching}$ is the matched best path.

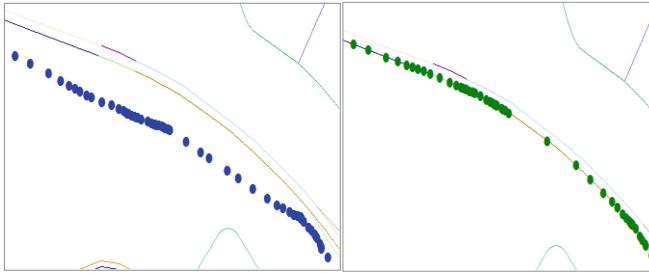
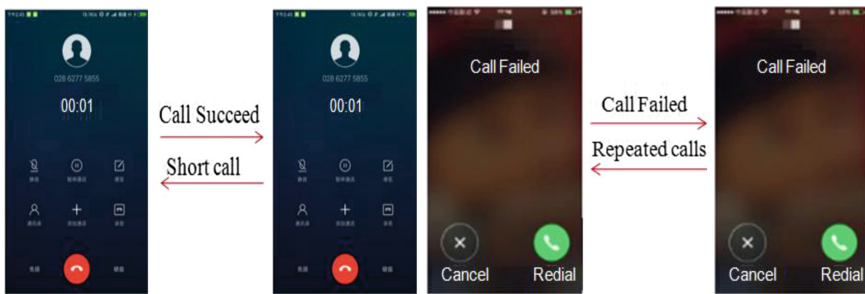


Fig. 3. The introduction of Road match

2.4 User Behavior Model Established

With the development of 4G system, users demand higher quality of data service experience, the basic voice business perception is ignored in a sense. So the biggest problem faced by operators is how to find the failed voice service place based in user behavior model. there are two main reasons that the Voice perception comes to poor: on the one hand, when there exists a two sides call, one of the phone is not connected. on the other hand, both of the phone connected to the system, but as the wireless situation quality is not good, unilateral users hang up, resulting in short talk time. Shanghai Unicom has established two kinds of user behavior model based on these situations. Combine the high-speed rail users with CDR, which establishes voice evaluation system for high-speed rail users:

- (1) Short talk: all users under the cell whose talking time is less than 3 s
- (2) Short call between two users: two users call in 3 s repeatedly (Fig. 4)



(a) Short talk

(b) Short call between two users

Fig. 4. The User behaviors from CDR

3 Experiment

3.1 Select Test Area

We set the inner ring area as test area and also consider the high traffic flow and frequency planning. In this paper we choose 100 stations to experiment.

3.2 Virtual Drive Test Based on CDR

(1) Shot talk

Step1: Output the top list through short calls from high-speed rail users

Step2: No configuration for SAC when testing on the spot

It sends the initial direct message Uplink, but does not receive the message from the downlink, and then receives the RRC release message from the downlink, causing the call to fail. For this phenomenon, it is found that SAC is not configured (Table 1).

Table 1. The top list of short calls

Cell name	Ratio of short talk ratio (<3 s)
li'an xi_2(1)	94.35%
li'an xi_2(2)	92.50%
min ming gu_3(3)	91.79%
min zhong hang_2(2)	90.18%
songwanchen_1(3)	87.39%
songwanchen_2(3)	85.78%
songzhiyuan_2(2)	85.28%

Step3: After SAC is set, the KPI for DT become normal

After SAC is set, the initial direct message is sent upstream and then the direct message sent by the downlink is received and the call is successful (Fig. 5).

After SAC is set, the test indicators show that the voice business back to normal, the problem is resolved.

(2) Short call ratio between two users

Step1: Converge the shortcomings ratio TOP list through the high-speed rail shortcuts cell:

Step2: HD voice encryption when the encrypted information changes, Nokia MSC in accordance with 3GPP 25.413 requirements to carry encryption information, then the problem is solved (Table 2).

3.3 Virtual Drive Test Based on MR

The RSRP trend of Virtual Drive testing and DT match very well, especially the problems in road (Table 3).

PC Time	Message
13:19:20.017	W->SIBType11
13:19:38.690	CM Service Request
13:19:38.690	W->RRCConnectionRequest
13:19:38.909	W->MIB
13:19:38.909	W->SIBType7
13:19:39.236	W->MIB
13:19:39.236	W->SIBType7
13:19:39.347	W->RRCConnectionSetup
13:19:39.564	W->RRCConnectionSetupComplete
13:19:39.564	W->InitialDirectTransfer
13:19:39.783	W->Measurement Control
13:19:39.891	W->Security Mode Command
13:19:39.891	W->Security Mode Complete
13:19:39.891	Setup
13:19:39.891	W->Uplink Direct Transfer
13:19:40.218	W->Downlink Direct Transfer
13:19:40.218	Call Proceeding
13:19:40.655	W->RadioBearerSetup
13:19:40.874	W->RadioBearerSetupComplete
13:19:41.201	W->Measurement Control
13:19:44.696	W->Downlink Direct Transfer
13:19:44.696	Alerting
13:19:48.084	W->Downlink Direct Transfer
13:19:48.084	Connect

Fig. 5. The normal signaling flow

Table 2. The top list of Short calls between two users

Station name	Ratio of short call between two users
Shen lan21	35.12%
Min tian shen31	24.52%
Hua yi_2(0)	22.22%
Shen guang_2(1)	18.55%

Table 3. The error rate of virtual drive test based on MR

Coverage ratio	Virtual drive testing	DT	Error rate
RSRP ≥ -90 dBm	79.59%	81.40%	1.81%
RSRP ≥ -95 dBm	83.78%	85.02%	1.24%
RSRP ≥ -100 dBm	92.73%	94.42%	1.69%
RSRP ≥ -105 dBm	97.12%	98.37%	1.25%
RSRP ≥ -110 dBm	99.10%	99.80%	0.70%
RSRP ≥ -115 dBm	100%	100%	0.00%

4 Conclusion

This paper mainly introduces a method of Virtual Drive Test which is based on MR and CDR to evaluate the user perception information. This method achieved good results in practical application. There are two reasons why Virtual Drive Test based on MR and CDR can replace traditional method of road test. On the one hand it can save a lot of manpower and material resources, and on the other hand it can enrich the index system through the establishment of users' behavior model.

References

1. Sun, R., et al.: Nonisotropic scattering characteristic in an alternant tree-blocked viaduct scenario on high-speed railway at 2.35 GHz. *Int. J. Antennas and Propag.* **2014** (2014)
2. Chang, C.-L., Wang, S.: Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **44**(6), 1575–1585 (2006)
3. He, R., Zhong, Z., Ai, B.: Path loss measurements and analysis for high-speed railway viaduct scene. In: *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*. ACM (2010)
4. Rogati, M., Yang, Y.: High-performing feature selection for text classification. In: *Proceedings of the Eleventh International Conference on Information and Knowledge Management*. ACM (2002)
5. Zhao, J., Ke, L., He, X.: Locality sensitive semi-supervised feature selection. *Neurocomputing* **71**(10), 1842–1849 (2008)
6. Hagenauer, J., Hoehner, P.: A Viterbi algorithm with soft-decision outputs and its applications. In: *Proceedings of IEEE Global Telecommunications Conference and Exhibition' Communications Technology for the 1990s and Beyond' (GLOBECOM)*, 1989. IEEE (1989)
7. Kumar, A., Manjunath, D., Kuri, J.: *Communication Networking: An Analytical Approach*. Elsevier, Amsterdam (2004)

Big Data Research on Driving Behavior Model and Auto Insurance Pricing Factors Based on UBI

Heng Zhang¹(✉), Lexi Xu¹, Xinzhou Cheng¹, Weiwei Chen¹,
and Xueqing Zhao²

¹ China Unicom Network Technology Research Institute, Beijing 100048, China
{zhangheng23, xulx29, chengxz11,
chenww10}@chinaunicom.cn

² University of California, Davis, CA 95616, USA
xqzzhao@ucdavis.edu

Abstract. With the popularization of cars in China, traditional auto insurance market experiences fierce competition. Auto insurance company concerns with insurance product innovation, thus improving the insurance service level. For the purpose of auto insurance innovation, this paper proposes a novel driving behavior model to evaluate the driving risk. Through logistic regression algorithm, we analyze the correlation between the driving score and the accident. Then, we discuss the reliability of this model. Furthermore, we employ this driving score to the auto insurance pricing model, in order to improve the risk identification of the pricing model. Both the novel driving behavior model and the auto insurance pricing model can achieve the effective risk segmentation and precise pricing, as well as assisting the auto insurance company to improve the market competition capability.

Keywords: UBI · Driving behavior · Insurance model · Logistic regression

1 Introduction

In recent years, the User-Behavior Insurance (UBI) market has been developing rapidly, especially in the American and Europe [1]. By analyzing the real driving behavior, UBI products can achieve the precise risk segmentation and improve the service quality of the insurance company. In addition, UBI products can provide a variety of value-added services. UBI products can also strengthen service innovation and enhance customer's perception [2]. Therefore, UBI leads to a revolution of the insurance industry, especially in product innovation and customer service.

In China, the auto insurance market becomes a competitive industry, meanwhile, the product homogenization phenomenon is serious. Hence, the commercial auto insurance reform is inevitable. This brings new challenges to both the product innovation and the product pricing. Auto insurance, which is based on UBI, is a promising type of auto insurance recently [3]. UBI auto insurance can meet the requirements of

commercial insurance reform. Furthermore, UBI auto insurance can improve the profits of insurance company and the insurance service level.

With the rapid development of mobile internet technology, intelligent mobile phones are widely used. Therefore, data collected by intelligent mobile phones is feasible and this provides the data sources for the auto insurance research and application. Based on the UBI data, vehicle data, user's mobile data and weather data, a novel driving behaviors model is proposed in this paper. In addition, this paper researches the relationship between driving model and the risk.

2 Data Source Composition

Data source is the key element for the big data analysis. In recent years, the rapid development of intelligent devices have made it possible for large-scale applications of UBI technology. The user's daily behavior can be obtained through the UBI terminals [4]. UBI data analysis and mining technology will help to research driving behaviors and differentiated pricing. The UBI data source can be divided into:

- (1) Car hardware box: The frequency of data collection from box is second level. High collection frequency will make pressure to data storage system. However, low collection interval will be inaccurate to the resulting of behavioral data. Data is as follows: Trip_Number, Trip_Begin, Trip_End, D_time, GPSPileag, Accel, GPSAverage_Speed, GPSSpeed_Max, etc.
- (2) Mobile operator data: Due to enhanced functions of mobile phone, operators can achieve some data to replace several types' data from hardware box (e.g., time, location, gravity induction), in order to enhance the ease of data acquisition [5, 6].
- (3) Road data: The actual road conditions will affect the driver's safety and convenience, such as the actual driving of different road types, the degree of congestion, historical archived location.
- (4) Weather data: Weather conditions (e.g., rainfall, snow, wind, visibility etc.) affect the travel environment, the visibility and road conditions.

3 Driving Behavior Model Analysis

3.1 Analysis Method

Different users have different driving behaviors, thus, the driving risks caused by different driving behaviors vary greatly. Therefore, the research of the user driving model is the basics to analyze the insurance risk. Model research consists of three aspects:

Aspect1: Basic UBI data analysis. The research obtains customer's driving behavior data per second, at the same time, deep tests are implemented to guarantee the data quality which can promise the required precision. And through the statistical analysis, UBI customer's characteristics are obtained, and we get an intuitive understanding of customer driving behavior.

Aspect2: Calculating the driving score. Based on driving behavior data and real claim data, a generalized linear model is established, and the model output can be translated into driving score. At present, we have the location information from operator, and other data rely on three party data.

Aspect3: Application of driving score. We will take the driving score as a factor, then, we employ this factor to the auto insurance pricing model to make it more accuracy. After increasing the driving score factor, the risk identification ability of the auto insurance pricing model will be improved.

3.2 Factor Composition

Factors are independent variables of driving score model, and they can influence driving behavior. Some factors can be obtained directly from the driving behavior data collected by the UBI equipment, and other factors need to match the external data acquisition (e.g., weather information, map information, etc.) (Table 1).

Table 1. The driver scoring model factor

Category	Factor	Influence
Time	Peak time/Driving time/Continuous driving time/Working time or weekend	The driving time will reflect driver’s status
Speed/acceleration/direction	Driving speed/Engine speed	Reflect the driver’s driving behaviors, belong to impatient, smooth or slow style
	Rapid acceleration -Longitudinal acceleration	
	Quick brake-Longitudinal acceleration/Quick play round-Lateral acceleration	
	Uphill-vertical/Downhill-vertical acceleration	
	Turning-direction change	
	Lane change- changes of direction and speed	
	Overtaking-changes of direction and speed	
Weather information	Rain and snow/Wind power/temperature/visibility	Reflect the weather’s impact on driving condition
Map information	Road type/Road flow/Speed limit	Reflect driving risks which may exist on the road
	Intersection or not/History road accident	
Car information	Car brand/Car model/Driving range	The car’s status will reflect the consumption ability
	Times of auto insurance accident/Car price	

3.3 UBI Driver's Features

We make an analysis of the customers who have installed the UBI equipment. Because male and young drivers are more interested in new electronic technology products, they are the main UBI equipment customers. As the statistics, the proportion of male drivers is higher than the female. For young drivers, the proportion is higher than the older ones. In contrast, the effect of gender is more obvious than age (Fig. 1).

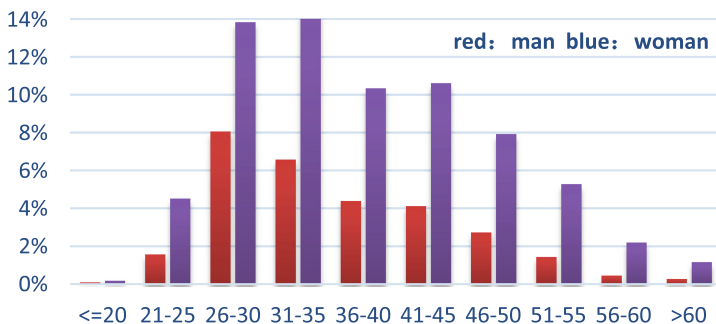


Fig. 1. UBI driver's age and gender distribution

From one day's perspective, the peak value of the traffic appears in the morning and evening peak. From a weekly perspective, morning peak only occurs on weekdays, while evening peaks occur on weekdays and weekends at the same time. However, foreign data generally show that neither morning nor evening peaks will occur on weekend (Fig. 2).

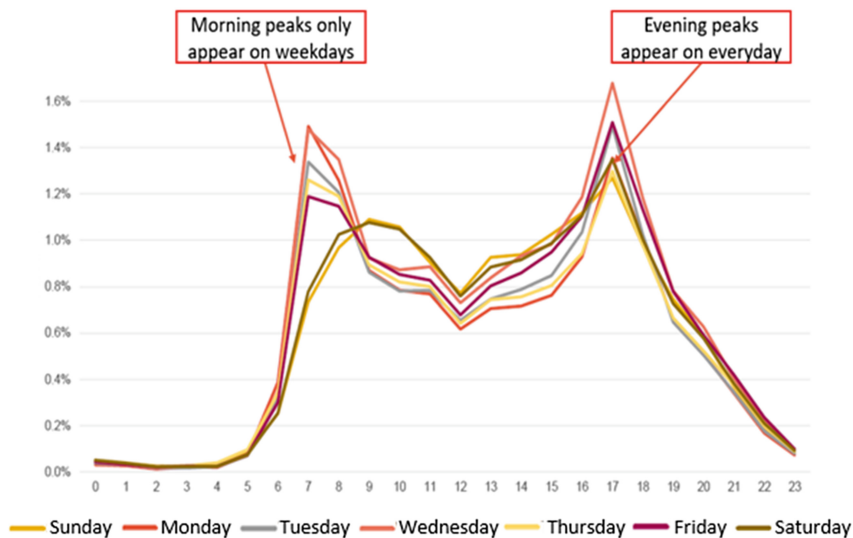


Fig. 2. UBI customer driving time statistics

4 Insurance Model Research

4.1 Driving Model

In this paper, an insurance driving model is established based on driving behavior, which uses the driving behavior factors as independent variable and accidents as dependent variable. The classification methods for classification variables include Logistic regression, decision tree, random forest, neural network etc. Considering that Logistic regression can show the relationship between dependent variable and independent variable by regression equation and choose variables, logistic model is used in modeling.

Supposing the sample is $\{x, y\}$, y is 0 or 1, which means positive or negative class, and x is the sample eigenvector of our m dimension. Then, this sample x belongs to the positive class, hence, the probability of $y = 1$ can be expressed by the following logical function (1):

$$p(y = 1|x; \theta) = \sigma(\theta^T x) = \frac{1}{1 + \exp(-\theta^T x)} \tag{1}$$

There are n independent training samples $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $y = \{0, 1\}$, x is the driving factor score model mentioned above, y means if accident happened or not, hence, the analysis of sample data are as Table 2 (x, y sample only for example):

Table 2. Sample data

x samples								y samples
Car id	Driving time	Driving distance	Driving section	Average speed	Emergency acceleration	Rain & snow	Dangerous area	Accident or not
2788	9:20	30	Ring road	100	3	0	No	No
2789	22:45	100	Township road	80	5	15	Yes	Yes
2790	6:10	10	expressway	120	6	20	No	No
...

The probability of the occurrence for each observed sample (x_i, y_i) is as (2).

$$P(y_i, x_i) = P(y_i = 1|x_i)^{y_i} (1 - P(y_i = 1|x_i))^{1-y_i} \tag{2}$$

For the whole sample set, the likelihood function for the n independent samples occurred is (since each sample is independent, the probability of the occurrence for the n samples is the multiplication of their respective probabilities):

$$L(\theta) = \prod P(y_i = 1|x_i)^{y_i} (1 - P(y_i = 1|x_i))^{1-y_i} \tag{3}$$

The maximum likelihood method is to find the value of theta in the model to make the likelihood function maximum. This maximum likelihood is the cost function. Then, the next step is to optimize the solution and solve equation, and the results are calculated as (4):

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \frac{e^{\theta^T x_i}}{1 + e^{\theta^T x_i}} x_i = \sum_{i=1}^n (y_i - \sigma(\theta^T x_i)) x_i \quad (4)$$

Here, the gradient descent method is used to find the local optimal solution of the function by employing the first order gradient information, the result is as (5):

$$\theta^{t+1} = \theta^t - \alpha \frac{\partial L(\theta)}{\partial \theta} = \theta^t - \alpha \sum_{i=1}^n (y_i - \sigma(\theta^T x_i)) x_i \quad (5)$$

4.2 Auto Insurance Pricing Model

The driving behavior factor is added to the auto insurance pricing model to make up for the lack of personal behavior information. Further it can improve the model prediction accuracy, and enhance the risk differentiation ability (Fig. 3).

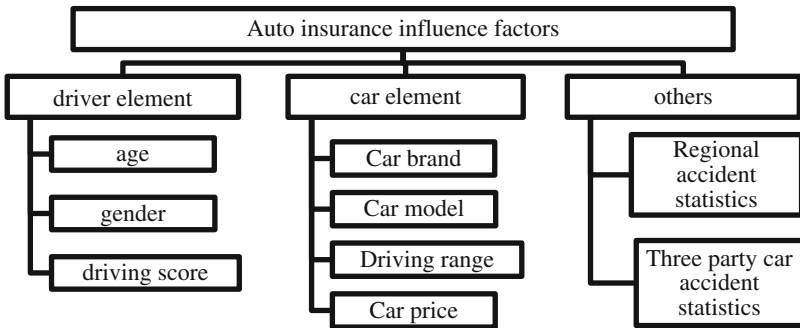


Fig. 3. Auto insurance influence factor

Customer’s driving scores represent the risk level. This score will be converted to a form which match user’s behaviors better. Higher the score, more safe the driving is. But it don’t has directly proportional relationship with risk level. Such as, 90’s score has a better driving behaviors than 80’s, but any specific difference can’t be judged.

4.3 Relevance of Driving Behavior and Risk

Figure 4 shows the one-dimensional analysis of car accident frequency. The abscissa is driving score segment (Note: the abscissa of each segment represents the score in the +2.5 range, such as 7.5 score means the range from 5 to 10), the histogram shows the degree of risk exposure, the line chart represents the actual accident frequency.

From Fig. 4, there is a negative correlation between occurrence frequency and driving score. A low score below 60 points have a big fluctuation due to less risk's exposure. Claim frequency decreases with the rising trend of driving score on the other section. Fluent driving behaviors indicates the low probability of accidents.

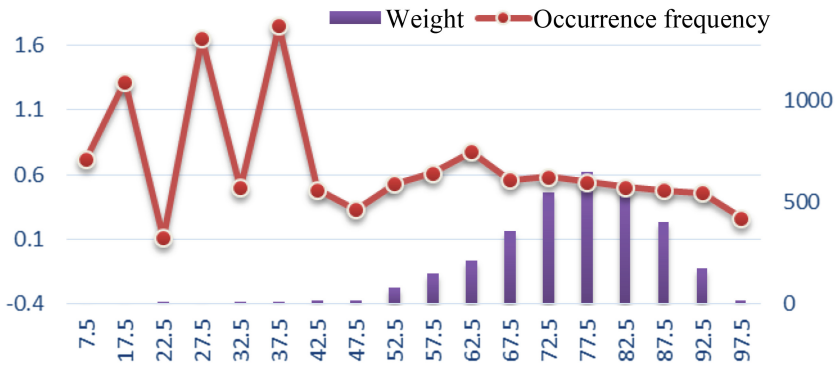


Fig. 4. One-dimensional analysis of car accident frequency

Driving behavior affects the risk level significantly. The increase of the driving score factor can improve the risk identification ability of the existing pricing model, and the optimization effect is quite significant, which helps to achieve risk segmentation and precise pricing.

5 Conclusions

UBI pricing research can enhance the market competition in the insurance pricing, as well as improve the company's capabilities. In addition, UBI contains abundant valuable information. This paper proposes both a novel driving behaviors model and an auto insurance pricing model. The proposed two models utilize and analyze UBI data, vehicle data, user's mobile data and weather data comprehensively. The proposed two models can achieve effective risk segmentation and precise pricing.

References

1. Fan, C.K., Wu, X., Zheng, D., Lin, W.: A market analysis of telematics-based UBI in Taiwan. *J. Appl. Finance Banking* **6**(6), 71–89 (2016)
2. Fan, C., Wang, W.: A comparison of underwriting decision making between telematics-enabled UBI and traditional auto insurance. *Adv. Manage. Appl. Econ.* **7**, 1–5 (2017)
3. Brocke, J.V., Schmiedel, T.: *BPM - Driving Innovation in a Digital World*. Springer, Heidelberg (2015)
4. Dijksterhuis, C., Lewisevans, B., Jelijs, B., Tucha, O., Waard, D.D.: In-Car usage based insurance feedback strategies. *Ergonomics* **59**(9), 1158–1170 (2016)

5. Xu, L., Chen, Y., Gao, Y., Cuthbert, L.: A self-optimizing load balancing scheme for fixed relay cellular networks. In: 2th IET International Conference on Communication Technology and Application, pp. 306–311. IET Press, Beijing (2011)
6. Xu, L., Luan, Y., Cheng, X., Fan, Y., Zhang, H., Wang, W., He, A.: Telecom big data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)

Compliance Testing for Data Quality Assurance: Definitions, Models and Applications

Xu Mao¹(✉), Fei Su², Kuitong Xian¹, and Kaicheng Xu¹

¹ China National Institute of Standardization, Beijing, China
maoxu@cnis.gov.cn

² China Unicom Network Technology Research Institute, Beijing, China

Abstract. Entering the 21st century, data are considered as nationally strategic resources for business innovations. Currently, it is a global trend to utilize big data to promote economic development, improve social governance, and enhance governmental services and regulatory capabilities. However, the low-quality data are seriously hindering applications for data analysis and decision support. Testing is a necessary part of quality assurance. This paper presents a theoretical survey on compliance testing for data quality assurance. Firstly, basic concepts of data quality and compliance testing are studied. Secondly, data quality testing models are proposed, involving testing element, testing object and testing process. Finally, the application model of documents data quality testing is demonstrated through requirement analysis of documents.

Keywords: Data quality · Data standards · Compliance testing · Documents

1 Introduction

In 1963, the Japanese scholar Tadao Umesao in his *Information Industry Theory: Dawn of the Coming Era of the Ectodermal Industry* first predicted the coming era of information revolution [1]. In 1967, the Japanese government defined informatization from the economic aspect. The informatization is an advancing process of information society, which reflects the fundamental change of leading role from tangible material products to intangible information products. Depending on description and evolution of informatization in the human history, this paper holds a view that informatization emphasizes the scale of information labor force caused by various applications of information technology. Furthermore, this paper divides informatization into four periods: electronic, networking, data quality and intelligent. The four periods of informatization are shown in Fig. 1.

Entering the 21st century, information technology and economic society are deeply integrated, which leads to huge data explosion from society, sensors and cyberspace. Data are considered as nationally strategic resources for business innovations [2]. Currently, it is a global trend to utilize big data to promote economic development, improve social governance, and enhance governmental services and regulatory capabilities. The premise for effective analysis of big data is to guarantee the quality of

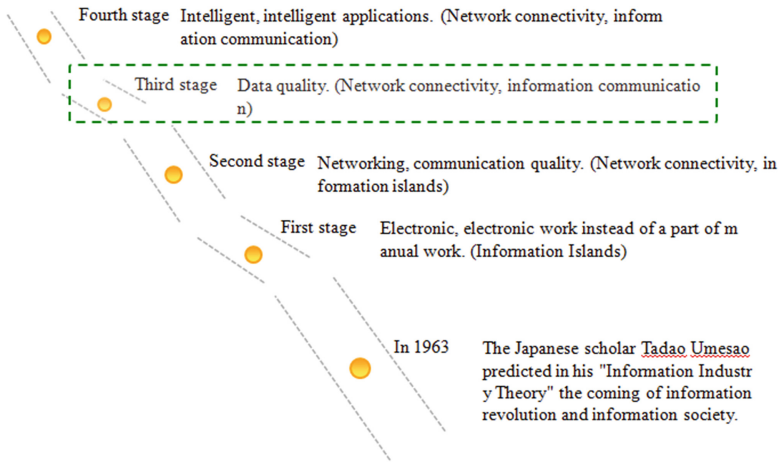


Fig. 1. The four periods of informatization

collected data. That is because professional data analysis tools extract implicit, accurate and useful information only in the big data environment of high quality. Based on high quality analysis results, enterprises making decisions will not deviate from the normal track to make business decisions. Otherwise, even if data analysis tools are advanced, a junk big data environment can only extract meaningless junk information. Therefore, data management must be conducted like traditional product or service management to ensure data quality. However, since the inherent attributes of intangible data differ from tangible products or services, the principles of data quality, management, and technology are different from the traditional product quality or quality of service.

Testing is a necessary part of quality assurance. Data quality testing is gradually emphasized by software developers and software users, which is also as a necessary engineering part to ensure data quality in the process of software system development,. Implementing conformance testing and monitoring data life cycle strictly aim to detect data defects, assess and improve data quality, meeting user expectations for measured data [3].

Although data quality testing can learn from the existing software testing methods in a certain extent, but they are not the same. The data quality testing is not standardized. For example, many software developers and users are still misunderstanding on data quality testing, and data quality testing theories are lacking, and the data quality testing process management is incomplete.

This paper focuses on building a set of top level theoretical models for data quality testing, which solves the following problems:

- (1) how to fully understand data quality testing?
- (2) how to select the testing object of data quality?
- (3) how to establish a basic data quality testing process?

2 Definitions

2.1 Software and Software Testing

Software, or computer software, is a part of a computer system that consists of data or computer instructions, in contrast to physical hardware from which the system is built [4]. Software quality is very important. If a software is buggy, and the bugs must be discovered and debugged through software testing.

Software testing is an investigation, which is conducted to provide stakeholders with information about the quality of software under test. In general, the properties of “fit for use” are listed as follows [5]:

- (1) meeting the requirements of its design and application.
- (2) responding correctly to inputs.
- (3) performing functions timely.
- (4) usable.
- (5) can be installed and run in intended environments.
- (6) achieves the general result that stakeholders desire.

To a great extent, software testing conventionally emphasizes on debugging computer instructions, however, the quality of data circulating in software are thoroughly ignored.

2.2 Data Lifecycle

Data are resources with life cycle like everything else in the real world. The data life cycle and the system development life cycle are shown in Fig. 2. The data life cycle (DLC) describes the management process of data resources. The system development life cycle (SDLC) describes the process of establishing information systems for system analysts, software engineers, programmers, and end-users. They are closely related, because data planning, specification, definition, and development implementation activities are parts of the SDLC.

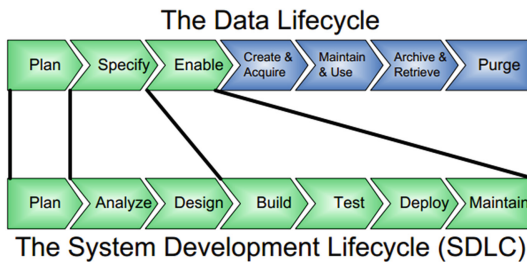


Fig. 2. The data life cycle and the system development life cycle [6]

2.3 Data Quality

Data quality (DQ) is one aspect of the quality category. It is a concept of the quality level of data. The definition of data quality “data are fit for their intended uses” is widely accepted and quoted [7]. Data quality is also known as information quality (IQ). They are both concerned about quality issues. But strictly speaking, there are two differences between them. DQ describes the characterization of quality characteristics of data from their objective features, and emphasizes the data that meet the objective facts. While from the aspect of data bearing semantic and pragmatic features of information, IQ emphasizes data bring people feeling, cognition and recognized degree, and focuses on depicting the data from the subjective aspects of the quality characteristics.

It can be seen that the quality of data can be considered low, if the data are not suitable for its intended use. In particular, there are special standards and methods for measuring quality of graphics, image and video. Therefore, the quality of data is usually referred to structured data, such as text and numbers, and does not involve unstructured multimedia data.

2.4 Data Standards

Data standards are agreements between parties on the definitions of common business terms and the ways those terms are named and represented in data [8]. A standard incorporates a set of rules that describe how data objects are stored, exchanged, formatted, or presented and encompasses the rules by which information is shared. This includes [8]:

- (1) the identification and definition of common business terms.
- (2) the determination of which data objects will be shared.
- (3) the list of data elements composing those data objects.
- (4) data element naming, format/structure, and presentation rules.

2.5 Compliance Testing

Compliance testing, also known as conformance testing, is a nonfunctional testing technique which is done to validate whether the system developed meets the organization’s prescribed standards or not [9].

Objectives of compliance testing should include [9]:

- (1) determining that the development and maintenance process meets the prescribed methodology.
- (2) ensures whether the deliverables of each phase of the development, meets the standards, procedures and guidelines.
- (3) evaluate the documentation of the project to check for completeness and reasonableness.

2.6 Compliance Testing on Data Standards

Compliance testing on data standards can be defined as validating whether the deliverables of each phase of the software development meets the organization’s prescribed data standards or not, such as Chinese and English terms, data types, data formats, data units, data ranges.

It is an inevitable product of the development of the big data industry economy in the current society. It aims to provide reliable data quality assurance for governmental data sharing and commercial data marketing.

3 Models of Compliance Testing on Data Standards

3.1 H Testing Model of Data Quality

This paper proposes a H testing model of data quality, as shown in Fig. 3.

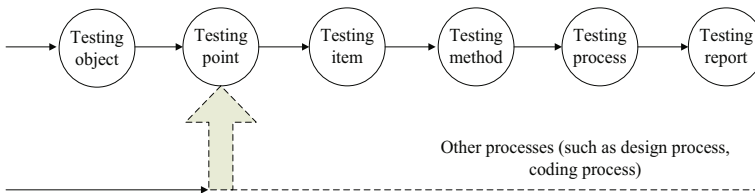


Fig. 3. H testing model of data quality

Data quality testing is a systematic project, usually involving six elements: testing object, testing point, testing item, testing method, testing flow and testing report.

(1) Testing object

Data is usually recorded in physical or electronic form. As a result, testing objects fall into two main categories: document data and electronic data.

The file data category includes documents, contracts, operation orders, project documents, records, faxes, financial reports, development plans, emergency plans, daily data generated by the undergraduate office, and various foreign inflow documents.

Electronic data includes system management, system files, documents, reports, technical scheme and work records, forms, configuration files, topology, system information table, user manuals, database, data operation and statistical data, the development process of the source code.

(2) Testing point

A testing point is a key monitoring point for data quality which is implemented in the data life cycle. Data quality tests should first select the best test point.

(3) Testing item

The testing item is the property of the testing object. For example, the testing items of data content include Chinese and English terms, data types, data formats, data units, data range, etc.

(4) Testing method

Select the appropriate testing method according to testing object and their attributes.

(5) Testing flow

It is important to use advanced standards, methods and tools for data quality testing. However, successful data quality testing cannot be separated from the management of the organization and process of the testing. The flow management of data quality testing is an important guarantee for successful testing.

(6) Testing report

In the current field of laboratory accreditation classification (CNAS-AL06: 2015), the “software product and information security product” category has not used data as a detection product.

Data quality testing is an independent process that runs throughout the system development life cycle and is concurrent with other processes (such as design processes, coding processes, etc.).

Data quality testing should be prepared as soon as possible and implemented as soon as possible.

Data quality testing should be conducted at different levels according to testing point. Different levels of testing activity can be done in a certain order, but may be repeated.

3.2 Testing Object Model of Data Quality

This paper proposes a testing object model of data quality, as shown in Fig. 4. This model reveals the deliverables as testing object of each phase of the data lifecycle.

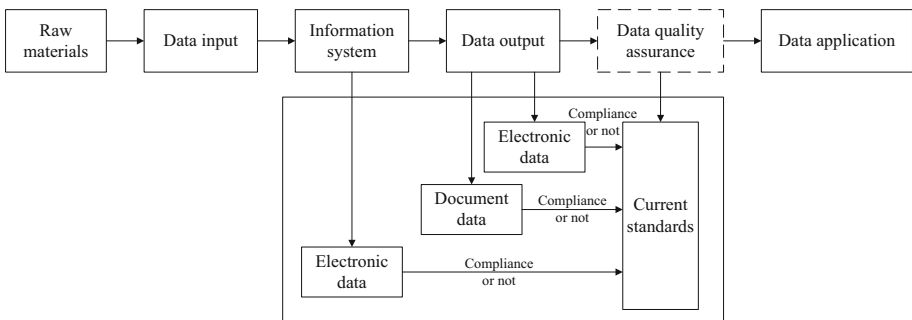


Fig. 4. Testing object model of data quality

For example, data dictionary as electronic data in computer systems, electronic data and document data as data output.

3.3 Testing Process Model of Data Quality

This paper proposes a testing object model of data quality, as shown in Fig. 5. Data quality testing process is generally divided into the establishment of testing plans, the establishment of the current standard library, data collection and analysis, testing implementation, testing logs, testing reports and other stages. The tasks, inputs and outputs of each phase are clearly defined so as to control and configure the entire testing process.

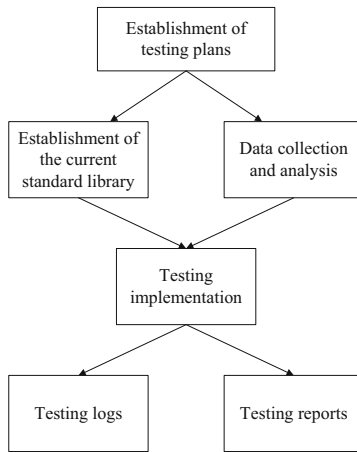


Fig. 5. Test object model of data quality

4 Document Data Quality Assurance

This paper proposes a testing framework of document data quality assurance, as shown in Fig. 6. It is constructed according to the modular architecture, which mainly consists of four parts: testing system, simulation training and evaluation system, security assurance system and hardware and software infrastructure.

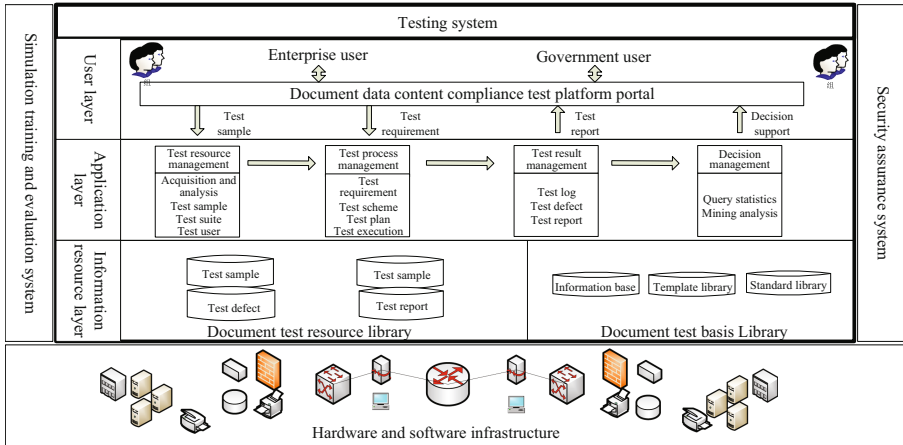


Fig. 6. Testing Framework of document data quality assurance

5 Conclusions

Compliance testing on data standard for data quality assurance is an inevitable product of the development of the big data industry economy in the current society. This paper helps well understanding of its definitions, models and applications. It can be exploited in a wide range of potential applications to make data high available. Still, it remains challenges to refine testing models into a unified view of complex and dynamic real world.

Acknowledgement. This work was supported by the Presidential Foundation of China National Institute of Standardization (Grant No. 532016Y-4689, and No. 712016Y-4942) and the National Key Research and Development Project (Grant No. 2016YFF0202103).

References

1. Tadao Umesao. https://en.wikipedia.org/wiki/Tadao_Umesao
2. Zheng, X.: Big data for social transportation. *IEEE Trans. Intell. Transp. Syst.* **17**(3), 620–630 (2016)
3. Copeland, L.: A practitioner's guide to software test design. Artech House Publishers, Massachusetts (2003)
4. Software. <https://en.wikipedia.org/wiki/Software>
5. Software testing. https://en.wikipedia.org/wiki/Software_testing
6. DAMA International: The DAMA guide to the data management body of knowledge (DAMA-DMBOK), 1st edn. Technics Publications, New Jersey (2010)
7. Data quality. https://en.wikipedia.org/wiki/Data_quality
8. Loshin, D.: The practitioner's guide to data quality improvement. Morgan Kaufmann Series on Business Intelligence, 1st edn. Morgan Kaufmann Publishers, San Francisco (2011)
9. <http://www.softwaretestinghelp.com/what-is-conformance-testing/>

Telecom Big Data Based Electromagnetic Wave Research Under Haze and Rainstorm

Xinzhou Cheng, Lexi Xu (✉), Tao Zhang, Chen Cheng, Weiwei Chen,
Heng Zhang, Yuwei Jia, and Haina Ye

China Unicom Network Technology Research Institute, Beijing 100048, China
{chengxz11, xulx29}@chinaunicom.cn

Abstract. Mobile cellular networks are experiencing fast development recently. The electromagnetic wave is the key factor to impact the cell coverage and the performance of mobile cellular networks. In this paper, we research the electromagnetic wave under the haze and the rainstorm. Initially, we study the basic theory of electromagnetic wave and research its propagation characteristic. Then, we analyze the theoretical impact of electromagnetic wave under the haze and the rainstorm. In order to verify the theoretical analysis, we employ the telecom big data in a city of China to analyze the cell coverage and the service access performance. Results show the haze has little impact on the cell coverage and the access performance, whilst the rainstorm degrades the cell coverage.

Keywords: Electromagnetic wave · Telecom big data · Haze · Rainstorm

1 Introduction

In the past decade, mobile cellular networks have experienced fast development [1–3]. Telecom operators deployed 2G GSM, 3G WCDMA and CDMA2000 networks worldwide [4, 5]. Due to users' requirements and service diversity, 4G LTE networks are deployed worldwide [6–8].

The transmission of electromagnetic wave plays an important role and impacts the mobile networks, including the cell coverage, the networks performance etc. [9, 10]. Different weather conditions have different influences on the transmission of electromagnetic wave, thus impacting the mobile networks. From the perspective of system operation and networks optimization, telecom operators should be aware of the detailed impact of mobile cellular networks under different weather conditions [11, 12].

This paper researches the electromagnetic wave under two weather conditions, including haze and rainstorm. Specifically, we study the definition and classification of electromagnetic wave. Then, we research the propagation characteristics. Furthermore, this paper theoretically analyzes the impact of haze and rainstorm. Finally, we collect the realistic 3G/4G data to analyze the cell coverage, the service access performance in Beijing under the haze condition and the rainstorm condition.

2 The Electromagnetic Theory

2.1 Definition and Classification of Electromagnetic Wave

The electromagnetic wave is a self-propagating transverse oscillating wave in the electric and magnetic fields. Figure 1 shows the spread of the electromagnetic wave (from the left to the right direction) [13]. The electric and magnetic fields in EMR wave are always in phase and at 90° to each other.

The electromagnetic wave varies in size, from long radio wave (the size of buildings) to short gamma rays (shorter than atom nuclei). Its frequency is inversely proportional to wavelength, as shown in (1):

$$v = f\lambda \tag{1}$$

where v is the speed of the wave, f is the frequency and λ is the wavelength [13].

A photon has an energy proportional to its frequency by (2):

$$E = hf = \frac{hc}{\lambda} \tag{2}$$

where E is the energy, h is Planck's constant λ and c is the speed of light.

The momentum p of a photon is also proportional to its frequency, as well as inversely proportional to its wavelength, as shown in (3):

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda} \tag{3}$$

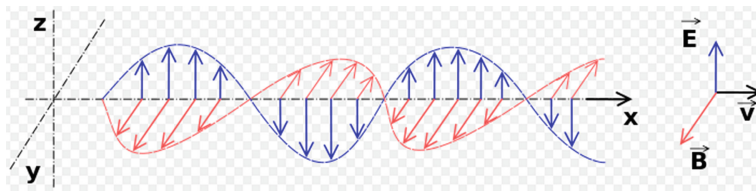


Fig. 1. Spread of electromagnetic wave

Table 1. Classification of electromagnetic wave

Class	Frequency	Wavelength
Low frequency (LF)	30–300 kHz	10 ⁴ –10 ³ m (10–1 km)
Medium frequency (MF)	300–3000 kHz	10 ³ –10 ² m (1000–100 m)
High frequency	3–30 MHz	10 ² –10 m (100–10 m)
Very high frequency (VHF)	30–300 MHz	10–1 m
Ultrahigh frequency (UHF)	300–3000 MHz	1–0.1 m
Ultrahigh frequency (SHF)	3–30 GHz	10–1 cm
Extremely high frequency (EHF)	30–300 GHz	10–1 mm
Tremendously High frequency (THF)	300–3000 GHz	1–0.1 mm
Visible Light	100–10000THz	3 × 10 ⁻³ – 3 × 10 ⁻⁵ mm

The classification of electromagnetic wave is as Table 1. The electromagnetic wave of mobile communication is Ultrahigh frequency, as shown in Table 2.

Table 2. Electromagnetic wave for telecom operators in China

Operator	Standard	Frequency (MHz)	Upload frequency (MHz)	Download frequency (MHz)	Wavelength (cm)
China Unicom	GSM900 (2G)	6	909–915	954–960	29.80–31.47
	DCS1800 (2G)	10	1745–1755	1840–1850	15.46–16.40
	WCDMA (3G)	15	1940–1955	2130–2145	13.33–14.75
	TD-LTE (4G)	40	2300–2320	2300–2320	12.33–12.44
			2555–2575	2555–2575	11.11–11.20
FDD-LTE (4G)	10	1755–1765	1850–1860	15.38–16.30	
China Mobile	GSM900 (2G)	19	890–990	935–954	29.99–32.15
	DCS1800 (2G)	15	1710–1725	1805–1820	15.72–16.73
	TD-SCDMA (3G)	15	2010–2025	2010–2025	14.13–14.23
	TD-LTE (4G)	130	1880–1890	1880–1890	15.14–15.22
			2320–2370	2320–2370	12.07–12.33
			2575–2635	2575–2635	10.86–11.11
China Telecom	CDMA (2G)	15	825–840	870–885	32.33–34.68
	CDMA2000 (3G)	15	1920–1935	2110–2125	13.46–14.90
	TD-LTE (4G)	40	2370–2390	2370–2390	11.97–12.07
			2635–2655	2635–2655	10.78–10.86
	FDD-LTE (4G)	15	1765–1780	1860–1875	15.26–16.21

2.2 Propagation Characteristic of Electromagnetic Wave

The propagation characteristics of electromagnetic wave include:

- (1) Propagation loss: Propagation loss and dispersion caused by variation of signal propagation distance.
- (2) Shadow fading: Propagation of wave in a transmitting environment, tall buildings, and prominent obstacles that cause the decline of electromagnetic wave.
- (3) Multipath fading: It contains reflection, diffraction and scattering.
- (4) Doppler effect: The expansion of frequency domain because of the movement/mobility of travelling carriage in high speed [14,15].

3 Theoretical Analysis of Impact of Haze and Rainstorm

Haze, which is also called PM_{2.5} (Particulate Matter 2.5), is a fine particulate matter suspended in the air. Its diameter is less than 2.5 μ . The electromagnetic wave length of China Unicom ranges of 11–32 cm, more than 40 thousand times the diameter of PM_{2.5}. As described in Sect. 2.2, haze has little effect on the propagation loss, shadow fading, Doppler effect of mobile communication, but may has effect on multipath fading. Then, we will analyze it in details.

The three parts of multipath fading and the trigger conditions are as follows [13]:

- (1) Reflection: The block is smooth and larger than the transmission wavelength.

$$R \gg \lambda \tag{4}$$

In (4), R represents the diameter of stumbling block, λ represents the diameter of electromagnetic wave.

- (2) Diffraction: the edge of the block is sharp and much larger than the transmission wavelength, as shown in (5).

$$R \gg \lambda \tag{5}$$

- (3) Scattering: The electromagnetic wave deviates from the original direction through the uneven medium. Trigger conditions are rough surfaces, small objects or other irregular objects. The scattering method varies with the relative relation between electromagnetic wave length and atmospheric molecular diameter and aerosol particle size, divided into mie scattering, non-selective scattering and Rayleigh scattering.

As the size of PM2.5 particle is 40000 times smaller than electromagnetic wave for telecom operators, it has little impact on the reflection and diffraction of electromagnetic wave. However, PM2.5 may affect the electromagnetic scattering. The trigger conditions of three kinds of electromagnetic scattering are as follows:

- (a) Mie scattering:

$$R \cong \lambda \tag{6}$$

In (6), R represents the diameter of stumbling block while λ represents the diameter of electromagnetic wave.

- (b) Non-selective scattering:

$$R \gg \lambda \tag{7}$$

- (c) Rayleigh scattering:

$$R < \frac{\lambda}{10} \tag{8}$$

The length of electromagnetic wave of mobile networks is 1–35 cm, which is much larger than the PM2.5 size. Hence, the Mie scattering and non-selective scattering is impacted slightly. The intensity of scattering wave is inversely proportional to the fourth power of wave length. Therefore, the impact to mobile networks is insignificant.

The diameter of rainstorm is 0.5–5.5 mm, which is much larger than the size of haze. Hence, rainstorm also impacts Rayleigh scattering. As the diameter of rainstorm

is 1000 times larger than haze, the intensity is 10^{12} times stronger than the impact of haze.

In addition, the road filled with water may causes reflection, which will impact the electromagnetic wave of mobile communication as well.

4 Experiments and Results

Telecom operators can collect telecom big data of mobile networks [16, 17]. Typical telecom big data includes the measurement report (MR) of users, key performance indicators of mobile networks, service condition etc. [18–20]. Section 4 presents our experiments of analyzing different weather condition's effect for electromagnetic wave of mobile networks. The process follows the typical experiment process in mobile networks, including the data collection stage, the data analysis stage for different evaluation indicators, as well as the results discussion stage [21, 22].

4.1 Experimental Results of Haze's Effect on Electromagnetic Wave

We analyze the telecom big data of 3G WCDMA outdoor macro-cells in Beijing between December 15, 2015 and December 22, 2015. During this period, the air pollution was becoming worse gradually. The experimental area consists of over 6100 macro-cells, which are located in Xicheng District and Haidian District.

We compare both the coverage and the accessibility performance between good weather condition and haze condition. From Table 3, as the air pollution became worse, the overall coverage rate did not change significantly. We can conclude that the haze weather condition has no significant effect on the quality of wireless coverage. Additionally, based on the weak coverage rate geographical visualization in Fig. 2, we can get the same conclusion that the proportion of weak coverage in each cell was not significantly deteriorated as the air pollution intensifies.

Table 3. Information of coverage rate at different air quality conditions

Date	Weather	Air quality	PM 2.5 concentration	RSCP > -100 dBm samples	Total samples	Coverage rate
12.15	Sunny	Good	28	17041092	17830878	95.57%
12.16	Sunny	Good	23	16964881	17739950	95.63%
12.17	Sunny	Moderate	84	16779586	17522143	95.76%
12.18	Sunny	Moderate	80	16964524	17726104	95.70%
12.19	Haze	Unhealthy	168	14684398	15330071	95.79%
12.20	Haze	Very Unhealthy	249	14729609	15375202	95.80%
12.21	Haze	Very Unhealthy	280	16808394	17553258	95.76%
12.22	Haze	Hazardous	347	16938851	17706373	95.67%

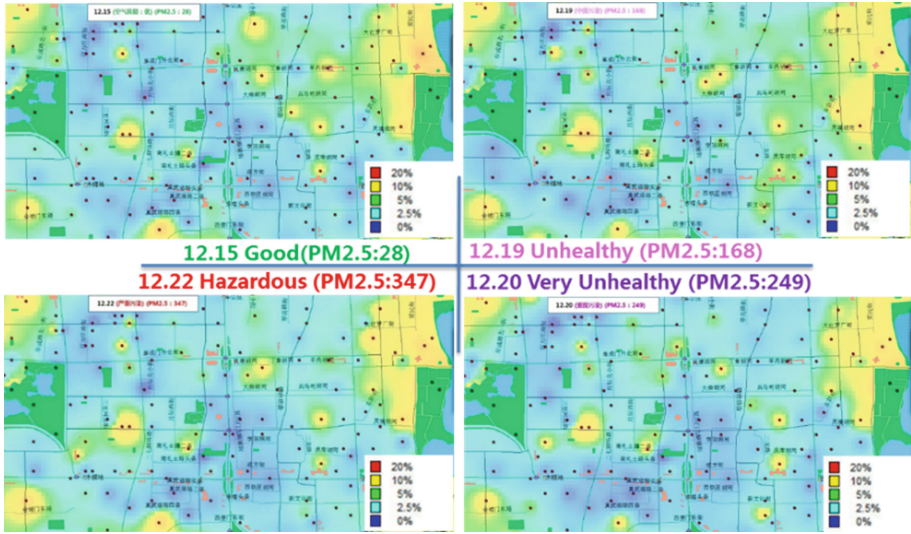


Fig. 2. Weak coverage rate geographical visualization at different air quality condition

Figure 3 shows that the Radio Access Bearer (RAB) establishing success rate did not decrease as air quality decreased. From above results, we can see that the coverage, accessibility of mobile networks were not affected by the haze weather condition.

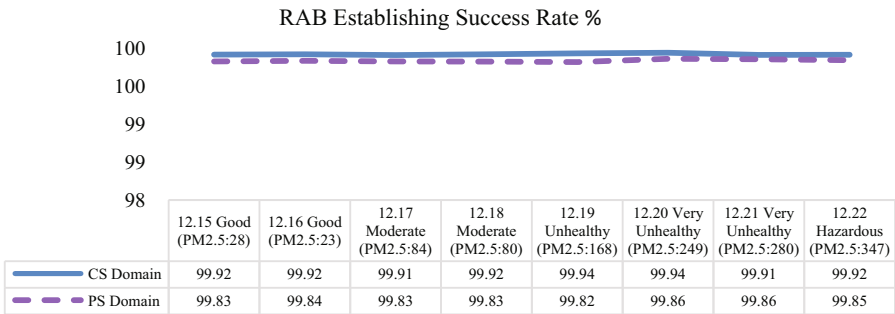


Fig. 3. Result of RAB establishing success rate

4.2 Experimental Results of Rainstorm’s Effect on Electromagnetic Wave

In Sect. 4.2, we analyze the 4G MR of over 2800 outdoor LTE macro-cells in Daxing District of Beijing between June 15, 2017 and June 24, 2017, thus verifying the actual effect of rainstorm on the transmission of electromagnetic wave [23].

Table 4 and Fig. 4 show that as the rain got heavier, the overall coverage rate had an obvious drop, especially on June 22 and June 23. Because of the heavy rain, the

Table 4. Horizontal comparison of coverage rate at different weather conditions

Date	Weather	RSRP > -110 dBm samples	Total samples	Coverage rate
6.15	Sunny	468171614	529158727	88.47%
6.16	Cloudy	470003196	530791848	88.55%
6.19	Cloudy	418457313	473350960	88.40%
6.21	Thunder rain	462485371	523875863	88.28%
6.22	Moderate rain	446711770	513630026	86.97%
6.23	Heavy rain	300126992	352168132	85.22%
6.24	Cloudy	429247922	490790139	87.46%

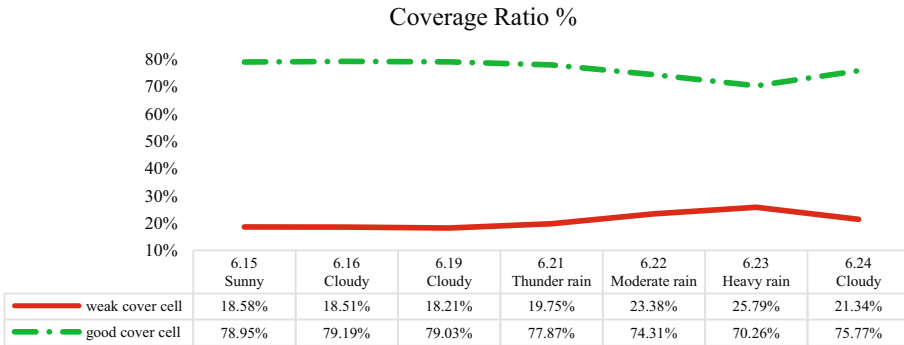


Fig. 4. Result of coverage ratio (good coverage cells, weak coverage cells)

coverage rate fell by over 3%. Meanwhile, the proportion of weak coverage cell increased greatly during the heavy rain day on June 23.

From the experimental results, we can conclude that the rain has a great influence on the transmission of the electromagnetic wave of mobile cellular networks. Due to the absorption and scattering of electromagnetic wave by rain drops, the transmission signal will attenuate.

5 Conclusions

This paper investigates the electromagnetic wave under the haze condition and the rainstorm condition. Firstly, this paper studies the basic theory of electromagnetic wave and researches its propagation characteristic. Secondly, we theoretically analyze the impact of electromagnetic wave under haze and rainstorm. Thirdly, we collect the 3G WCDMA data and 4G LTE data in Beijing. Finally, we use the collected data to analyze the cell coverage and the service access performance under haze and rainstorm. Experimental results show the haze has little impact on the cell coverage and service access performance. Results also show the rainstorm degrades the cell coverage.

References

1. Xu, L., Chen, Y., Gao, Y., Cuthbert, L.: A Self-optimizing load balancing scheme for fixed relay cellular networks. In: 2th IET International Conference on Communication Technology and Application, pp. 306–311. IET Press, Beijing (2011)
2. Cao, Y., Sun, Z., Wang, N., Riaz, M., Cruickshank, H., Liu, X.: Geographic-based Spray-and-Relay (GSaR): an efficient routing scheme for DTNs. *IEEE Trans. Veh. Technol.* **64**(4), 1548–1564 (2015)
3. Ge, C., Wang, N., Skillman, S., Foster, G., Cao, Y.: QoE driven DASH video caching and adaptation at 5G mobile edge. In: 3rd ACM Conference on Information-Centric Networking, pp. 237–242. ACM Press, Kyoto (2017)
4. Xu, L., Cheng, X., Chen, Y., Chao, K., Liu, D., Xing, H.: Self-optimised coordinated traffic shifting scheme for LTE cellular systems. In: 1st EAI International Conference on Self-Organizing Networks, pp. 67–75. Springer press, Beijing (2015)
5. Cao, Y., Wang, N., Sun, Z., Cruickshank, H.: A reliable and efficient encounter-based routing framework for delay/disruption tolerant networks. *IEEE Sens. J.* **15**(7), 4004–4018 (2015)
6. Cao, Y., Sun, Z., et al.: Routing in delay/disruption tolerant networks: a taxonomy, survey and challenges. *IEEE Commun. Surv. Tutorials* **5**(12), 654–677 (2013)
7. Xu, L., Chen, Y., Chai, K. K., Luan, Y., Liu, D.: Cooperative mobility load balancing in relay cellular networks. In: 2nd IEEE/CIC International Conference on Communication in China, pp. 141–146. IEEE press, Xi’an (2013)
8. Cao, Y., Wang, T., Kaiwartya, O., et al.: An EV charging management system concerning drivers’ trip duration and mobility uncertainty. *IEEE Trans. Syst. Man Cybern. Syst.* **99**, 1 (2016)
9. Cheng, X., Xu, L., Zhang, T., et al.: A novel big data based telecom operation architecture. In: 1st International Conference on Signal and Information Processing, Networking and Computers, pp. 385–396. CRC Press Taylor & Francis Group, Beijing (2015)
10. Fan, X., Lu, J., et al.: Analysis of key wireless factors affecting on VoLTE user-perceived performance based on measured data. *Mobile Commun.* **41**(14), 5–12 (2017)
11. Zhang, X., Wang, N., Cao, Y., et al.: A stochastic analytical modelling framework on ISP-P2P collaborations in multi-domain environments. *IEEE Syst. J.* **99**, 1 (2016)
12. Xu, L., Luan, Y., Cheng, X., Xing, H., Liu, Y., et al.: Self-optimised joint traffic offloading in heterogeneous cellular networks. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 263–267. IEEE press, Qingdao (2016)
13. Wikipedia. https://en.wikipedia.org/wiki/Electromagnetic_radiation. Accessed 19 Jul 2017
14. Cao, Y., Wei, K., Min, G., et al.: A geographic multicopy routing scheme for DTNs with heterogeneous mobility. *IEEE Syst. J.* **99**, 1 (2016)
15. Xu, L., Chen, Y., Yue, et al.: User relay assisted traffic shifting in LTE-advanced systems. In: 77th IEEE Vehicular Technology Conference, pp. 1–6. IEEE Press, Dresden (2013)
16. Xu, L., Luan, Y., et al.: WCDMA data based LTE site selection scheme in LTE deployment. In: 1st International Conference on Signal and Information Processing, Networking and Computers, pp. 249–260. CRC Press Taylor & Francis Group, Beijing (2015)
17. Wang, W., Xu, L., Zhang, Y., Zhong, J.: A novel cell-level resource allocation scheme for OFDMA system. In: 1st International Conference on Communications and Mobile Computing, pp. 287–292. IEEE Press, Kunming (2009)
18. Xu L., Chen, Y., et al.: Cooperative load balancing for OFDMA cellular networks. In: 18th European Wireless Conference, pp. 1–7. VDE Verlag GMBH, Poznan (2012)

19. Xu, L., Cheng, X., Liu, Y., et al.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: 16th IEEE International Conference on Communication Technology, pp. 806–812. IEEE Press, Hangzhou (2015)
20. Kaiwartya, O., Abdullah, A., Cao, Y., et al.: T-MQM: testbed based multi-metric quality measurement for sensor deployment for wireless sensor networks—a case study. *IEEE Sens. J.* **16**(23), 8649–8664 (2016)
21. Xu, L., Chen, Y., Gao, Y.: Self-organizing load balancing for relay based cellular networks. In: 10th IEEE International Conference on Computer and Information Technology, pp. 791–796. IEEE Press, Bradford (2010)
22. Miao, Y., Sun, Z., Wang, N., et al.: Time efficient data collection with mobile sink and vMIMO technique in wireless sensor networks. *IEEE Syst. J.* **99**, 1 (2016)
23. Liu, Y., Xu, L., Chen, Y., Fan, Y., Xu, B., Nie, J.: A novel power control mechanism based on interference estimation in LTE cellular networks. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 397–401. IEEE Press, Qingdao (2016)

Analysis and Optimization of Video Fluency Based on Big Data

Mingjun Mu^(✉), Yuwei Jia, Weiwei Chen, and Yongfeng Wang

Network Technology Research Institute, China United Network
Communications Corporation Limited, Beijing, China
mumj5@chinaunicom.cn

Abstract. In the age of 4G, mobile video has become the key development strategy of Internet companies and telecom operators. Because the user perception of video service has a direct impact on the reputation of the operators' networks, the evaluation of video traffic quality becomes one of the most important tasks for operators. Video is different from other types of service and it needs to inspect every packet to analyze video fluency which is a challenge for both the analytic algorithm and the server. This paper proposes a pre evaluation algorithm to evaluate video fluency which can dramatically reduce the inspection overhead as well as the server pressure.

Keywords: Pre evaluation algorithm · Video evaluation · Big data

1 Introduction

With the rapid development of mobile communication and video technology, video transmission network extends from broadcast TV network to information network, and mobile video service will gradually become a very valuable business in the communication and Internet industry. The prospect of video service will be beyond our imagination in the future, and it will run through all walks of people's life. We will be gradually entering the video society, and accordingly video service will become the mainstream of people's life experience. Faced with the huge development space of video service, telecom operators at home and abroad are turning video into their key strategy.

At present, the traffic of mobile video service is more than half of that of the 4G mobile service. The support capability of mobile network directly affects the quality of the mobile video and the user perception, therefore, the quality of video service is an important measure to evaluate the network capability and the competitiveness of operators in the 4G era. Researchers such as Cheng et al. [1] and Wang et al. [2] have proposed relevant metrics to evaluate video quality and user perception, including success rate of playing, initial playout delay, times of lag and duration of lag. Times of lag and duration of lag belong to the fluency assessment index, which are significantly different from other types of business. Zhang et al. [3] puts forward the accumulation algorithm to calculate the times of lag and duration of lag. Its core idea is to combine the relevant signaling with every video packet to restore the user's download process as much as possible, thus inferring the fluency index of videos.

As the video traffic is improving rapidly, the server is getting more and more stressed while analyzing video traffic and outputting relevant evaluation index. This will lead to a phenomenon that the demo of a product can work correctly but correct results can't be outputted in the practical network. In this paper, a pre evaluation scheme is proposed to reduce calculation amount and server pressure by judging whether the video is smooth or not and then calculating indices of the video's lag.

2 Calculation of Lag Index

The calculation of video fluency index is the most difficult during the video evaluation. It can be calculated from test data, Internet log data and operators' core network data. Video fluency index can be accurately calculated from test data, but the amount of test data is smaller than others and it also consumes a lot of manpower, material resources, time and other resources. It's hard to cooperate with every video site even the Internet log data can provide accurate indicators and its volume is relatively large. Core network data belongs to the operator's own data which is easy to implement and can provide large amounts of data but requires DPI tools and mass calculation. Considering the above three methods, operators mainly use the core network data for video evaluation.

Based on core network data, the basic scheme for computing video lag is as follows.

We assume that the actual download rate is V_i , the download period of one packet is T_i , the basic download rate required by video is V_a , the number of packets in video is M . For $n \in (1, 2, 3 \dots M)$

When

$$\sum_0^n (v_i * T_i) < v_a * \sum_0^n T_i \tag{1}$$

Lag will appear in the video, else the video can be viewed as smooth.

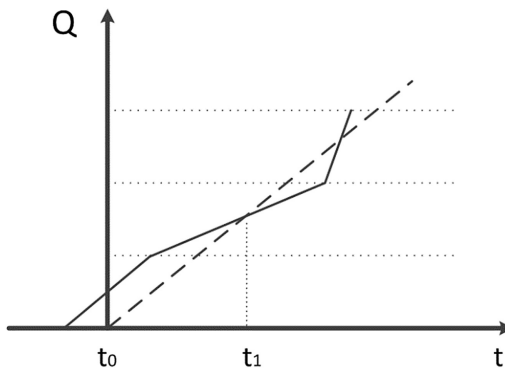


Fig. 1. Method of calculating lag.

The horizontal axis represents the download time, and the vertical axis represents the download data. The solid line represents practical download data, while the dotted line represents the required download data which should meet the basic requirement for smoothly video playing. In Fig. 1, it shows that lag will appear at the moment of t_1 .

The video evaluation index calculated by this method is basically consistent with the actual situation, but every packet is to be calculated, the calculation amount of this method is enormous. The data size of a standard definition video is about 200 MB, while a TCP packet only has 1500 B data, just one video’s fluency assessment requires up to 140000 times of packet count. So the amount of resources needed to evaluate video business is enormous for a city’s network.

3 Pre Evaluation Algorithm

In the age of 4G, networks provided by operators own the capability of high download rate which can reach 100 Mbps in theory, especially the download rate is as high as 900 Mbps in indoor and outdoor scenario on the China Unicom’s 4G+ test network. More than half of the video service meets the requirement of high-definition video from actual statistics in Fig. 2. It is necessary to take the evaluation method mentioned in the previous section when lag appears in a video. If there is no lag, it will save a great deal of computational resources by an efficient method rather than the traditional evaluation method.

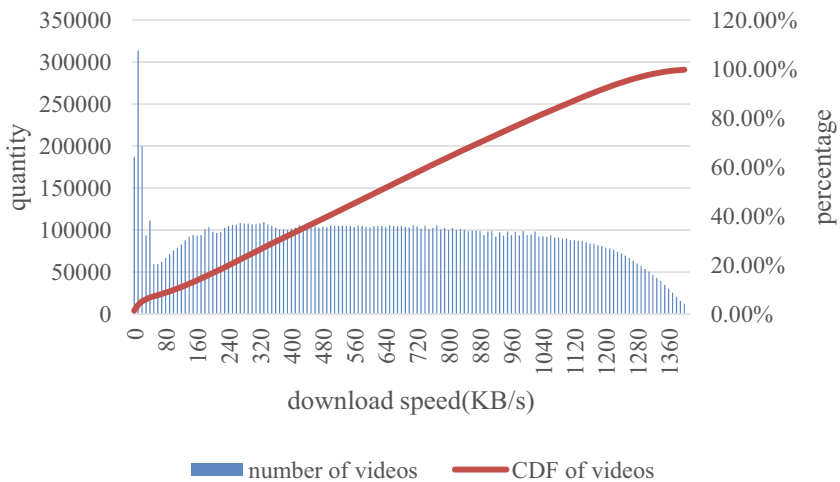


Fig. 2. The distribution of video download rate in practice.

The current streaming protocol is to separate video into segments, the data amount of each segment is about 10 MB or less, the segment will not be downloaded until the previous one is nearly finished. We can come to a conclusion for every segment first of all, if there is no lag, the evaluation index will be outputted directly. If lag exists in one segment, the traditional evaluation method will be used to get the index.

Here a method is introduced to judge whether there is lag or not in a segment. For simplicity, if the download time of a segment (t_d) is less than the viewing time (t_w), we can roughly conclude that there is no lag in this segment. Ideally, the actual download rate is always higher than the required download rate like the solid line Q_1 and the dotted line Q_0 , but we can't rule out what's shown by curve Q_2 because we have no idea about the details of each packet (Fig. 3).

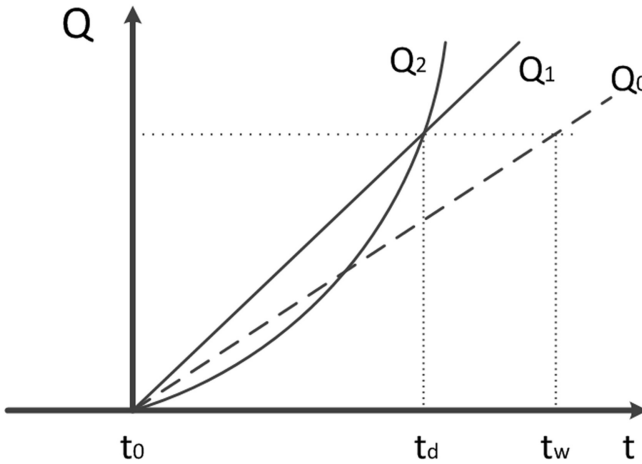


Fig. 3. Unsatisfactory of the judgment by average download rate.

In the process of playing the video, the download data will be stored in a buffer before it is played. In order to ensure the fluency of video, the media player typically sets a minimum threshold for the buffer, when the data in the buffer is smaller than this threshold, the next segment will be requested to be downloaded.

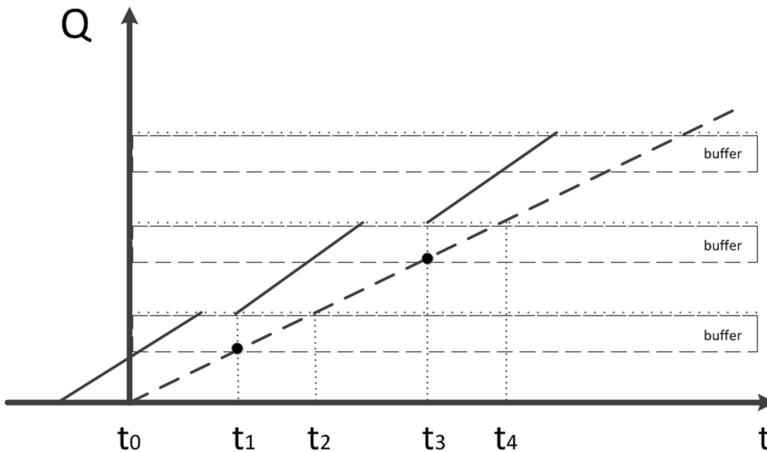


Fig. 4. Three stages of video download

The Fig. 4 shows how the video is downloaded separately. At the moment of t_1 and t_3 , the remaining data is insufficient which triggers the minimum buffer threshold, and then the next segment is requested to be downloaded. The minimum buffer is available for normal viewing during the period between t_1 and t_2 , a whole segment can play normally and smoothly during the period between t_2 and t_4 .

If the video can be normally viewed with no lag, there will be a relationship as

$$(t_3 - t_1) - (t_2 - t_1) + (t_4 - t_3) = t_4 - t_2 \tag{2}$$

Assume that ΔQ is the minimum buffer threshold, Q_i is the i th segment of the video, T_i is the play duration of the i th segment, C_i is the rate of i th segment, t_i is the moment that the i th segment is requested, we have

$$(t_{i+1} - t_i) - \frac{\Delta Q}{C_{i-1}} + \frac{\Delta Q}{C_i} = T_i \tag{3}$$

Because

$$C_i = \frac{Q_i}{T_i} \tag{4}$$

We finally have

$$t_{i+1} - t_i - \frac{\Delta Q}{Q_{i-1}/T_{i-1}} + \frac{\Delta Q}{Q_i/T_i} = T_i \tag{5}$$

The above parameters can be obtained without detailed information of each packet. If the relevant parameters obtained by inspecting signaling satisfy formula (5), it shows that this segment of video has been played normally and smoothly. And the final conclusion can be drawn directly that there is no lag in the segment and analyzing each packet deeply is unnecessary.

To show the advantage of the Pre Evaluation Algorithm compared with the traditional cumulative algorithm, simulation is realized according to different video standards and network environment (Fig. 5).

The simulation result shows that the number of packets required by the traditional cumulative algorithm in any network condition is constant. However, the number of packets required by the Pre Evaluation Algorithm decreases gradually as the network condition improves. With the development of mobile network technology, the network environment is getting better, and the Pre Evaluation Algorithm will be of greater value. In brief, the pre evaluation algorithm can reduce the complexity of the video evaluation system and reduce the pressure of the evaluation server.

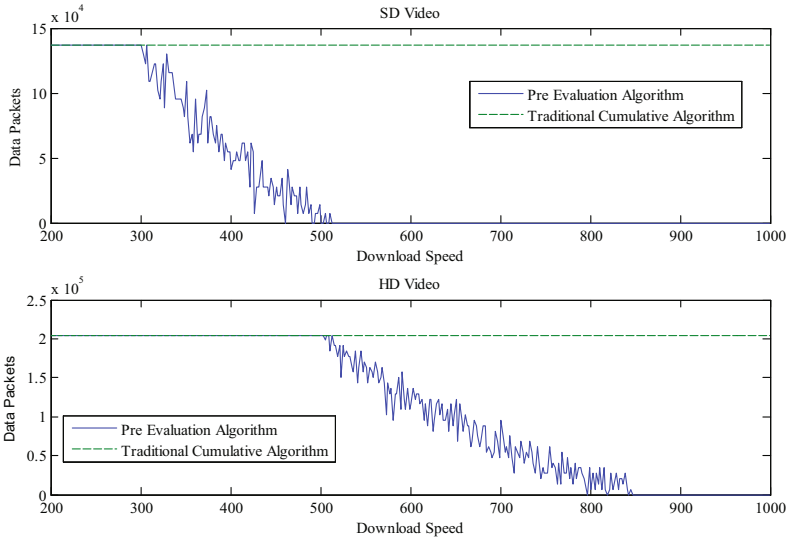


Fig. 5. The difference between the two algorithms in the calculation of packets.

4 Conclusion

In the era of video traffic explosion, the evaluation of video quality is very important, operators have invested a lot of resources to build video evaluation platform. It is not applicable for the video evaluation system without considering the algorithm and the server pressure in the background of big data. The pre evaluation algorithm proposed in this paper is an improvement on the traditional evaluation method of video lag, which can greatly reduce the pressure of the evaluation server, especially for the high bandwidth network.

References

1. Cheng, S., Wang, H., Wang, W., Zhan, L.: Research on the quality analysis method of video experience based on deep packet inspection. *Softw. Guide* **172**(02), 141–143 (2017)
2. Wang, X., Liao, Z., Hu, Y.: Research on user perceived evaluation system of 4G video service. *Inf. Commun.* **147**(03), 230–231 (2015)
3. Zhang, P., Cheng, Q., Tang, J.: Optimization on mobile video business perception evaluation based on big data packet-by-packet accumulation algorithm. *Des. Tech. Posts Telecommun.* **492**(02), 10–16 (2017)
4. Geng, Y., Jin, J., Meng, L.: Video QoE assessment considering encoding and transmission damage. *J. Beijing Univ. Posts Telecommun.* **39**(S1), 55–58+71 (2016)
5. Mu, M., Wang, Y., Chen, W.: Telecom big data based investment strategy of value areas. In: *Signal and Information Processing, Networking and Computers*, pp. 281–287 (2015)
6. Ge, C., Wang, N., Skillman, S.: QoE driven DASH video caching and adaptation at 5G mobile edge. In: *ACM ICN IC5G Workshop, Kyoto* (2017)

7. Zhang, X., Wang, N., Cao, Y.: A stochastic analytical modelling framework on ISP-P2P collaborations in multi-domain environments. *IEEE Syst. J.* **99**, 1 (2017)
8. Xu, L., Luan, Y., Cheng, X., Cao, X., Chao, K., Gao, J., Jia, Y., Wang, S.: WCDMA data based LTE site selection scheme in LTE deployment. In: 1st International Conference on Signal and Information Processing, Networking and Computers, pp. 249–260. CRC Press Taylor & Francis Group, Beijing (2015)
9. Xu, L., Chen, Y., Chai, K.K., Luan, Y., Liu, D.: Cooperative mobility load balancing in relay cellular networks. In: 2nd IEEE/CIC International Conference on Communication in China, pp. 141–146. IEEE press, Xi'an (2013)

Clothing Recommendation System Based on Advanced User-Based Collaborative Filtering Algorithm

Yu Liu^{1,2(✉)}, Jingwen Nie^{1,2}, Lexi Xu^{3,4}, Yue Chen²,
and Bingyu Xu²

¹ Beijing University of Posts and Telecommunications, Beijing 100876, China
{liuyul23123, niejingwen}@bupt.edu.cn

{yu.liu, jingwen.nie}@sel4.qmul.ac.uk

² Queen Mary University of London, London E1 4NS, UK

³ China Unicom Network Technology Research Institute, Beijing 100048, China

⁴ National Engineering Lab of Next Generation Internet Broadband Service,
Beijing 100048, China

Abstract. With the development of e-commerce technology, a growing number of people prefer to purchase clothes on the e-commerce websites. Therefore, an effective recommendation system is necessary for customers. User-based Collaborative Filtering (UCF) algorithm is widely utilized to predict the preferences of customers. However, UCF algorithm employs the sparse matrix and the recommendation has low precision. In this paper, an improved recommendation algorithm named Advanced User-based Collaborative Filtering (AUCF) algorithm is proposed and implemented in the clothing recommendation system. The proposed AUCF algorithm introduces user-item linked list, which can overcome the problem of large time complexity. Considering the impact of different popularity of items, AUCF algorithm is capable of publishing the negative influence of popular items, which can increase the recommendation coverage. Experiment results show the AUCF algorithm significantly increases the recommendation coverage and precision.

Keywords: Collaborative filtering algorithm · Online clothing Recommendation system

1 Introduction

In the past couple of decades, mobile communication systems experience fast development [1–3]. Recently, 4G LTE systems are widely deployed worldwide [4–6]. Due to the high data rate of LTE systems, mobile Internet and e-commerce experiences fast development [7–9]. In addition, online stores can provide customers with a large number of products. Hence, online clothing purchasing becomes very popular choice for customers. According to the statistics in 2016, the ratio of transactions of online clothing stores in China reaches nearly seventy percent [10]. However, customers can not accurately find favorite clothes and their time would be wasted when browsing.

Therefore, it's necessary for online clothing stores to recommend appropriate clothes to customers [11].

Collaborative Filtering Recommendation algorithm is utilized to recommend items to target users by employing the groups which have common preferences. User-based Collaborative Filtering (UCF) algorithm is a typical way to make recommendation [12]. The basic principle is to use the historical records to generate Top-K neighbors, and then take the recommendations to the target user based on the Top-K neighbors. However, UCF algorithm has its limitations such as sparseness of matrix, poor efficiency, low precision as well as low recommendation coverage. This paper proposes an Advanced User-based Collaborative Filtering (AUCF) algorithm, which can reduce the sparseness of matrix, and increase the recommendation coverage as well as the efficiency. The experiment results of 10-fold cross-validation in dataset from Tmall.com show that AUCF algorithm increases the coverage and the precision.

2 Advanced User-Based Collaborative Filtering (AUCF)

2.1 Summary of AUCF Algorithm

This paper tries to improve the traditional UCF algorithm via calculating the similarity among users, and we also propose user-item linked list to reduce sparseness of matrix.

This paper denotes $U = \{u_1, u_2, u_3, \dots, u_n\}$ as the set of all n users, $U_k = \{u_1, u_2, u_3, \dots, u_k\}$ as the Top-K neighbors, $I = \{I_1, I_2, I_3, \dots, I_k\}$ as the set of historical items purchased by Top-K neighbors. This paper also denotes $W = \{W_1, W_2, W_3, \dots, W_k\}$ as the set of weights of items for the target user. The similarity between the target user and other users is calculated firstly. U_k are selected from all users based on similarity. We calculate W based on Top-K neighbors. Based on W , we select the Top-N items as recommendation to the target user. Note that the parameter K and N can be adjusted according to application scenario.

2.2 Improvement of Representation

In the typical recommender system based on UCF algorithm, data is presented as an $m \times n$ matrix R , m represents the number of users, n represents the number of items, R_{ij} represents the evaluation value of the i^{th} user to the j^{th} item. In this paper, evaluation value $R_{ij} = 1$ represents the item is purchased, whilst $R_{ij} = 0$ represents the item is not purchased.

Table 1 shows the sparseness of matrix is large. Therefore, we propose the linked list to store whether the user has purchased certain clothes or not. As shown in Fig. 1, each user is a linked list head, and the clothes purchased by the user will be connected to the user in sequence. The clothes that are not purchased by the user will not be stored in the user's user-item linked list.

Table 1. Data sparseness of user-rating-data matrix

User\Item	Item1	Item2	Item3
User1	R_{11}	–	R_{13}
User2	–	R_{22}	–
User3	–	R_{32}	R_{33}

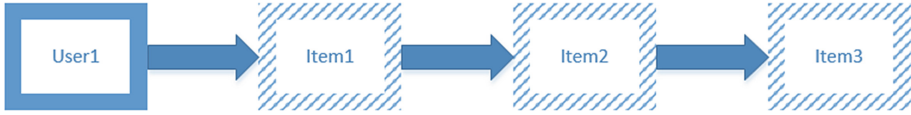


Fig. 1. User-item linked list

2.3 Improvement of Similarity Calculation

This paper tries to improve the calculation method of the similarity among users, and we develop two methods. One method is derived from the vector model and the other method is an improved version of the first one.

2.3.1 Cosine-Based Similarity

Cosine-based similarity (CS) is widely used. This paper assumes v and u are two different customers, which can be represented as vectors. Each dimensionality of the vector represents the purchasing frequency of a certain item (Purchasing frequency indicates the number of certain item purchased by a user).

$$sim(v, u)_{CS} = \cos(\vec{v}, \vec{u}) = \frac{\vec{v} \cdot \vec{u}}{\|\vec{v}\| \|\vec{u}\|} \tag{1}$$

where ‘.’ denotes the vector dot-product operation.

The cosine-based similarity considers the purchasing frequency of different users. However, purchasing frequency is less important than purchasing coverage in this scenario. According to the statistics of on-line store (Tmall.com), around 80% users purchase certain clothes once or not. Therefore, most users only purchase certain clothes once, and the purchasing frequency is 1 or 0 under 80% scenarios [7]. From the analysis above, purchasing frequency is not a key factor to analyze the purchase behavior. The cosine-based similarity method can be simplified by using the *intersection principle* [13]. I_v is the item set, which consists of all items purchased by v . I_u is the item set purchased by u . Based on the number of items (both v and u purchased), the modified cosine-based similarity (MCS) between v and u can be calculated as (2).

$$sim(v, u)_{MCS} = \frac{|I_v \cap I_u|}{\sqrt{|I_v| |I_u|}} \tag{2}$$

where $|I_v|$ denotes the number of elements in set I_v . By using (2), the impact of purchasing frequency can be overcome.

2.3.2 Inverse Item Frequency Similarity

The MCS method is effective to calculate the similarity. However, the MCS method employs the same weight for different items. To our knowledge, popular items have less influence on the similarity than unpopular items. It is because most users will purchase popular items. Therefore, compared with unpopular items, popular items should have lower weight.

The recommendation coverage reflects the range of items recommended by the similarity method. This paper improves the recommendation coverage by introducing item frequency factor, as shown in (3).

$$sim(v, u)_{AUCF} = \frac{\sum_{j \in |I_v \cap I_u|} \frac{1}{\log(1 + |U_j|)}}{\sqrt{|I_v| |I_u|}} \quad (3)$$

where $|U_j|$ is the number of items purchased by both u and v . $sim(v, u)_{AUCF}$ uses Log function on $|U_j|$ and calculates the reciprocal, thus effectively decreasing the impact of popular items on the similarity calculation. Employing (3), the similarity between two users will be high if these two users purchase the same unpopular items.

2.4 Recommendation Through Neighbors

Initially, there is a candidate item set in which the weight of each item is 0. For each item, we compute the sum of similarity between the target user and its neighbors who purchase the item. Therefore, each neighbor adds similarity to the weight of item.

$$Wu(j) = \sum_{v \in S \cap U_j} sim(u, v) \quad (4)$$

Finally, we can obtain an array of weights for items and arrange them in decreasing order. We can select the top N items as our recommendation to the target user. Note that the parameter K and N can be adjusted according to application scenario.

3 Experiment and Evaluation

This section discusses an experiment of recommendation using UCF algorithm and AUCF algorithm. The typical experiment process includes data preprocessing, experimental design, and the comparison of algorithms performance [14, 15].

3.1 Data Preprocessing

This experiment is based on the dataset provided by website Tmall.com, including 3667 users, 4212 cloths and 63502 transactions. Initially, we filter out the users who

Table 2. Tmall.com datasets used in UCF algorithms

Dataset	Clothing-purchasing information	User number	Clothes number
Total dataset	63502	3667	4212
Training dataset	44451	3667	3921
Test dataset	19051	3212	2331

only browse. Then the clothes that are not purchased by any customers in the dataset are also filtered out. The precision of the resulting recommendation would be improved. Table 2 shows the specific user purchasing records in total dataset, training set and testing set.

3.2 Evaluation Metric

The aim of evaluation is to assess result of recommendation. We use UCF algorithm and ACUF algorithm to predict the user's preference. We divide the data set into two separate parts: training set and testing test. 70% of data set is used as training set and the rest is used as testing set. Three metrics are used to assess the performance.

Recall: R_u are recommendations that are predicated by algorithms. T_u are all items favored by the target user. Recall represents the percentage of items that are predicated by algorithms and also favored by the target user.

$$Recall = \frac{\sum_u |R_u \cap T_u|}{\sum_u |T_u|} \quad (5)$$

Precision: Precision represents the percentage of items that are predicated by algorithms and also favored by the target user among all items. R_u are recommendations that are predicated by algorithms. T_u are all items favored by target user.

$$Precision = \frac{\sum_u |R_u \cap T_u|}{\sum_u |R_u|} \quad (6)$$

Coverage: Coverage represents the percentage of items that recommended by algorithm among all items. R_u represents the items that are recommended by algorithm, and I represents all items.

$$Coverage = \frac{|\cup_{u \in U} R_u|}{|I|} \quad (7)$$

This paper runs eight experiments with different datasets. For each experiment we perform four different runs by using cross validation so that our results are statistically accurate. In all of trials, we set $N = 10$ as the number of items recommended by the

top-K recommendation algorithms. We also adjust the K value (number of most similar users) to observe the change of performance.

3.3 Results of Experiments

The experiment is utilized to evaluate the performance of recommendation by using UCF algorithm and AUCF algorithm. Figs. 2 and 3 show the comparison of precision and recall using these two algorithms.

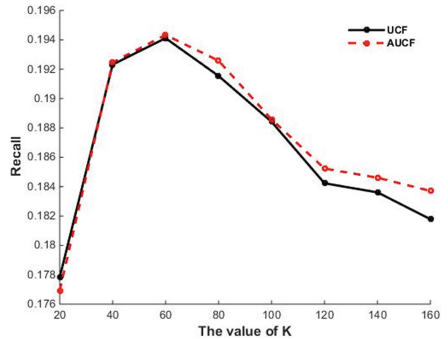
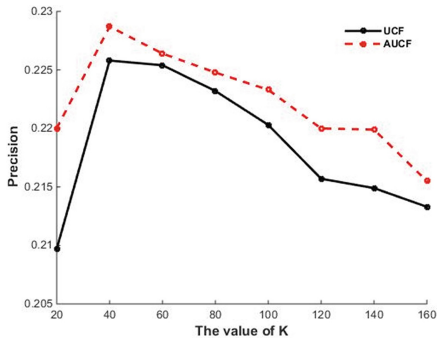


Fig. 2. Comparison of precision (UCF, AUCF) **Fig. 3.** Comparison of recall (UCF, AUCF)

From Fig. 2, AUCF algorithm achieves higher precision and the highest precision is increased from 22.58% to 22.87%. As shown in Fig. 3, AUCF algorithm achieves similar recall with UCF algorithm when K is from 20 to 60. When K is greater than 60, AUCF algorithm achieves higher recall. According to Fig. 4, AUCF algorithm achieves better coverage than UCF algorithm. The highest recommendation coverage is increased from 26.03% to 27.71%.

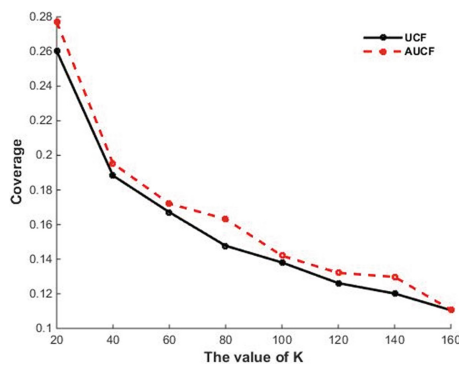


Fig. 4. Comparison of recommendation coverage (UCF, AUCF)

From above results, AUCF algorithm can increase the precision, recall and coverage. In addition, the similarity calculation among users by UCF algorithm takes 166 s. This is longer than time to calculate similarity by AUCF algorithm, which is 132 s. Therefore, using user-item linked list can save lots of time because of reduction of amount of useless calculation, which can make algorithm more efficient. Our future work is to utilize telecom big data to further investigate on the recommendation system and improve the performance of AUCF algorithm [16, 17].

4 Conclusion

In e-commerce, the recommendation system is widely used to assist customers to find their interested products. This paper evaluates the traditional UCF algorithm and discusses its application in recommendation systems of online clothing stores. This paper designs the AUCF algorithm. The AUCF algorithm can improve the calculation efficiency of the similarity among users. The AUCF algorithm can also reduce the sparseness of matrix. Therefore, it can make better recommendation to the customers.

References

1. Liu, Y., Xu, L., Chen, Y., Fan, Y., Xu, B., Nie, J.: A novel power control mechanism based on interference estimation in LTE cellular networks. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 397–401. IEEE Press, Qingdao (2016)
2. Cao, Y., Wang, N., Sun, Z., Cruickshank, H.: A reliable and efficient encounter-based routing framework for delay/disruption tolerant networks. *IEEE Sens. J.* **15**(7), 4004–4018 (2015)
3. Xu, L., Chen, Y., Gao, Y., Cuthbert, L.: A self-optimizing load balancing scheme for fixed relay cellular networks. In: 2th IET International Conference on Communication Technology and Application, pp. 306–311. IET Press, Beijing (2011)
4. Wang, W., Xu, L., Zhang, Y., Zhong, J.: A novel cell-level resource allocation scheme for OFDMA system. In: 1st International Conference on Communications and Mobile Computing, pp. 287–292. IET Press, Kunming (2009)
5. Xu, L., Cheng, X., Liu, Y., Chen, W., Luan, Y., Chao, K., Yuan, M., Xu, B.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: 16th IEEE International Conference on Communication Technology, pp. 806–812. IEEE Press, Hangzhou (2015)
6. Cao, Y., Sun, Z., Wang, N., Riaz, M., Cruickshank, H., Liu, X.: Geographic-based spray-and-relay (GSaR): an efficient routing scheme for DTNs. *IEEE Trans. Veh. Technol.* **64**(4), 1548–1564 (2015)
7. Xu, L., Luan, Y., Cheng, X., Xing, H., Liu, Y., Jiang, X., Chen, W., Chao, K.: Self-optimised joint traffic offloading in heterogeneous cellular networks. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 263–267. IEEE Press, Qingdao (2016)
8. Cao, Y., Wang, T., Kaiwartya, O., et al.: An EV charging management system concerning drivers' trip duration and mobility uncertainty. *IEEE Trans. Syst. Man Cybern. Syst.* **PP**(99), 1 (2016)

9. Xu, L., Luan, Y., Cheng, X., Cao, X., Chao, K., Gao, J., Jia, Y., Wang, S.: WCDMA data based LTE site selection scheme in LTE deployment. In: 1st International Conference on Signal and Information Processing, Networking and Computers, pp. 249–260. CRC Press Taylor & Francis Group, Beijing (2015)
10. Chinese Ecommerce Research Centre. <http://www.100ec.cn/>. Accessed 5 June 2017
11. Guo, G., Zhang, J., Zhu, F., Wang, X.: Factored similarity models with social trust for top-N item recommendation. *Knowl. Based Syst.* **122**, 17–25 (2017)
12. Wang, W., Zhang, G., Lu, J.: Member contribution-based group recommender system. *Decis. Support Syst.* **87**, 80–93 (2016)
13. Fulton, W.: *Intersection Theory*, 2nd edn. Springer, New York (1998)
14. Xu, L., Cheng, X., Chen, Y., Chao, K., Liu, D., Xing, H.: Self-optimised coordinated traffic shifting scheme for LTE cellular systems. In: 1st EAI International Conference on Self-Organizing Networks, pp. 67–75. Springer Press, Beijing (2015)
15. Cao, Y., Sun, Z., et al.: Routing in delay/disruption tolerant networks: a taxonomy, survey and challenges. *IEEE Commun. Surv. Tutor.* **5**(12), 654–677 (2013)
16. Zhang, X., Wang, N., Cao, Y., et al.: A stochastic analytical modelling framework on ISP-P2P collaborations in multi-domain environments. *IEEE Syst. J.* **PP**(99), 1 (2016)
17. Xu, L., Luan, Y., Cheng, X., Fan, Y., Zhang, H., Wang, W., He, A.: Telecom big data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)

OSS Data Based LTE Wireless Coverage Efficiency Analysis Method

Xingyu Fan^(✉), Jun Lu, and Weiwei Chen

China Unicom Network Technology Research Institute, Beijing 100048, China
fanxy@dimpt.com

Abstract. Traditional methods are generally based on single data source or several simple indicators to achieve coverage efficiency analysis, which makes them neither effectively reflect the full reality of network nor provide sufficient restrictions. In response to this problem, an OSS data based LTE wireless coverage efficiency analysis method is proposed. This method is capable of reflecting network status from the granularity of cell-level indicators and the dimension of multi-source data. Using this method, the improvement in the effectiveness of network state reflection and the enhancement of ability to coverage efficiency analysis have been demonstrated by practical application.

Keywords: Coverage efficiency · Key performance indicator · OSS data

1 Introduction

Coverage efficiency analysis plays an important role in wireless network planning and optimization. The coverage efficiency, which reflects the network resource availability and the network structure rationality, represents the ability to provide basic service to users in a certain area [1–3]. For network status evaluation and network problem discovery, coverage efficiency analysis is considered as a fundamental requirement. However, traditional methods are generally based on single data source or several simple indicators to achieve analysis, which makes them unable to fully reflect network state [4–6]. In response to this problem, a method capable of reflecting network status from granularity of cell-level indicators and dimension of multi-source data is proposed to improve the effectiveness of coverage efficiency analysis.

2 Contents of the Coverage Efficiency Analysis Algorithm

2.1 Algorithm Overview

Based on the key indicators of cell status, combining multi-source OSS data such as MR, counter, configuration and so on, this algorithm uses perception-related KPIs to evaluate the coverage quality and the efficiency of wireless network. Three main processes are included in the algorithm flow, namely data acquisition, analysis decision and result clustering.

Data acquisition process is the basis for coverage performance analysis. With the formal commercialization of LTE networks, the status and interactive information generated by huge amount of devices and massive online users has become the key factor in accurately grasping network status and effectively supporting user perception. To ensure the accuracy, completeness and validity of acquired data, the acquisition process directly collects original data file generated by online devices.

The cleanup and analysis of the acquired raw data is the primary purpose of analysis decision process. After eliminating the influence of invalid data through cleaning processing, the business anomalies are targeted based on O-side data fusion, and the problem cell and its main influencing factors are identified.

The purpose of the result clustering process is to classify the cells into the problem category and establish a cell clustering table with the corresponding problem type for further analysis. Such clustering result directly reflects the coverage level of wireless network and can assist the relevant departments to provide support for LTE wireless planning and optimization.

The algorithm flow is shown in Fig. 1.

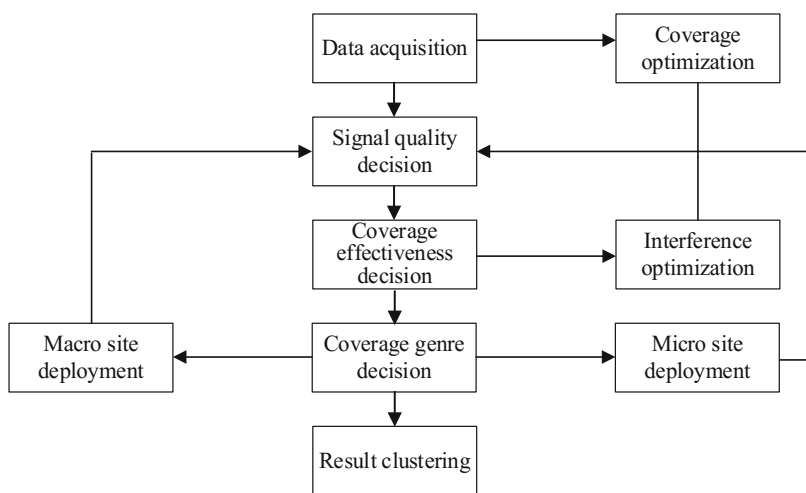


Fig. 1. The algorithm flow of coverage efficiency analysis

2.2 Algorithm Flow

The algorithm includes the following core steps: data acquisition, signal quality decision, coverage effectiveness decision, coverage genre decision, result clustering.

Data acquisition. Data acquisition is the basis for coverage efficiency analysis. To ensure data accuracy, raw data should be taken directly from devices and those acquired row data should be cleaned.

The goal of data cleansing is to identify unavailable cells. Here the so-called “unavailable” represents that the statistical significance of data from such cells is too

low to support coverage efficiency analysis. The main sources of such data include the following three, namely, low time-availability cell, low user-amount cell, and low context-built cell.

Low time-availability cell. As an important indicator of cell availability, the available time reflects the overall effective operation of cell. Data acquired from low time-availability cells need to be cleaned from the analysis process.

Identify low time-availability cells with the “Cell Available Duration” indicator. Define the set of all cells in local network as \mathcal{C} , then the set of low time-availability cells $\mathcal{C}_{unavail}$ can be defined as

$$\mathcal{C}_{unavail} = \{\mathcal{C} | L.Cell.Avail.Dur \leq Th_1\} \quad (1)$$

where the $L.Cell.Avail.Dur$ stands for “Cell Available Duration”, Th_1 stands for the selected threshold.

Low user-amount cell. The maximum number of users in a particular cell calculates the maximum number of UEs in RRC connection state during the statistical period. MR data obtained from low user-amount cells needs to be cleaned from the analysis process, for is shortage on statistical significance.

Identify low user-amount cells with the “Traffic User Maximum” indicator. The set of low user-amount cells \mathcal{C}_{fewUE} can be defined as

$$\mathcal{C}_{fewUE} = \{\mathcal{C} | L.Traffic.User.Max \leq Th_2\} \quad (2)$$

where the $L.Traffic.User.Max$ stands for “Traffic User Maximum”, Th_2 stands for the selected threshold.

Low context-built cell. The context building situation, which reflects the cell’s ability of allocating resources to users, is another important indicator of cell performance. Failure of context establishment means the user’s business is not supported, therefore the statistical significance of data from corresponding cells is affected due to the lack of operational information. Such data needs to be cleaned from analysis process.

Identify low context-built cell with the “UE-Context Successfully Establish” indicator. The set of low context-built cells $\mathcal{C}_{UEContxFail}$ can be defined as

$$\mathcal{C}_{UEContxFail} = \{\mathcal{C} | L.UECNTX.SuccEst \leq Th_3\} \quad (3)$$

where the $L.UECNTX.SuccEst$ stands for “UE-Context Successfully Establish”, Th_3 stands for the selected threshold.

Through the above-mentioned processing, all unavailable cells are identified and the corresponding data is cleaned. The residual data with sufficient statistical significance can be further implemented to coverage efficiency analysis.

Signal quality decision. As a direct reflection of wireless quality problem and the decisive factor influencing user’s rate perception, SINR is regarded as the main indicator to evaluate wireless signal situation. The purpose of signal quality decision process is to distinguish the defective cell and the normal cell according to their signal

quality level, that is, to identify the suspected low-quality cells according to SINR. However, measurement value of SINR is not reported in the current network, therefore the channel quality indicator (CQI), which reflects SINR directly, is adopted to characterize the wireless signal quality.

Identify suspected low-quality cells with the “Cell Measured CQI-mean” and “Cell Measured CQI-proportion” indicators. The set of suspected low-quality cells $C_{Lowqual}$ can be defined as

$$C_{Lowqual} = \{C | L.CMeas.CQI_{mean} \leq Th_4 \text{ or } L.CMeas.CQI_{prop} > Th_5\} \quad (4)$$

where the $L.CMeas.CQI_{mean}$ and $L.CMeas.CQI_{prop}$ stands for “Cell Measured CQI-mean” and “Cell Measured CQI-proportion” respectively. Th_4 and Th_5 stands for the selected threshold respectively.

As the average value of cell’s CQI, $L.CMeas.CQI_{mean}$ can be defined as

$$L.CMeas.CQI_{mean} = \sum_{k=0}^{15} (k \times L.CMeas.CQI.DL.k) \quad (5)$$

where the $L.CMeas.CQI.DL.k$ stands for the reported times of full bandwidth CQI equal to k.

As the proportion of low CQI value, $L.CMeas.CQI_{prop}$ can be defined as

$$L.CMeas.CQI_{prop} = \sum_{k=0}^{Th_6} L.CMeas.CQI.DL.k / \sum_{k=0}^{15} L.CMeas.CQI.DL.k \quad (6)$$

where the Th_6 stands for the selected threshold for low CQI value.

After the above-mentioned processing, all suspected low-quality cells are identified. The remaining cells can be classified as normal cells, for no serious quality problem is left behind.

Coverage effectiveness decision. To identify the main influencing factor of the suspected low-quality cells is the main purpose of coverage effectiveness decision process. In current network, the two primary causes of poor quality are insufficient coverage and interference existence. For the former case, detected signal strength of many users in the corresponding coverage area can hardly meet the basic requirements, therefore resulting in insufficient SINR level. For the latter case, high interference level is the corresponding cause of insufficient SINR. In short, the former is proportional to SINR, while the latter is inversely proportional to SINR.

RSRP, which represents the absolute signal strength received by UE, is positively correlated with SINR level and considered to be key indicator of the network coverage strength. For those suspected low-quality cells caused by insufficient coverage, high proportion of low RSRP level sampling value is appeared. Meanwhile for those interference-caused low-quality cells, no similar phenomenon exists. Therefore, the coverage effectiveness decision can be achieved through the cells’ RSRP distribution, that is, the “Cell Measured RSRP-proportion” indicator.

The interference-caused low-quality cells are classified as interference cells. The set of interference cells, which is represented by $C_{Interfere}$, can be defined as

$$C_{Interfere} = \{C | L.CMeas.RSRP_{prop} \leq Th_7\} \quad (7)$$

where the $L.CMeas.RSRP_{prop}$ stands for the ‘‘Cell Measured RSRP-proportion’’, Th_7 stands for the selected threshold.

As the proportion of cell’s low RSRP values, $L.CMeas.RSRP_{prop}$ can be defined as

$$L.CMeas.RSRP_{prop} = \sum_{k=0}^{Th_8} (L.CMeas.RSRP.k) / \sum_{k=0}^{16} L.CMeas.RSRP.k \quad (8)$$

where the $L.CMeas.RSRP.k$ stands for the number of RSRP values within index k . Th_8 stands for the selected threshold for Low RSRP range.

After the above-mentioned processing, all suspected low-quality cells satisfying the criterion are classified as interference cells, while the other are identified as suspected insufficient coverage cells and will be further proceed in next step.

Coverage genre decision. To identify the genre of suspected insufficient coverage cells is the main purpose of coverage genre decision process. In other words, the main cause of insufficient coverage is judged. In current network, according to the types of scenes, the causes of poor coverage can be divided into two categories, namely shallow problem and depth problem. The SINR inferiority in shallow problem cells is due to the fact that RSRP obtained by most users in the coverage area cannot meet the basic needs because of the users’ large access distances. Meanwhile the SINR inferiority in depth problem cells is caused by direct signal path blocking.

In all suspected insufficient coverage cells, those cells with depth problem are classified as depth problem cells. The set of depth problem cells is represented by C_{NLOS} .

$$C_{NLOS} = \{C | |P_{ant} - P_{PL} - L.CMeas.RSRP_{mean}| \geq Th_9\} \quad (9)$$

where the P_{ant} stands for the actual signal strength transmitted over antenna, P_{PL} stands for the idea road loss, $L.CMeas.RSRP_{mean}$ stands for the average value of cell’s measured RSRP, Th_9 stands for the selected threshold for depth problem judgment.

The idea road loss can be defined as

$$P_{PL} = 32.4 + 20 \log(\sqrt{h_{eNB}^2 + \mathcal{L}_{TA}^2}) + 20 \log(f_M) \quad (10)$$

where the h_{eNB} stands for antenna’s height, f_M stands for the frequency of signal.

The average value of cell’s measured RSRP is defined as

$$L.CMeas.RSRP_{mean} = \sum_{k=0}^{16} (L.CMeas.RSRP.k \times \mathcal{B}_k) / \sum_{k=0}^{16} L.CMeas.RSRP.k \quad (11)$$

where the $L.CMeas.RSRP.k$ stands for the number of measured RSRP value within index k . The set of RSRP correction constants are represented as \mathcal{B} , and its elements are denoted as \mathcal{B}_k .

$$\begin{aligned} \mathcal{B} &= \{\mathcal{B}_k, k = 0, 1, \dots, 16\} \\ &= \{-132.5, -123, -120, -117, -114, -111, -108, -105, -102, -99, -96, -93, \\ &\quad -90, -87, -84, -81, -61\} \end{aligned} \quad (12)$$

All suspected insufficient coverage cells satisfying the criterion above are classified as depth problem cells. The other are classified as shallow problem cells.

Result clustering. Through the above mentioned steps, all local cells are classified into five categories according to their characteristics, that is, unavailable cells, normal cells, interference cells, depth problem cells, and shallow problem cells. Each type of cell can be analyzed separately and optimized for the actual scenario. Further, the listed cells can be mapped into GIS (Geographic Information System) by conjunction with configuration information, so as to form distribution layers for each type of cell. These layers present the regional aggregation characteristics of problem cells from the geographic latitude, thus providing a reference for network planning and operations.

Give an example, the aggregation area of shallow problem cells has the necessity of macro site supplement, meanwhile the area of depth problem cells is not recommended to deploy macro site but indoor distribution system or micro site. Similarly, for the interference aggregation area, the urgency of interfering control is higher than that of site deployment. Based on such principles, sites that are unreasonable or need to be carefully verified can be identified during planning and operations.

3 Implementation Effects

The threshold parameters involved in the algorithm flow directly affect the overall performance of the algorithm. Therefore, impact aspects such as data anomalies, index distribution, false alarms, missing rate, and so on, need to be considered and carefully weighed during the parameter configuration [7–9].

Take a city whose urban area contains about 2000 cells as an example. Based on theoretical analysis, combined with the city's specific business statistics, the threshold configuration shown in Table 1 is adopted.

Table 1. Threshold configuration for analysis

Threshold	Configuration	Unit
Th_1	3% of max value	s
Th_2	0	–
Th_3	0	–
Th_4	4	–
Th_5	30%	–
Th_6	4	–
Th_7	30%	–
Th_8	5	–
Th_9	20	dB

Based on the parameters shown in Table 1, coverage efficiency is analyzed, and the proportions of each type of cells are shown in Table 2.

Table 2. Proportions of each type of cells

Type of cells	Number of cells	Proportion
Unavailable cells	57	3.00%
Normal cells	1320	69.55%
Shallow problem cells	266	14.01%
Depth problem cells	184	9.69%
Interference cells	71	3.74%

Analytical results are geographic mapped to gain the distribution of each type of cells, as shown in Fig. 2.

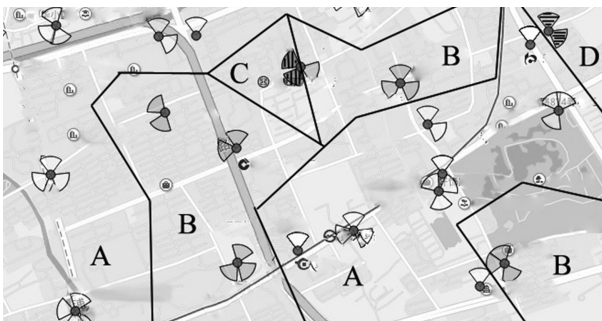


Fig. 2. Result clustering of each type of cells

In Fig. 2, normal cells, shallow problem cells, depth problem cells, interference cells are expressed in white, gray, vertical stripes and horizontal stripes respectively. The corresponding regions, that is, the normal region, the shallow problem region, the depth problem region, and the interference problem region are denoted by A, B, C, and D, respectively. The macro site should be deployed mainly in shallow problem region. Meanwhile for interference region, the urgency of interfering control is higher than that of site construction. Sites violate such basic principles are unreasonable and need to be carefully verified.

4 Conclusion

Coverage efficiency analysis plays an important role in wireless network planning and optimization, and is considered as a fundamental requirement especially in network status evaluation and network problem discovery. Traditional methods cannot adequately meet such requirements. In response, a method capable of reflecting network status from granularity of cell-level indicators and dimension of multi-source data is proposed to improve the effectiveness of coverage efficiency analysis. The improvement in the effectiveness of network state reflection and the enhancement of ability to coverage efficiency analysis have been demonstrated by practical application. Further improvement of the indicator system will be the next challenge.

References

1. Dahlman, E., Parkvall, S., Sköld, J., et al.: 4G - LTE/LTE-Advanced for Mobile Broadband. Academic Press, Oxford (2011)
2. Buenestado, V., Toril, M., Luna-Ramirez, S., et al.: Self-tuning of remote electrical tilts based on call traces for coverage and capacity optimization in LTE. *IEEE Trans. Veh. Technol.* **66**, 4315–4326 (2017)
3. Kaya, A.Ö., Calin, D.: On the wireless channel characteristics of outdoor-to-indoor LTE small cells. *IEEE Trans. Wirel. Commun.* **15**(8), 5453–5466 (2016)
4. Miao, Y., Sun, Z., Wang, N., et al.: Time efficient data collection with mobile sink and mMIMO technique in wireless sensor networks. *IEEE Syst. J.* **PP**(99), 1 (2016)
5. Xu, L., Luan, Y., Cheng, X., et al.: Telecom big data based user offloading self-optimization in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)
6. Xu, L., Cheng, X., Liu, Y., et al: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: 16th IEEE International Conference on Communication Technology, pp. 806–812. IEEE Press, Hangzhou (2015)
7. Fan, X., Lu, J., Miao, S., et al.: Analysis of key wireless factors affecting on VoLTE user-perceived performance based on measured data. *Mob. Commun.* **41**(14), 5–12 (2017)
8. Kaiwartya, O., Abdullah, A., Cao, Y., et al.: T-MQM: testbed based multi-metric quality measurement for sensor deployment for wireless sensor networks-a case study. *IEEE Sens. J.* **16**(23), 8649–8664 (2016)
9. Valizadeh, S.R., Abouei, J.: An adaptive distributed coverage optimization scheme in LTE enterprise femtocells. In: 22nd Iranian Conference on Electrical Engineering, pp. 1723–1728. IEEE, Tehran (2014)

A Radio Network Differentiation Parameter Optimization Algorithm Based on Clustering

Guanggen Guo¹, Baisong Ren², Gang An^{3(✉)}, Wendong Wu²,
Lexi Xu⁴, Hongxing Bai¹, and Zhongxi Zhao¹

¹ China United Network Communication Corporation, Cangzhou Branch,
Cangzhou, People's Republic of China

² China United Network Communication Corporation, Hebei Branch,
Shijiazhuang, People's Republic of China

³ Nokia Solutions and Networks System Technology (Beijing) Co., Ltd.,
Beijing, People's Republic of China
gang.an@nokia.com

⁴ Network Technology Research Institute,
China United Network Communication Corporation,
Beijing, People's Republic of China

Abstract. This article presents a new parameter optimization algorithm based on radio coverage grids clustering analysis. First, the coverage zone is divided into some grids based on drive test data and measure reports including GSM/WCDMS/LTE RF signal performance index. Next, using system integration clustering algorithm, all grids are clustered in different groups based on the degrees of RF signal deviation. Finally, we select the optimized parameter configurations based on GSM/WCDMA/LTE signal strength and quality in each group. The performance of the inter-RAT cell reselection and handover are improved.

Keywords: Clustering · Radio network · Optimization · Coverage grid

1 Background

Now, the wireless access network of an operator has been developed into a new era. It operates 2G, 3G and 4G three networks at the same time. Voice and data are always the basic traffic types carried on radio bearers. Before the VoLTE service is launched in the network, voice call is carried on 2G and 3G networks. The voice calls of 4G user is fallen back to 3G or 2G radio bearers through CSFB procedure. As 4G networks continue to develop fast, the user demands on data traffic as well as the diversity of applications are increasing.

Meanwhile from the analysis of network optimization viewpoint, there are more discrepancies among the coverage status of 2G/3G/4G radio networks. It is the most significant research subject in radio network optimization field to make the three networks' traffic load develop synergistically. The best method is to use the most appropriate parameters in order to optimize user service perception. Traditionally, we set the configurations of the inter-RAT interaction parameters with differentiation based

on macro scenarios such as urban, suburban, rural and indoor. So many micro scenarios are not considered. It's hard to make sure that all users from all scenarios have the best service perception. Although, most of those interaction parameters can be configured by cell level.

According to 3GPP UU interface protocols [1], in the connection status, 2G, 3G and 4G terminals need to send measure reports to the networks periodically or event triggered. The measurement reports carry the RF signal strength, quality of the serving cell and the neighbors. So, there are tremendous measurement reports from all terminals in the radio networks, which can be considered as big data. Each measurement report is located in certain spot through geographic position algorithm. After distributed in geographic grids, usually the grids are 50 m * 50 m, the measurement reports can be weighted based on RF signal strength and quality. Each grid includes one strongest serving cell and one neighbor cell. If we configure inter-RAT parameters for every grid, the work is impossible to be finished. Certainly, it's unnecessary. Using big data analysis, we cluster the tremendous grids based on coverage characteristics. It's essential and reasonable to set inter-RAT parameters for the clustered groups. It's a good method to precisely optimize inter-RAT interaction performance.

This paper designs a system clustering method and applies the method to data mining analysis for coverage characteristics of radio network grids. First, we analyze measurement reports of radio network through geographic position and grid distribution algorithms, and label the grids based on RF signal coverage characteristics. Next, we design correlation coefficient between grids. By deeply analyzing the differences of the coverage grids, we find out the system clustering condition. Then we finish the coverage grids clustering through iterative clustering computations. Finally, for different grid groups, based on their 2G/3G/4G coverage characteristics of clustering center, we set the appropriate inter-RAT interaction parameters. The method is designed to improve inter-RAT interaction performance.

2 Coverage Grid Labeling

From the definitions in 3GPP about the RF coverage characteristic indexes involved in 2G/3G/4G interaction parameters, we confirm that RSSI and C/I in 2G system, RSCP and Ec/Io in 3G system, RSRP and SINR in 4G are considered as main factors. Because the inter-RAT parameters are based on cell level, the coverage grids include 2G/3G/4G cell IDs. In order to reflect the facticity and effectiveness of the network RF coverage excluding occasionality, we export at least 7 days * 24 h measurement report original data to analyze.

By using measurement reports analysis tool, we finish the computation of measurement report geographic positioning and grid distribution. Figure 1 shows RF signal's strength and quality of the best serving cell and one neighbor cell, including 2G, 3G and 4G cells. Specially, in some grids, a certain network cell's signal will be empty because there is no RF signal coverage in the location.

GRID_X	GRID_Y	WCELL1	SCR1	AVG_RSC_P1	AVG_ECN_O1	GSMCELL_1	AVG_RSSI	AVG_CI	LTECELL1	AVG_RSC_P	AVG_SINR
40.776077	111.69561	30591	292	-73.6	-5.86	11891	-75.38	2.92	45589267	-82	8
40.777079	111.69511	30591	67	-60.8	-4.06	11891	-70.33	3.39	45589267	-89	8.9
40.77808	111.69511	30591	292	-84.3	-7.78	11891	-76.44	2.19	45589267	-94	12.4
40.778581	111.70262	32673	51	-70	-5.5	65271	-69.5	0.71	45232913	-55	16.9
40.779082	111.6901	32842	324	-62	-2.92	14322	-75.73	3.35	34348819	-92	10

Fig. 1. 2G/3G/4G measurement report framework

Representing the coverage characteristic indexes involved in coverage grids:

The 2G coverage characteristic index is represented by (1)

$$A = (A_1, A_2) \tag{1}$$

The 3G coverage characteristic index is represented by (2):

$$U = (U_1, U_2) \tag{2}$$

The 4G coverage characteristic index is represented by (3)

$$L = (L_1, L_2) \tag{3}$$

Where, the variates of A_i , U_i and L_i mean RF signal strength and quality.

Using weighted computation method to represent 2G/3G/4G coverage characteristics indexes by formula (4) (5):

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, p) \tag{4}$$

Where, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ is the sample mean of No.j variate. S_j is the sample variance of No.j variate.

$$S_j = \frac{1}{n - 1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \tag{5}$$

3 A System Clustering Algorithm Based on Coverage Characteristic Indexes

The main principle is to design the all grids in the optimization zone as clustering targets and make random distribution on them first. Next, analyze 2G/3G/4G coverage characteristic indexes deeply by computing coverage strength, coverage quality, coverage strength difference, coverage quality difference and sample mean and variance. Then, cluster the sample mean and variance based on density, select several targets as

the clustering center according to 2G/3G/4G inter-RAT interaction parameter setting principles and use iterative computations on all targets to get the final optimized clusters. Finally, set the appropriate inter-RAT parameters for all cells based on coverage characteristics of each cluster.

According to usual 2G/3G/4G interaction scenario parameter setting experience, we divide the grid targets into two layers based on good or bad 4G network coverage, and then compute the next clustering through 2G and 3G coverage performance.

First step, we calculate the distance between the correlation coefficient of 3G and 2G RF signal strength variations:

Where c_{ij} represents the correlation coefficient of variations x_i and x_j , usually, it must satisfy following three conditions:

- (1) $c_{ij} = \pm 1$, if and only if $x_{ij} = ax_j + b$,
 $a(\neq 0)$ and b are constants
- (2) $|c_{ij}| \leq 1$, for all i, j ;
- (3) $C_{ij} = c_{ji}$, for all i, j .

$$c_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\left\{ \left[\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \right] \left[\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2 \right] \right\}^{1/2}} \tag{6}$$

We design distance through the above correlation coefficient:

$$d_{ij}^2 = 1 - c_{ij}^2$$

Second step, clustering:

Here, the basic principle of clustering the grid samples is to make n samples in one category each other, provide the distance between the samples and provide the distance between the clusters first. Then combine the closest clusters as a new cluster, computing the distance between the new cluster and other clusters. Finally, combine the closest clusters as a new cluster repeatedly, reducing one cluster each time, till all samples are combined into one cluster.

Let us denote d_{ij} as the distance between sample i and j , and denote G_1, G_2, \dots as the clusters, and D_{KL} as the distance between G_K and G_L . In the method introduced in this paper, each grid makes itself as one cluster at the beginning and the distances between the clusters equals to the distance between the samples. Therefore, $D_{KL} = d_{KL}$, and the distance matrixes are the same, represented by:

$$D_{(0)} = (d_{ij}) \tag{7}$$

We define the sum of the squared Euclidean distance (inside clusters) from each sample to the cluster center as sum of squares of deviations, or SSD. If G_K and G_L are combined into the new cluster G_M , then the SSDs of G_K, G_L and G_M are calculated as (8), (9) and (10), respectively:

$$W_K = \sum_{i \in G_K} (\mathbf{x}_i - \bar{\mathbf{x}}_K)^T (\mathbf{x}_i - \bar{\mathbf{x}}_K) \tag{8}$$

$$W_L = \sum_{i \in G_L} (\mathbf{x}_i - \bar{\mathbf{x}}_L)^T (\mathbf{x}_i - \bar{\mathbf{x}}_L) \tag{9}$$

$$W_M = \sum_{i \in G_M} (\mathbf{x}_i - \bar{\mathbf{x}}_M)^T (\mathbf{x}_i - \bar{\mathbf{x}}_M) \tag{10}$$

If the distance between G_K and G_L is closer, the added SSD after the combination $W_M - W_K - W_L$ will be smaller. Otherwise it will be bigger. So, the squared distance between G_K and G_L is defined as:

$$D_{KL}^2 = W_M - W_K - W_L \tag{11}$$

It can be used as another presentation for computing conveniently.

$$D_{KL}^2 = \frac{n_L n_K}{n_M} (\bar{\mathbf{x}}_K - \bar{\mathbf{x}}_L)^T (\bar{\mathbf{x}}_K - \bar{\mathbf{x}}_L) \tag{12}$$

Formula derivation of squared distance from SSDs is expressed as:

$$D_{MJ}^2 = \frac{n_J + n_K}{n_J + n_M} D_{KJ}^2 + \frac{n_J + n_L}{n_J + n_M} D_{LJ}^2 - \frac{n_J}{n_J + n_M} D_{KL}^2 \tag{13}$$

Third step, determining the amount of the clusters:

During the iterative computation procedure, based on the requirements of setting 2G/3G/4G interaction parameters, we assess the result of each iterative computation and select the most appropriate amount of clustering, as well as the number of scenarios of the parameter setting.

If the number of samples is n , all samples are combined into G_1, G_2, \dots, G_k during clustering computation. The sample amount of cluster G_i and cluster center are n_i and $\bar{\mathbf{x}}_i, i = 1, 2, \dots, k$. Then $\sum_{i=1}^k n_i = n$, the total cluster center of all samples is

$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^k n_i \bar{\mathbf{x}}_i$. The total SSD of all samples is defined as:

$$W = \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})^T (\mathbf{x}_j - \bar{\mathbf{x}}) \tag{14}$$

The SSD of samples in group G_i is:

$$W_i = \sum_{j \in G_i} (\mathbf{x}_j - \bar{\mathbf{x}}_i)^T (\mathbf{x}_j - \bar{\mathbf{x}}_i) \quad (15)$$

The SSD of k groups (inside) is:

$$p_k = \sum_{i=1}^k W_i \quad (16)$$

After computing, W can be disassembled as:

$$W = P_k + \sum_{i=1}^k n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \quad (17)$$

Defining

$$R^2 = 1 - P_k/W = \sum_{i=1}^k n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) / W \quad (18)$$

Then, the value of P_k/W is smaller, (as well as the value of R^2 is bigger),

Which means the proportion of inside-group SSD is smaller in the total SSD, as well as k groups are separated further. Hence, statistics of R^2 can be used to assess the effect of clustering k groups. The value of R^2 is bigger, the result of clustering is better.

4 Experimental Verification

We exported at least 7 days * 24 h measurement report original data to analyze.

Our algorithm was implement in MATLAB environment. The experiments were completed on a HP server with Windows operating system and 2*Intel Xeon E5-2699 2.2 GHz, 16 * 16 GB RDIMM, 2 * 400 GB SSD.

After clustering, we identify five clusters of scenarios with different RF coverage.

Figure 2 shows the results of clustering. Table 1 shows the main radio network parameter configurations of inter-RAT interaction.

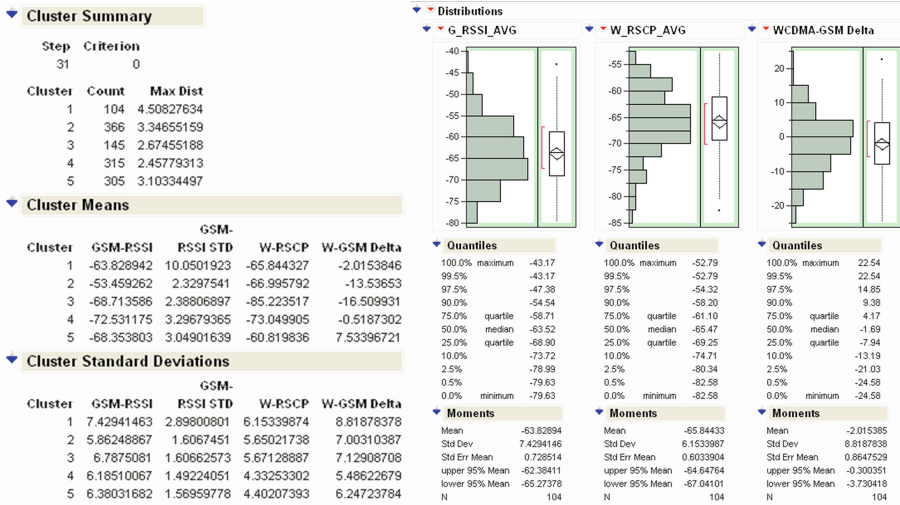


Fig. 2. Clustering results for X City A operator RF coverage grids.

Table 1. Main Inter-RAT interaction radio network parameters setting

Cluster number after clustering	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Scenario description	Good W and G coverage, little difference	W little poor, more difference	W poor, G good, more difference	Normal W and G, little difference	Good W, normal G, more difference
<i>2G Cell reselection</i>					
FDD_Qmin	7(-12 dB)	7 (-12 dB)	5(-10 dB)	5(-10 dB)	7(-12 dB)
<i>3G Cell reselection</i>					
QqualMin	-18 dB	-18 dB	-18 dB	-18 dB	-18 dB
QrxlevMin	-111dBm	-111dBm	-111dBm	-111dBm	-111dBm
<i>3G Handover</i>					
HHoRscpThreshold	-105	-105	-95	-100	-105
<i>4G Handover</i>					
b1ThresholdCSFBUltraRscp	-95	-95	-85	-90	-95
qRxLevMinUltra	-105	-105	-95	-100	-105

Figure 3 shows the optimization improvement after using the appropriate parameter settings. The main KPIs such as inter-RAT handover success rate and 4G CSFB success rate are increased.

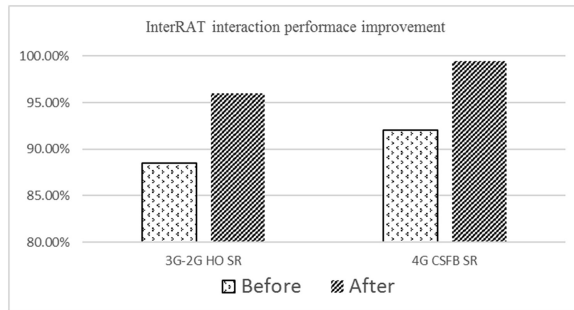


Fig. 3. Network optimization performance after parameter setting

5 Summary

This article presents a novel radio network differentiation parameter optimization algorithm based on radio coverage grids clustering analysis. Big data analysis and radio network RF signal elements clustering enable the telecom operator to optimize wireless networks performance accurately and improve user's perception efficiently. Due to the rapid development of the radio network, there are so much network elements which can be considered as big data. The cluster analysis for network performance optimization, user perception improvement can also be used, which will be the research topic of our future work.

References

1. Xu, L., Luan, Y., Cheng, X., Cao, X., Chao, K., Gao, J.: WCDMA data based LTE site selection scheme in LTE deployment. In: International Conference on Signal and Information Processing, Networking and Computers (ICSINC), Beijing, China, pp. 249–260, October 2015
2. Baisong, R.: An intelligent swarm clustering algorithm using swarm similarity measure. In: International Symposium on Communications and Information Technologies (ISCIT) (2016)
3. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv.* **31**(3), 264–323 (1999)
4. Li, C., Jing, L., Hou, S.: Smart data mining and technology research. Xi'an University of Electronic Science and Technology Publishing House, Xi'an (2006)
5. Han J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco (2001)
6. Tongming, L.: Data mining technology and applications. National Defence Industry Press, Changsha (2001)

A Value-Added Service Strategy for 3G Mobile Network Based on Network Resource Utilization

Lijuan Cao^(✉), Yuwei Jia, Chuntao Song, Jie Gao,
and Xinzhou Cheng

Department of Network Optimization and Management, China Unicom Network
Technology Research Institute, Beijing, People's Republic of China
{caolj68, jiaYW18, songct2, gaojie49,
chengxz11}@chinaunicom.cn

Abstract. With the gradual improvement of mobile network technology and the vigorous development of mobile Internet business, data traffic revenue has become telecom operators' main driver of profit growth. In this paper, a novel value-added service strategy is proposed for 3G mobile communication, in which the priority quality of service (QoS) will be provided in the case of network resource constraints if the user order this service. The main objective of this service is to improve user perception in network resource limitation. The experimental results reveal that the proposed scheme is able to achieve good performance in addressing the user satisfaction problem.

Keywords: Value-added service · Network resource · QoS · User perception

1 Introduction

The newly developed mobile broadband infrastructures, associated with the rapidly growing smartphones penetration is boosting the mobile Internet usage. Telecom operators have entered the era of traffic management. To counter this trends, telecom operators need to explore new value-added service to monetize mobile broadband data and increase user perception [1–4].

Some deployed value-added data services [5–8] like Sponsored data, Time based service and Shared wallet service have been proved notable success among the world. The data growth contributes to the interest of telecom operators by means of different value-added service. However, to our knowledge, little attention has been devoted to solve the quality of service (QoS) problem caused by the lack of network resource. As a result, all of the users' perception will deteriorate under bad network resource constraints in certain areas, and the user satisfaction will be badly effected as well as the interest of telecom operators. In this paper, we proposed a value-added service strategy based on network resource utilization for 3G mobile networks, and it also can used for 4G network after some proper adjustment.

The proposed strategy is based on the network resources and QoS characteristics in WCDMA. Specifically, in the case where the utilization rate of the network resource

reaches the threshold, the QoS negotiation mechanism between SGSN and GGSN will be used in the establishment process of 3G PS service, which provide users with differentiated business guarantee mechanism. So as to solve the poor user perception problems caused by congestion or resource constraints and then promote the use of user traffic.

This paper is organized as follows. Section 2 briefly presents the network resource and QoS configuration in WCDMA. Section 3 describes the network resource utilization based value-added service strategy proposed in this paper. Section 4 presents experimental results, which demonstrate the validity of our strategy. Section 5 gives the conclusion and future work of this strategy.

2 Key Network Resource Indicators and QoS in WCDMA

2.1 Network Resource Indicators in WCDMA

Code Resource Utilization: WCDMA is a kind of CDMA spread spectrum communication system. The scramblers are used to distinguish different UE, while the OVSF channelization codes are used to spread spectrum and to distinguish different downlink channels in the same cell. High utilization rate of code will greatly increase the possibility of abnormal events, and then effect user experience.

Power Utilization: WCDMA power utilization represents the use of carrier downlink power. TCP will be limited and downlink power utilization get higher when the users increase, but the downlink power distributed to single user will be short. As a result, the user experience will be effected.

Uplink/Downlink CE Utilization of Base Station: 1CE is the sum of the coding hardware and software resources needed to process a 12.2K voice service, and it is a logical concept instead of physical attributes. The CE data configured by each base station characterizes the service processing capability of the base station itself. When the CE resource is insufficient, the new access to the user will be limited, which results in switching calls and other issues.

Uplink/Downlink Iub interface Bandwidth Utilization of Base Station: Uplink Iub interface Bandwidth Utilization of Base Station = Iub interface uplink traffic/Iub interface uplink configuration bandwidth *100%.

Downlink Iub Bandwidth Utilization of Base Station = Iub interface downlink traffic/Iub interface downlink configuration bandwidth *100%.

3 QoS Control in WCDMA

All packet flows mapped to the same PDP receive the same packet-forwarding treatment. To provide different QoS [9] defined a series of parameters including: Transfer delay, Guaranteed bit rate, Allocation/Retention priority and so on.

In the PDP activation process [10], SGSN sends the PDP QoS parameter value to GGSN based on the user's contract information in the HLR, and the GGSN allocates negotiated QoS to SGSN. As a result, the SGSN will perform the RAB setup process

according to the negotiated QoS. Different QoS parameters can be set in the GGSN as static rule so that priority service can be provided for certain users.

4 Network Resource Utilization Based Service Strategy

4.1 Network Resource Utilization and PS Dropping Rate Evaluation

Collect the KPI data for D days, which contains network resource information and PS performance. In this step, network resource utilization and PS dropping rate are calculated in both time and geographic dimensions for each cell.

We define RU as the network resource utilization considering 6 key network resource indicators as mentioned above. Taking (Cell_ID, Hour, Date) as the granularity, the network resource utilization and the dropping rate for the PS service in WCDMA can be obtained using formula (1) and (2):

$$RU = w_1 \times U_{TxPower} + w_2 \times U_{CE_UL} + w_3 \times U_{CE_DL} + w_4 \times U_{Code} + w_5 \times U_{Iub_UL} + w_6 \times U_{Iub_DL} \quad (1)$$

w_i is the weight for the indicator. w_i can be adjusted according to the actual situation.

$$w_i \in [0, 1], \sum_{i=1}^6 w_i = 1. \quad (2)$$

The dropping rate for the PS service can be obtained using formula (3):

$$DR = \frac{RAB \text{ abnormal release number}}{RAB \text{ successful establishment number}} \quad (3)$$

4.2 Time and Date Map

In view of the utilization changes of network resources and different load scenario in time and date, the classification of the date and time are carried out to obtain the certain map.

Date_Type: According to the resource utilization rate with the date of the law, the date type Date_Type can be divided into three categories {weekdays, weekends, other holidays}, but not limited to this classification.

Period_Type: 24 h can be divided into N Period_Types to ensure that the network resource utilization of the same (Date_Type, Period_Type) for the same Area is similar.

To simply the performing, we define Date_Type and Period_Type in a provided way in Table 1.

With regard to implementation, Time and date map should be set when different characteristics are considered for each city.

Table 1. Time and date map.

Date_Type	Period_Type				
	01	02	03	04
01 (weekdays)	6:00–12:00	12:00–17:00	17:00–19:00	19:00–02:00
02 (weekends)	10:00–16:00	16:00–20:00	20:00–02:00	02:00–04:00
03 (holidays1)	10:00–16:00	16:00–20:00	20:00–02:00	02:00–04:00
04 (holidays2)	10:00–16:00	16:00–20:00	20:00–02:00	02:00–04:00
.....					

4.3 K-means Clustering for the Cell

We consider two elements: resource utilization rate and PS dropping rate in the clustering. It should be noted that, the clustering is performed for certain (Date_Type, Period_Type), which means for each (Date_Type, Period_Type) the clustering scheme need to be performed once. To simply the performing, we omit the Date_Type ‘d’ and Period_Type ‘p’ in the clustering algorithm.

The value of K and initial center for each cluster are needed for clustering analysis, so as to generate K cluster sets by iterative calculation. We set K = 3 as default value. 3 initial centers can be obtained by formula (4):

$$\begin{cases} C_1 \text{ center initialize : } \{ \overline{DR} \times (1 + \sigma), \overline{UR} \times (1 + \rho) \} \\ C_2 \text{ center initialize : } \{ \overline{DR} \times (1 + \sigma), \overline{UR} \times (1 - \rho) \} \\ C_3 \text{ center initialize : } \{ \overline{DR} \times (1 - \sigma), \overline{UR} \times (1 - \rho) \} \end{cases} \quad (4)$$

\overline{DR} is the average resource utilization of all the cells, while \overline{UR} represents the average PS dropping rate of all the cells. Index σ and ρ to represent the efficiency of this clustering, and they will influence the centers’ locations (Fig. 1).

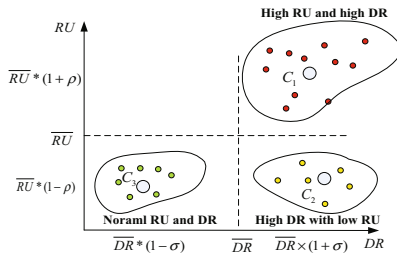


Fig. 1. The structure of K-means clustering analysis.

C1 is defined to represent the cluster with high resource utilization and high PS dropping rate, and C2 is high PS dropping rate and low resource utilization cluster and C3 is the low PS dropping rate and low resource utilization cluster (Table 2).

Table 2. K-means algorithm for the cells.

INPUT: K=3, and database contains resource utilization and PS dropping rate for C cells.
OUTPUT: three cluster subsets.

1. Determine each cluster center as (4);
2. Assign the data samples in the database to the nearest neighbor clustering according to the minimum distance principle by Euclidean distance model as (5);
3. Using the sample mean of each cluster as the new clustering center;
4. For i from 1 to C, repeat the procedure 2 and 3 until the Euclidean distances of each k(1 to 3) samples space converge to constant.

$$d(RU_c, DR_c) = \sqrt{(RU_c - C_{kx})^2 + (DR_c - C_{ky})^2}, c \in [1, C] \tag{5}$$

where RU_c is the average resource utilization rate for Cell_ID ‘c’.

By the result of clustering algorithm and optimization, Table 3 reveals the target cells of the proposed service strategy for each (Date_Type, Period_Type), as the high resource utilization probably greatly increase the PS dropping in C1.

Table 3. Target cells map.

Date	Date_Type	Hour	Period_Type	Cell_ID
2017/7/11	01	09:00	02	10231

4.4 Priority QoS Service Activation in GGSN

When the data packets arrive the GGSN, the service identification will be formed, so that (Date, Hour, Cell_ID, IMSI) of data packets can be obtained by the GGSN. If the data packets information matches Table 4, GGSN will allocate priority QoS in this PDP activation according to the service rules and perform the corresponding billing rules, otherwise the packet data will be forwarded using default QoS, as shown in Fig. 2.

Table 4. Target users map.

Date	Date_Type	Hour	Period_Type	Cell_ID	IMSI
2017/7/11	01	09:00	02	10231	46001141860XXXX

The premise of the proposed service is that, users order this service so that the operator will mark the IMSI of this consumer as advanced users with priority QoS in Table 4.

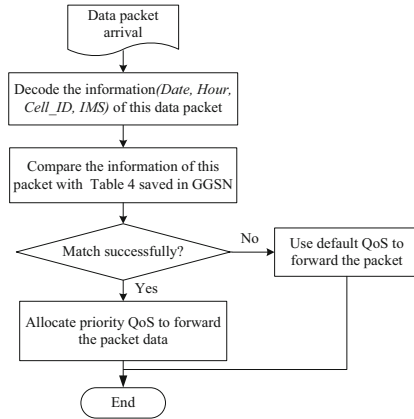


Fig. 2. Priority QoS Service in GGSN.

4.5 QoS and Charging Rule Set

(GBR, ARP, Transfer delay) are the most important parameters to guarantee the QoS of the PDP. The QoS design needs be decided by the algorithmic game theory, as QoS with low priority could cause user experience problems, while QoS with too high priority will cause the resource waste. The telecom operator provides a serials of QoS set for the users, and unit price ‘Pc’ can be set based on the QoS priority. So that users could choose a QoS set based on their experience and payment monthly.

$$Bill = P_c \times Traffic \tag{6}$$

Our goal in this article is to define a basic function, so we define (GBR, ARP, Transfer delay) in Table 5.

Table 5. QoS set and price rule for the RAB set.

Number	(GBR, ARP, Transfer delay)	Price
1	(12800 bit/s, 5, 150 ms),	P1
2	(12800 bit/s, 5, 100 ms),	P2
...

5 Experiment Results

This section shows the experiential results of our strategy. The dataset contains 2.8 millions of data of KPI in city S for during D = 7 days period (2017/3/1–2017/3/7). The resource utilization and PS dropping rate at 20:00–23:00 in weekdays of RAN1 and RNC2 are relatively high because of the crowd gathered, so we treat all the cells in these two RNC as the target cells. So we define (Date_Type, Period_Type) = (Weekends, 20:00–23:00). Target users are selected by random.

The results are demonstrated by the improvement of user perception, as we couldn't value the business profit generated by this strategy for the telecom operators. In Fig. 3, the proposed strategy could achieve less bad user perception, because the stricter QoS set could guarantee the users the GBR transfer delay and ARP.

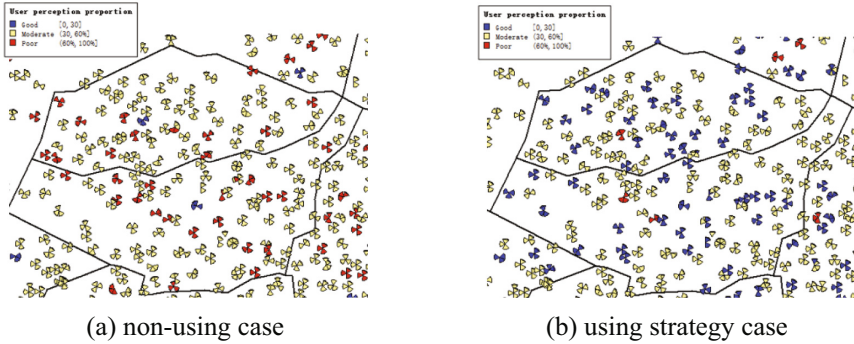


Fig. 3. Bad user perception proportion comparison.

Different results can be expected when different reference parameters are considered. Figure 4 shows that better the user experience could be achieved when the stricter QoS set is exploited by the bad user perception cumulative distribution function comparison.

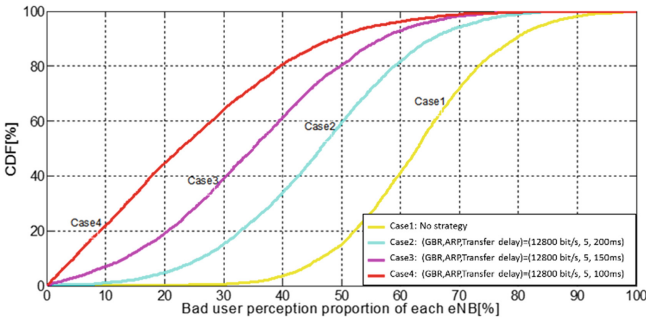


Fig. 4. Bad user perception proportion comparison of 3 different QoS sets.

6 Conclusion

This paper presents a new value-added service based on network resource utilization by means of QoS negotiation in 3G networks. The main objective of this service is to improve user perception in network resource limitation. Experimental result shows that

the proposed strategy is reliable and could achieve good performance in improving the user experience.

References

1. Xu, L., Luan, Y., Cheng, X., Xing, H., Liu, Y., Jiang, X., Chen, W., Chao, K.: Self-optimised joint traffic offloading in heterogeneous cellular networks. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 263–267. IEEE Press, Qingdao (2016)
2. Xu, L., Chen, Y., Gao, Y., Cuthbert, L.: A self-optimizing load balancing scheme for fixed relay cellular networks. In: 2th IET International Conference on Communication Technology and Application, pp. 306–311. IET Press, Beijing (2011)
3. Salem, A., Elhingary, A., Zerek, R.: Value added service for mobile communications. In: Conference 2013, Powereng, pp. 1784–1788. IEEE (2013)
4. Du, H., Zhu, G., Zhao, L., et al.: An empirical study of consumer adoption on 3G value-added services in China. In: Conference 2011, ICMSS, pp. 1–4. IEEE (2011)
5. Nwanga, M., Onwuka, E., Aibinu, A., et al.: Impact of big data analytics to Nigerian mobile phone industry. In: Conference 2015, IEOM, pp. 1–6. Taylor (2015)
6. Šerval, D., Marković, C., Kovačević, S.: 4G mobile internet, services, regulation and mobile operators in Bosnia and Herzegovina. In: Conference 2014, MIPRO, pp. 432–435. IEEE (2014)
7. Cao, L., Zhou, Y., Cheng, X., et al.: An user perception based value-added service strategy for 4G mobile networks. In: Conference 2015, ICSINC, pp. 299–306. Taylor (2015)
8. Gao, J., Cheng, X., Xu, L.: An interference management algorithm using big data analytics in LTE cellular networks. In: Conference 2016, ISCIT, pp. 246–251. IEEE (2016)
9. 3GPP TS 23.107: Quality of Service (QoS) concept and architecture (Release 6)
10. 3GPP TS 23.060: General Packet Radio Service (GPRS) Service description

Big Data and Location Based Dynamic Power Control for Small Cell Networks

Yi Li^(✉), Yucang Yang, Xingyu Fan, Jun Lu, and Weiwei Chen

China Unicom, Network Technology Research Institute, Beijing 100048, China
{liyiz, yangyc, fanxy, lujun, chenww}@dimpt.com

Abstract. To deal with the ever-growing demand for data service, the small cell network (SCN), a new structure network has been proposed. Although through spatial reuse SCNs may greatly increase the networks' throughput, it may also lead to severe intra-layer interference and excessive handover (HO) overheads. Besides, the large-amount small cell scenery can also lead to severe network energy consumption. In the article, a dynamic power control scheme is proposed for small cells to handle the interference and HO problems. It can not only guarantee the cell edge users' perception but also reduce the network energy consumption. The set of experiments in the big data analysis shows the scheme can bring a substantial growth in cell edge users' throughput as well as decrease system outage probability.

Keywords: Small cell networks · Dynamic power control · Big data analysis

1 Introduction

Confronted with the ever-growing requirement for wireless data services, the existing network structure can no longer meet the explosive data service growth. While macrocells have been effective in providing outdoor and indoor data services, they are not enough for high-capacity hotspots.

Small cell network is a novel and different network type. They could bring an energy-efficient solution to the high-capacity area. SCNs consists of low-cost, low-power, self-organizing and dense base stations. The significant use of SCNs is to offload the capacity from macrocells and improve the cell edge users' perception. However, due to the dense deployment, they may seriously interfere with each other. Moreover, the reduced cell radius of SCNs can also lead to more frequent HO procedures.

In this article, a location based dynamic power control scheme is proposed. We set the transmit power according to the users' location. We jointly optimize the network throughput and handover (HO) performances. The paper structure is as follows. Section 2 proposes the model hypothesis and analyses the existing co-interference scenario. Section 3 describes the proposed dynamic power control scheme. In Sect. 4, results for big data analysis are illustrated. Finally, Sect. 5 summarizes the essay.

2 System Model

According to 3GPP LTE-A standardization, the existing network should give priority to different frequency bands strategy when assigning the macrocells and small cells deployment. Small cell enhancements should be applicable to all existing cellular bands. Since there has been quite amount of research on macrocells before, herein we pay our attention on the network performance of SCNs.

Although small cells are exempted from macrocells interference due to different operating frequencies, the dense small cell deployment may cause severe co-interference. Figure 1 shows the specific deployment scenario. We can see that each macrocell is contained with several small cell clusters. Meanwhile, each cluster is contained with a number of small cells [4]. Freq. 1 and Freq. 2 are the different carriers for the macrocells and small cells.

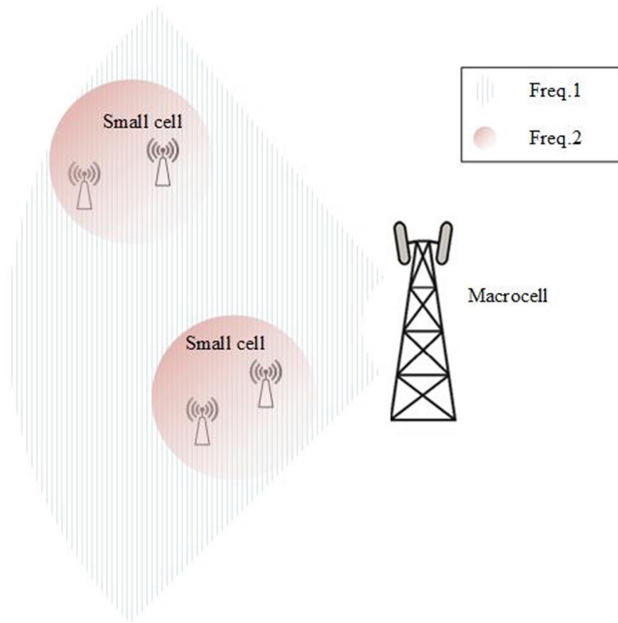


Fig. 1. Deployment scenario for SCNs

Downlink (DL) LTE-A is an Orthogonal Frequency Division Multiplexing (OFDM) system. On the basis of the formula for Signal to Interference plus Noise Ratio (SINR), $SINR_{n,k}$ for the small cell user u on carrier c can be shown as follows:

$$SINR_{u,c} = \frac{Power_c^{Server} \times Gain_{u,c}^{Server}}{\sum_{neighbour} Power_c^{neighbour} \times Gain_{u,c}^{neighbour} + Noise \times \Delta f} \tag{1}$$

$Power_c^{Server}$ is the transmitting power of the serving small cell on carrier c while $Power_c^{neighbour}$ belongs to the neighbor cells. $Gain_{u,c}^{Server}$ and $Gain_{u,c}^{neighbour}$ are the path loss between the user u and the small cell on carrier c . $Noise$ is the white noise and Δf is the width of LTE carrier.

According to $SINR_{n,k}$, the Channel Quality Indicator (CQI) of the user can be calculated. By looking for the 3GPP 36.213, we can assure the Modulation and Coding Scheme (MCS) and the efficiency of the small cell user. When using $Efficiency_{n,c}$, $Throughput_u$ of the small cell user can be shown as follows:

$$Throughput_u = \sum_{c=1}^{N_{RB}} Efficiency_{n,c} \quad (2)$$

Assuming $U = \{u = 1, \dots, U \mid SINR_u < 0\}$ is the set of cell edge users who suffered severe interference. And $\|U\|$ is the number of out-of-service users. The outage probability of the network can be shown as follows:

$$Outage_probability = \|U\|/U \quad (3)$$

From the formulae above, we see that the transmit power of small cells can significantly influence the network performance. Therefore, to minimize interference and enhance system throughput, an intelligent dynamic power control is well needed.

3 Dynamic Power Control Scheme

If set with the right power, SCNs can play a great role in the existing network, especially for the hotspot areas. In this section, we use the location based dynamic power control scheme to get small cells' just right transmit power. The transmit power should be just enough to serve the coverage area, not too big or too small. Thus, we may get less co-interference and more network capacity.

The purpose of the power control scheme is to achieve better network throughput and HO performance. Herein, we propose a location based sequential dynamic power control scheme. Since there is no overlap between the small cell clusters in a macrocell, we choose the small cells in one cluster as illustration and the power assignment for small cells in other clusters can be obtained as the same way [6].

The power control scheme first calculates the total interference each small cell suffers and then sequence them in order. To decrease the algorithm complexity, our scheme begins with the heaviest interfered small cell. Let BS1 be the heaviest interfered small cell and BS2, BS3 and BS4 be its neighbor cells. We can see from Fig. 2 that there are several coverage intersections between those four small cells. Based on the intersections and the original coverage area of BS1, we can obtain the central area of BS1. If there are no users in the central area, the small cell will transit from active mode to sleep mode and its power can be assumed as zero. Meantime, the users

operating at the edge of the small cell will hand over to the cell from which their received power is second largest. Otherwise, when there is data transmission in the central area, the network will find the user whose distance to the small cell is largest. Assume the distance is $d(BS_i)$, the power setting $P(BS_i)$ of the small cell can be calculated as follows:

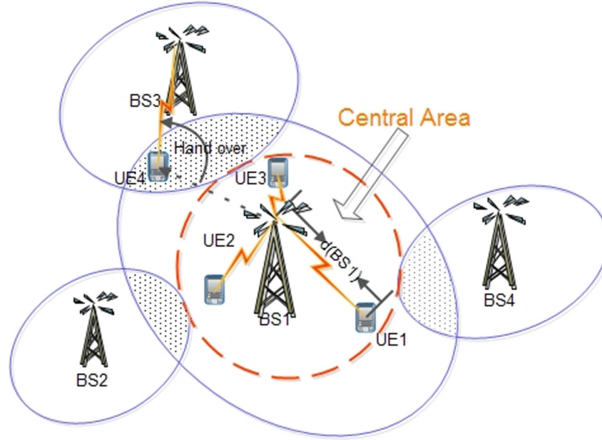


Fig. 2. Location based sequential dynamic power control scheme

$$P(BS_i) = \text{Median}(PL(d(BS_i)) + P_S(BS_i) + P_{offset}, P_{\max}(BS_i), P_{\min}(BS_i)) \quad (6)$$

Where $PL(d(BS_i))$ is the path loss of a certain distance $d(BS_i)$, which is the range between the small cell and farthest user. This parameter is proposed to adjust the coverage of the small cell so that it can appropriate satisfy the demands of users. $P_S(BS_i)$ is to make sure that the intra-layer interference is compromised. P_{offset} can be a predetermined value and is based on the penetration loss of walls. P_{offset} is set to be 0 dB for outdoor users and 20 dB for indoor users. $P_{\max}(BS_i)$ is the maximum power of the small cell [7]. And $P_{\min}(BS_i)$ is the lower boundary of $P(BS_i)$.

Our dynamic power control scheme sequentially operates above procedures as depicted in Fig. 3. During one iteration, the scheme extracts the users in the central area. If no users exist, the small cell will be shut down. Otherwise, it will adjust the transmit power. Consequently, the transmit power for all small cells will be just large enough to cover its users and there will be little cell intersection. The algorithm terminates until all the small cells have finished the power assignment.

```

1: Let  $S$  be the set of site indices
2: Let  $P$  be the array of assigned power
3: Let  $N$  be the array of the number of users in the small
   cell central area
4: Let  $d$  be the array of the distance between the small cell
   and farthest user in the central area
5: for all  $BS_i$  in  $BS$  do
6:    $i \leftarrow \arg \max_{1 \leq i \leq n} P_{interf}(BS_i)$ 
7:   Add index  $i$  into  $S$ 
8: end for
9: while  $S$  is not empty do
10:  target  $\leftarrow S(1)$ 
11:  Find the central area of  $S(\text{target})$ 
12:  if  $N(\text{target}) = 0$  do
13:     $P(\text{target}) = 0$ 
14:  else do
15:     $P(\text{target}) \leftarrow d(\text{target})$ 
16:  end if
17:  Remove the target for  $S$ 
18: end while
19: End

```

Fig. 3. Algorithm for location based dynamic power control scheme

4 Big Data Analysis Results

According to the small cell parameters given in [8], we execute ten thousands random instantiations. And Based on the parameters, we analysis our power control scheme's performance. The network data rates and cell edge users' throughput are collected and used to show the scheme performance. The article also analyzes the outage probability to illustrate the network edge users' improvement.

The following three schemes are used for comparison. Conventional power allocation: The same transmitting power is assigned to all small cells [1]. Sleep mode based algorithm: When there is no user in the small cell, it will transit into sleep mode and the power will be set at 0 [9]. The proposed power control scheme in the essay: Find the central area according to the location of the small cell users. And allocate the transmit power from the farthest user first, then proceed other users' power assignment in turn.

Figure 4 shows the network data rates and cell edge users' throughput of various small cell densities. We can see that the best network capacity is obtained with our dynamic power control scheme. Although the implementation of our power scheme provides little enhancement over the sleep mode based algorithm in the system data rates aspect, it is important to observe that the cell edge user data rates have a

significant improvement, i.e. at most 32.2% over the sleep mode based algorithm and 42.7% over the conventional power allocation. This is because the intersections between small cells are almost eliminated. Thus, cell edge users get less interfered, making the data rates enhanced greatly.

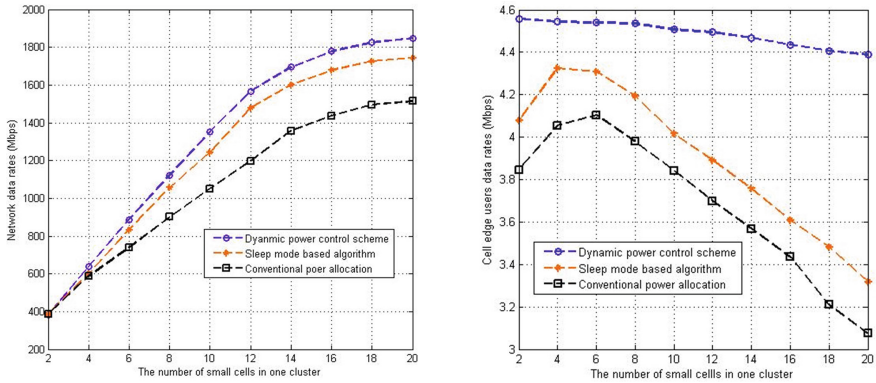


Fig. 4. Data rates for the network and cell edge users

Figure 5 illustrates the outage probability for the network. The network outage probability is defined as the user ratio who receives SINR < 0 [10]. The figure shows that our scheme has the lowest outage probability. Since the dynamic power control scheme highly values the cell edge users and remarkably enhance the network performance, the QoS of the users improves, leading to a decrease of drop users.

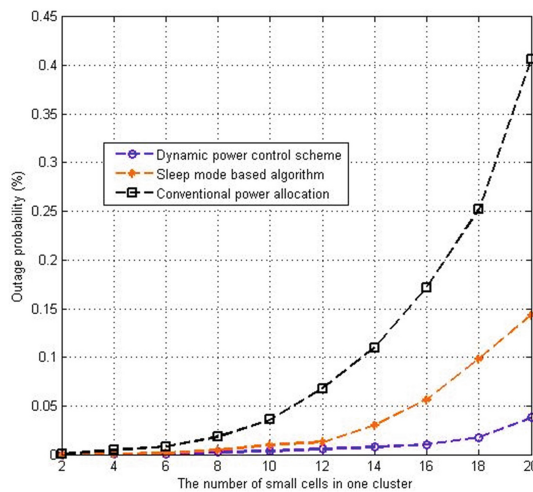


Fig. 5. Outage probability for the network

5 Conclusion

This paper introduces a novel dynamic power control scheme for small cell networks. The scheme assigns actual power to the small cell according to its received intra-layer interference and the location of its farthest users. Since the intra-layer interference of cell edge users is highlighted, their QoS is well guaranteed and the drop users in the network are sharply declined. Besides, with the tight control of the transmit power, the scheme can provide high energy savings. So we draw the conclusion that the proposed power control scheme is a meaningful strategy.

References

1. Federico, B., Lester, H.: Sleep mode techniques for small cell deployments. *Commun. Mag. IEEE* **49**(8), 72–79 (2011)
2. Hoydis, J., Kobayashi, M., Debbah, M.: Green small-cell networks. *Veh. Technol. Mag.* **6**(1), 37–43 (2015)
3. Brau, M., Corre, Y., Lostalen, Y.: Assessment of 3D network coverage performance from dense small-cell LTE. In: *IEEE International Conference Communications (ICC)*, pp. 6820–6824, June 2014
4. 3GPP TS 36.213V 12.5.0: Physical Layer Procedures, March 2015
5. 3GPP TR 36.942V 10.2.0: E-UTRA Radio Frequency System Scenarios, September 2012
6. Ge, C., Wang, N., Skillman, S., Foster, G., Cao, Y.: QoE driven DASH video caching and adaptation at 5G mobile edge. In: *3rd ACM Conference on Information-Centric Networking*, pp. 237–242. ACM Press, Kyoto (2017)
7. Xu, L., Luan, Y., Cheng, X., Xing, H., Liu, Y., Jiang, X., Chen, W., Chao, K.: Self-optimised joint traffic offloading in heterogeneous cellular networks. In: *16th IEEE International Symposium on Communications and Information Technologies*, pp. 263–267. IEEE Press, Qingdao (2016)
8. Xu, L., Cheng, X., Liu, Y., Chen, W., Luan, Y., Chao, K., Yuan, M., Xu, B.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: *16th IEEE International Conference on Communication Technology*, pp. 806–812. IEEE Press, Hangzhou (2015)
9. Kelly, F.: Charging and rate control for elastic traffic. *Eur. Trans. Telecomm.* **8**, 33–37 (1997)

Big Data Based Recommendation Scheme in APP Marketing Field

Yuwei Jia^(✉), Kun Chao, Lijuan Cao, Mingjun Mu,
and Xinzhou Cheng

China Unicom Network Technology Research Institute, Beijing 100048, China
jiaywl8@chinaunicom.cn

Abstract. In the traditional marketing activities, the key to success is to identify the potential needs of target users. An excellent recommendation system can help improve the effectiveness and efficiency of the precision marketing. In this paper, an Apriori algorithm based recommendation scheme is proposed, and it's then applied to the mobile APP marketing field. The scheme is realized through two processes, the frequent item sets generation process and the strong association rule generation process. Experimental results have shown that the scheme can achieve high advertising arrival rate, as well as superior exposure and click conversion rate. The effectiveness of the mobile APP product promotion has been improved dramatically.

Keywords: Big data · Association rule · Apriori algorithm

1 Introduction

With the improvement of enterprise informatization in all walks of life, people are living in an era of information explosion [1, 2]. How to make full use of the enormous information to assist the marketing strategy has become the research concern of many enterprises. According to Weber in [3], big data in enterprises includes user commercial transaction data, user behavior data, user social relationship data, and user temporal and spatial data. Large-scale personalized marketing will be a core competitiveness for enterprises. At present, data collection has arrived at a relatively high level, therefore how to integrate the data resource and transform it into applicable schemes has become a key task for enterprise operators.

One way to bring precision marketing into reality is the recommendation system, and the most critical part of the recommendation system is the recommendation algorithm. Under the circumstance of massive data, the efficiency and effectiveness to match target items with target users has become an important merit to measure the performance of the recommendation algorithm. An excellent recommendation algorithm will filter the most relative information and present it to the users. At present, the mainstream recommendation algorithms include: recommendation based on user analysis, recommendation based on content analysis, and recommendation based on association rules generation. They are described respectively in the following paragraphs.

Recommendation Based on User Analysis

As the name implies, user based recommendation algorithm relies on the basic information of the existing users and potential users, and then tries to find the relevance between them, so that preferable items of the existing users can be recommended to the potential users according to their relevance. This is the simplest recommendation algorithm. User attributes are the basis of the algorithm, which include user's age, gender, interest, and preference etc., then the similarity calculation model is built based on those attributes. For example, the system finds that user A and user C have high degree of similarity, then it will recommend favorite items of user A to user C.

Recommendation Based on Content Analysis

Similar to user based recommendation mentioned above, the focus of content based recommendation is the items. The algorithm model is built according to the relevance between items rather than users. Through the analysis of the items bought or browsed by a user, the system will set up and update an item preference list for the user, so that it can directly recommend highly relevant items to the target user. For example, in a movie recommendation system, user A is fond of movie A. Then, based on the analysis of the movie content, movie A and movie C are highly relevant in actors, directors, language, style, etc., so that movie C is recommended to user A as a potential choice.

Recommendation Based on Association Rules Generation

Association rule based recommendation algorithm is a machine learning method, which is aimed at identifying strong rules within a transaction database. Therefore, the internal relationship between items in a large item set can be discovered. For example, the strong rule {onions, tomato} \rightarrow { hamburger} implied in the supermarket transaction database would indicate that if a customer buys onions and tomatoes together, then they are likely to buy hamburgers as well. Such information can be employed as the basis of a series of marketing activities.

Each of these algorithms mentioned above has advantages and disadvantages. User based recommendation and content based recommendation are relatively less complex compared to the association rules based recommendation. However, these two algorithms either consider the similarity of users or the relevance between items, they haven't considered the influence of users' historical choice to the users' potential choice. Association rules based recommendation can identify the strong rules among items in the database, which also indicates the potential choice of users. In this paper, it adopts the association rules based recommendation algorithm to implement the recommendation system, and then applies the system to the precision marketing in mobile APP promotion field.

The following of this paper is organized as follows. Section 2 illustrates the basic principal of the Apriori algorithm as well as the process to generate frequent item sets and strong association rules. This will be the foundation of the recommendation system. Then in Sect. 3, it describes in detail the application of the recommendation system in mobile APP marketing field. In Sect. 4, the experimental results are demonstrated to prove the efficiency and effectiveness of the recommendation system. Finally, conclusions are drawn in Sect. 5.

2 The Apriori Algorithm

The Apriori algorithm is the basis of the association rules based recommendation. The core of this algorithm is the frequent item sets generation and association rules generation. It works in the way that the system will find the small frequent items in the database and then extend the small frequent items to larger item sets as long as they appear with high frequency in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database [4].

2.1 Basic Concepts of the Apriori Algorithm

In order to demonstrate the process of finding interesting association rules from the database, a series of important indicators involved in the algorithm are illustrated first.

Suppose that, $I = \{i_1, i_2, \dots, i_m\}$ is a full set containing m items. The set $X = \{i_1, i_2, \dots, i_k\}$ is a subset of $I = \{i_1, i_2, \dots, i_m\}$, which contains k items ($k \leq m$), and it's also called k -item set. Let $D = \{d_1, d_2, \dots, d_k\}$ be a set of transactions, which is called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I . A rule is defined as an implication of the form $X \rightarrow Y$ ($X, Y \subseteq I$), which means X is antecedent and Y is consequent [5].

The indicator "support" is used to measure how frequently the item set appears in the database. And it's can be interpreted as

$$support_{(X)} = occur_{(X)}/count_{(D)} = P_{(X)} \quad (1)$$

Where $occur_{(X)}$ is the occurring times of X in the database D , and $count_{(D)}$ is the transaction number of database D . In essence, the "support" of X is the probability of X in database D .

The indicator "confidence" is used to reflect how often a rule has been found to be true. The expression of "confidence" is

$$confidence_{(X \rightarrow Y)} = support_{(X \cup Y)}/support_{(X)} = P_{(X \cup Y)}/P_{(X)} = P_{(Y|X)} \quad (2)$$

Where $support_{(X \cup Y)}$ can be understood as the probability of set X and set Y occurring at the same time, and $support_{(X)}$ is the occurring probability of set X alone. The "confidence" of the rule $X \rightarrow Y$ can be considered as the conditional probability of Y under the condition of X .

The indicator "lift" can be used to measure the recommendation effectiveness of the rule $X \rightarrow Y$. It is expressed as

$$lift_{(X \rightarrow Y)} = confidence_{(X \rightarrow Y)}/support_{(Y)} = P_{(X \cup Y)}/(P_{(X)}P_{(Y)}) \quad (3)$$

2.2 The Process of Frequent Item Sets Generation

In this process, the frequent item sets are defined as sets which satisfy the minimum “support” threshold in a database. Efficient search is possible using the downward-closure property (also called anti-monotonicity) which guarantees that for a frequent item set, all its subsets are also frequent and thus no infrequent item set can be a subset of a frequent item set [6, 7].

The process of finding frequent item sets can be interpreted as follows (Table 1):

Table 1. The process of finding frequent item sets

Input: a transaction database D ; the minimum support threshold Min_sup
Output: frequent item sets L_k in D ; candidate items sets C_k in D
Process pseudo code:
Let L_1 be the frequent item sets of length 1.
For $k = 1; L_k \neq \emptyset; k++$
C_{k+1} is the candidate item sets generated from L_k
For each transaction d in database D
Increase the count of all candidates in C_{k+1} that are contained in d
L_{k+1} is the frequent item sets generated from C_{k+1} with minimum support
End for
Return all the frequent item sets
End for

2.3 The Process of Association Rules Generation

After finding the frequent item sets, strong association rules are generated through the maximum frequent item sets.

Suppose the frequent item sets generated in the previous process is L_k . For each non empty proper subset of L_k , it’s denoted as S , and $L_k - S$ is the complementary set of S accordingly.

If the following expression is satisfied, then $S \rightarrow L_k - S$ can be called a strong association rule.

$$confidence_{(S \rightarrow L_k - S)} = support_{(S \cup L_k - S)} / support_{(S)} = P_{(L_k)} / P_{(S)} \geq min_conf \quad (4)$$

The strong association rule can be used to recommend potential items to users with more precision.

3 Implementation of the APP Recommendation Scheme

In the traditional marketing activities, the essence of precision marketing is to employ the historical data to acquire the characteristics of target users as well as attributes of items, so as to establish an association relationship between them.

In this section, it takes the APP install information of a user group, and then by employing the Apriori algorithm, it can help sellers get a good knowledge of user potential needs. Meanwhile, because users' needs are often vague and unclear, users can easily get what they really want even at an unconscious state.

Suppose that, the APP install information of a user group is stored in a database D . The detailed information of a single user is called a transaction, and it has a unique transaction id denoted as Tid . The APP name in this section is called items, and accordingly the sets of APPs are called item sets.

Figure 1 gives an illustration of the APP frequent item sets generation. And the minimum support in this illustration is set as 0.5.

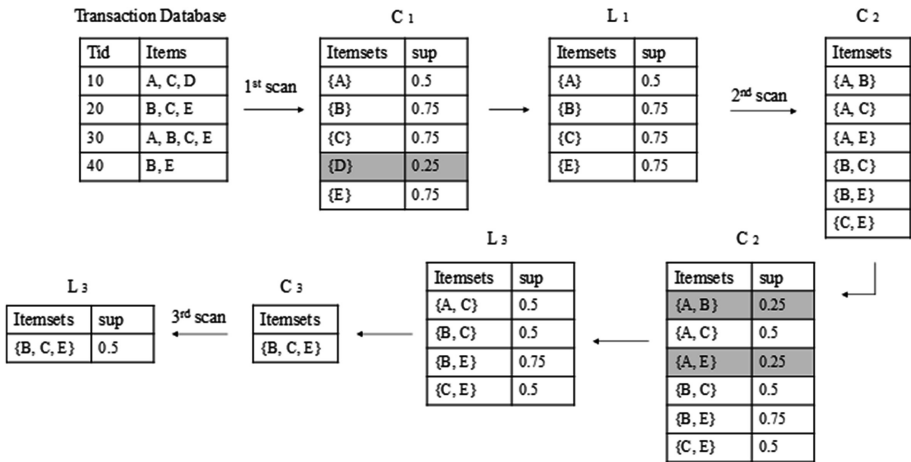


Fig. 1. Generation of frequent item set

From illustration above, it can be figured out that the combination of APP $\{B, C, E\}$ is the largest frequent item set, and the strong association rule will be derived from this item set.

The proper subsets of $\{B, C, E\}$ are $\{B\}\{C\}\{E\}\{B, C\}\{B, E\}$ and $\{C, E\}$. The possible association rules and their confidence are listed below.

$$confidence_{(B \rightarrow \{C, E\})} = support_{(\{B, C, E\})} / support_{(\{B\})} = 0.5 / 0.75 = 0.67 \quad (5)$$

$$confidence_{(C \rightarrow \{B, E\})} = support_{(\{B, C, E\})} / support_{(\{C\})} = 0.5 / 0.75 = 0.67 \quad (6)$$

$$\text{confidence}_{(E \rightarrow \{B, C\})} = \text{support}_{(\{B, C, E\})} / \text{support}_{(\{E\})} = 0.5 / 0.75 = 0.67 \quad (7)$$

$$\text{confidence}_{(\{C, E\} \rightarrow B)} = \text{support}_{(\{B, C, E\})} / \text{support}_{(\{C, E\})} = 0.5 / 0.5 = 1 \quad (8)$$

$$\text{confidence}_{(\{B, E\} \rightarrow C)} = \text{support}_{(\{B, C, E\})} / \text{support}_{(\{B, E\})} = 0.5 / 0.75 = 0.67 \quad (9)$$

$$\text{confidence}_{(\{B, C\} \rightarrow E)} = \text{support}_{(\{B, C, E\})} / \text{support}_{(\{B, C\})} = 0.5 / 0.5 = 1 \quad (10)$$

If the threshold of the minimum confidence is set as 0.8, then only rules $\{C, E\} \rightarrow B$ and $\{B, C\} \rightarrow E$ are filtered out as the strong association rules.

$$\text{lift}_{(\{C, E\} \rightarrow B)} = \text{confidence}_{(\{C, E\} \rightarrow B)} / \text{support}_{(B)} = 1 / 0.75 = 1.3 \quad (11)$$

$$\text{lift}_{(\{B, C\} \rightarrow E)} = \text{confidence}_{(\{B, C\} \rightarrow E)} / \text{support}_{(E)} = 1 / 0.75 = 1.3 \quad (12)$$

For the strong association rule $\{C, E\} \rightarrow B$, it means that if a user already possesses APP C and APP E, then he will adopt the promotion of APP B with high probability. The successful rate of recommending APP B to APP {C E} user groups is 30% higher than that of recommending APP B to the general crowd. The rule $\{B, C\} \rightarrow E$ works in the same way as $\{C, E\} \rightarrow B$, and it's omitted here for simplicity.

4 Experimental Results

In this section, the APP recommendation scheme is applied in a typical city in China, and it aims at the APP promotion and popularity improvement among the target users. During this process, the transactions, which are the APP install information of the target users, are stored in a database. Based on this information, the recommendation scheme employs the Apriori algorithm to find the frequent item sets, and this indicates the APP sets that have high probability of co-occurrence. Then according to the frequent APP sets, strong rules are generated under the constraint of minimum confidence. The strong association rules shows the direction of the precision marketing to target user groups.

Figure 2 illustrates the arrival rate of the APP recommendation scheme. Arrival rate is an important indicator about what percent of target users will receive the advertising information in a certain time period, and it is used to measure the availability of the recommendation scheme.

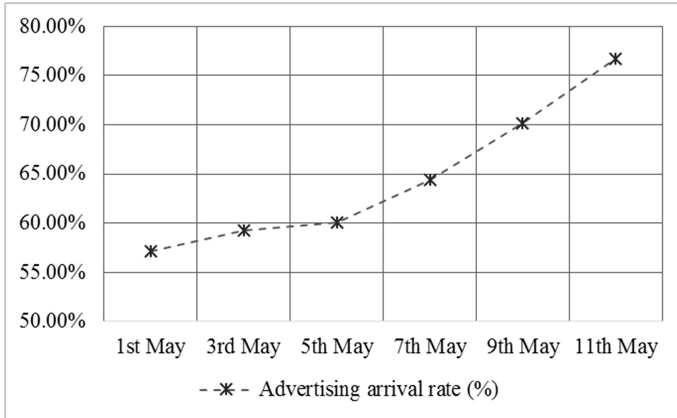


Fig. 2. Arrival rate of the APP promotion

Figure 3 shows another two important indicators, i.e. exposure conversion rate and click conversion rate. Exposure conversion rate refers to the ratio between the number of users who actually download the recommended APP and the number of users who receive the advertising information. While, click conversion rate is the proportion of users who actually download the recommended APP towards those who just click the advertisements without any other operation. These two indicators can be used to measure the effectiveness of the advertisements.

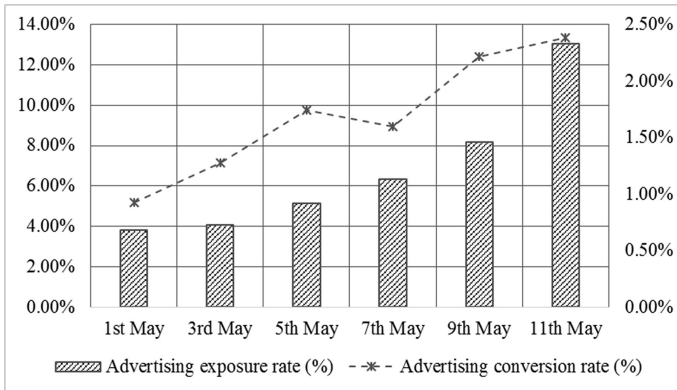


Fig. 3. Exposure/click conversion rate of the APP promotion

According to the experiment results shown above, effectiveness of the APP recommendation scheme is obvious. The advertising arrival rate, the exposure conversion rate and the click conversion rate are all dramatically improved. This precise APP

recommendation scheme helps APP developers save considerable advertising costs, while still achieve great effectiveness in product promotion and brand recognition.

5 Conclusion

This paper proposes an APP recommendation scheme, based on one of the classical recommendation algorithms, Apriori algorithm. It's realized through two processes, the frequent items sets generation process and the strong association rule generation process. By employing this APP recommendation scheme, the APP developers and promoters can identify users' preference of the APP combinations, thus recommend target APPs to target potential users. Experimental results have shown that the scheme can reach high advertising arrival rate, as well as superior exposure/click conversion rate. The effectiveness and efficiency of APP product promotion has been improved dramatically.

References

1. Xu, L., Cheng, X., Liu, Y., Chen, W., Luan, Y., Chao, K., Yuan, M., Xu, B.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: 16th IEEE International Conference on Communication Technology, pp. 806–812. IEEE Press, Hangzhou (2015)
2. Xu, L., Chen, Y., Gao, Y., Cuthbert, L.: A self-optimizing load balancing scheme for fixed relay cellular networks. In: 2th IET International Conference on Communication Technology and Application, pp. 306–311. IET Press, Beijing (2011)
3. Weber, L.: Marketing to the Social Web: How Digital Customer Communities Build Your Business, 2nd edn. Wiley, Hoboken (2009)
4. Ashish, S.: Association rule mining with modified Apriori algorithm using top down approach. In: 2nd International Conference on Applied and Theoretical Computing and Communication Technology, pp. 747–752. IEEE Press, Bengaluru (2016)
5. Feng, D., Zhu, L., Zhang L.: Research on improved Apriori algorithm based on MapReduce and HBase. In: 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, pp. 887–891. IEEE Press, Xi'an (2016)
6. Liu, C., Zhu, X., Nakata M.: On Apriori based rule generation in SQL a case of the deterministic information system. In: 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems and 2016 17th International Symposium on Advanced Intelligent Systems, pp. 178–182. IEEE Press, Sapporo (2016)
7. Ito, Y., Kato, S.: An Apriori-based approach to product placement in order picking. In: 2016 IEEE International Conference on Agents, pp. 114–115. IEEE Press, Buenos Aires (2016)

Research on Propagation Prediction Model Localization Over the Tropical Maritime Environment

Lianbo Song¹(✉), Xiaofei Chen¹, Jian Wang², Jialin Huang³,
Yafang Xu³, Chao Dou¹, and Cheng Yang²

¹ China Academy of Information and Communications Technology,
Beijing 100191, China
songlianbo@caict.ac.cn

² China Research Institute of Radiowave Propagation, Qingdao 266107, China

³ Bureau of Radio Supervision and Regulation of Hainan Province,
Haikou 570203, China

Abstract. It is a convention to verify the validity of a universal propagation model before its local application with the aid of mass local experimental data and make the universal propagation model further localized. This paper proposes detailed comparisons between theoretical predictions of field strength based on ITU-R P.1546 with the data collected from the experiments conducted on the South China Sea, which proved that the recommendation ITU-R P.1546 is valid in China tropical maritime region. Three optimization methods are also proposed for tuning the parameters from ITU-R P.1546 recommendation to improve accuracy of the propagation prediction in China tropical maritime region. At the end, future researches on tropical maritime propagation prediction are suggested.

Keywords: Propagation prediction · Tropical maritime
Propagation experiment · Localization research

1 Introduction

China is vast in territory, and its radio propagation environments vary in regions. The climate and geographical environment of the tropics are rather special compared to the non-tropics, characterized by low latitude, high temperature, wide sea area, complex landforms, rainy climate, and significant monsoon. The theoretical models of radio propagation in the non-tropics should not be applied directly to the tropics [1–4]. Therefore, tropical maritime radio wave propagation research is a research focus in recent years.

In the current research of the maritime radio wave propagation model home and abroad, two modelling methods are generally applied: empirical/semi-empirical model and deterministic model. The empirical/semi-empirical model is easily understandable and is more suitable for maritime radio propagation over a large covering area, while the deterministic model is more suitable for an accurate theoretical analysis of radio propagation characteristics over sea.

Existing researches of empirical statistics have found that the simulation based on Longley-Rice model is extremely close to the actual measurement data in the maritime environment [5, 6].

ITU-R P.1546 [7] has a narrower application scope than Longley-Rice, but it is equally applicable to maritime radio propagation of long distance transmitting in the VHF/UHF band. Unlike the Longley-Rice model adopting a mathematical formula form, ITU-R P.1546 introduces the propagation curves of cold and warm seas (e.g. the North Sea and the Mediterranean Sea), so that by looking up table and by interpolation/extrapolation, the radio wave pass loss can be calculated. ITU-R P.1546 also provides corrections applicable to different climate conditions. Therefore, ITU-R P.1546 also does well in analyzing the maritime radio propagation. The ITU-R P.1546 model is used in this study to predict the propagation of the maritime VHF/UHF band [8–11].

2 Tropical Maritime Radio Wave Propagation Experiment

Considering main maritime radio services, the frequency 160 MHz for wireless intercom and the frequency 830 MHz for mobile communications are selected as the key test frequencies. And according to the capacity of the test equipment, the nominal frequencies of 100 MHz and 600 MHz suggested by ITU-R P.1546 are selected. In addition, a fifth frequency of 1000 MHz (upper limit of the effective working frequency generated by the power amplifier used in this study) is also selected as a test frequency.

The transmitting site is set on the rooftop of a hotel located on the Sanya Bay coast line, and an azimuth angel of 210° is selected as the antenna direction. The test link established by the selected azimuth angel is not obstructed by any topographic conditions. The route of the receiving vessel is set to match the azimuth angel, which cuts through between the Western Isle and Easter Isle. Along the route there are 25 observing points, the first of which is located at the 8th Anchorage which is the nearest anchorage to the coast, and the furthest point is located 60 km away from the first one. As for the total 25 observing points, the rest are equidistantly distributed along the route. The transmitting site and observing points are illustrated in Fig. 1.



Fig. 1. Illustration of transmitting site and observing points

Given the particularity of the maritime test, measurements are carried out in two phases: horizontal polarization and vertical polarization. Average detection is used to collect the measurement data. Median field strength is calculated with one measurement per second and no less than 101 measurements for one test site.

Root mean square error (RMSE) is adopted to represent the dispersion of the random fluctuation of the receiving signal's field strength. It is generally accepted that the RMSE of a highly precise localized radio wave propagation model should be lower than 6 dB [12].

The RMSE representing the difference between the observed results and the predicted results by propagation model is:

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N (L_o - L_p)^2} \quad (1)$$

where N is the number of observing points, L_o is observed loss and L_p is predicted loss.

3 Analysis of the Research Results

Results of both horizontal and vertical polarized signals from the 25 observing points are gathered. The comparison is firstly used to analyze the applicability of the ITU-R P.1456 radio propagation model to the tropical South China Sea. Observation data and predicted data in two phases are shown in Figs. 2 and 3.

The results from horizontal and vertical polarizations show that:

Except for several observing points that produce deviated results, the test curves is generally smooth, complying with the propagation pattern of the radio wave.

The horizontally polarized test curve is smoother in general, but the data collected from Observing Point 13 deviates notably from the general pattern. From the on-site record, it is preliminarily concluded that the deviation results from a direction mismatch of the transmitting and receiving antenna caused by the vessel's turn due to random ocean currents. Data of Observing Point 4 and 5 also show a certain degree of deviation, but the reason needs to be analyzed with further tests.

During the vertical polarization test, multiple test points on the frequency 100 MHz have generated clear deviated results, and the statistics of test loss show a major fluctuation, while the data from the rest points comprise a comparatively smooth curve with no obvious deviation. The reason need to be analyzed with further tests.

The observed results and predicted results are compared in Table 1. The difference for the frequency of 100 MHz is greater than those of the other frequencies. The Mean Error and RMSE for frequencies other than 100 MHz are all under 5 dB. The overall predicted mean square error is 5.69 dB, which would be lower to 3.71 dB when excluding results for 100 MHz.

For both polarizations, prediction errors for the frequency of 600 MHz are the smallest. The RMSE of horizontal polarization is 4.57 dB, and that of the vertical polarization is only 1.39 dB.

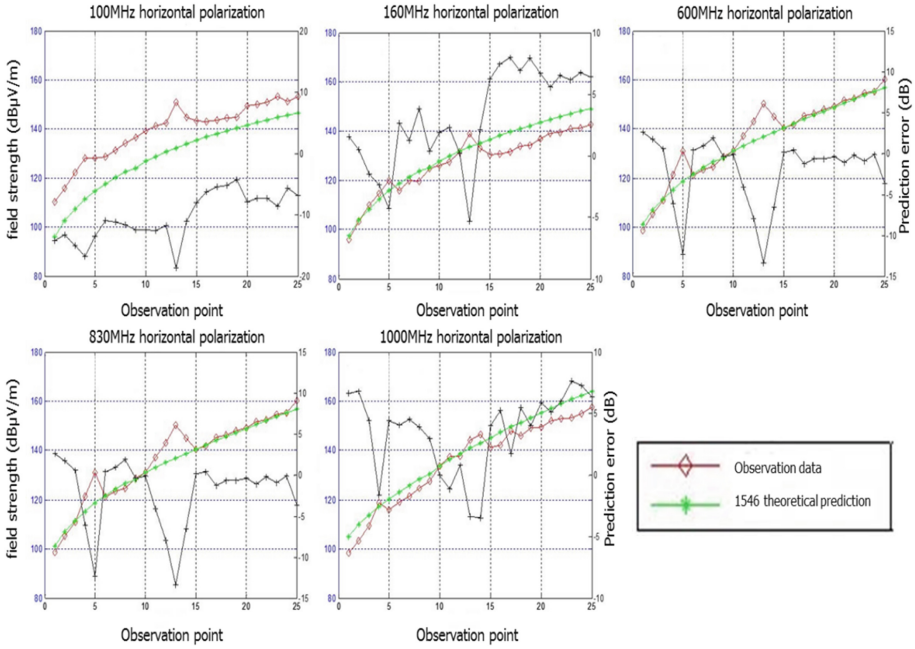


Fig. 2. Field strength comparison of horizontal polarization mode.

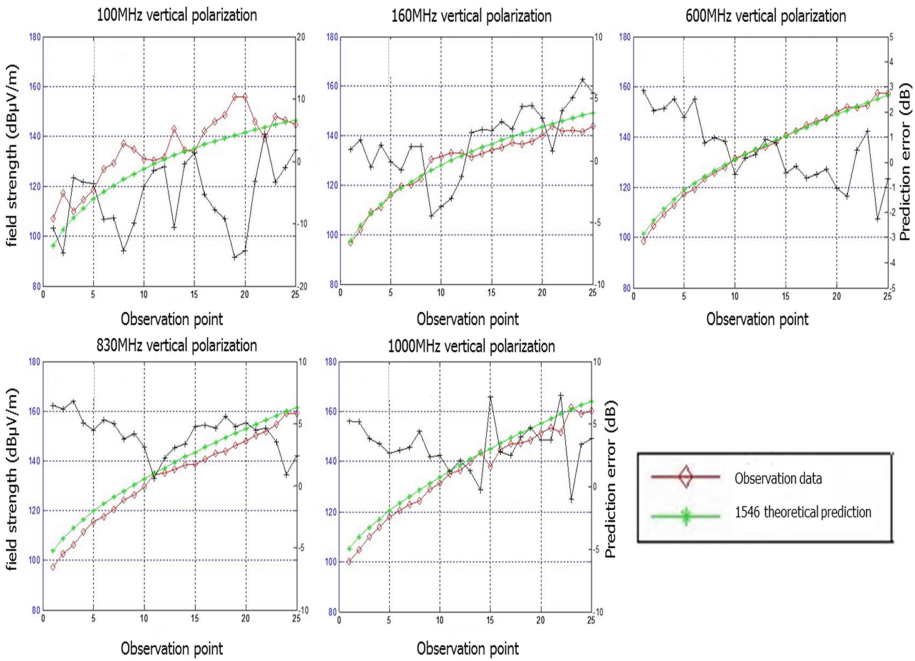


Fig. 3. Field strength comparison of vertical polarization mode.

Table 1. Path loss error of observed data and predicted data

Types of polarization	Prediction error	Mean square error of different frequency (MHz)					Total
		100	160	600	830	1000	
Horizontal	Mean error	-10.33	3.12	-2.04	3.47	3.50	-0.456
	Root-mean-square error	10.99	4.93	4.57	4.87	4.73	6.5155
Vertical	Mean error	-5.86	1.55	0.51	4.21	3.27	0.736
	Root-mean-square error	8.06	3.17	1.39	4.48	3.77	4.7131

The RMSE for all frequencies sums up to 6.52 dB in horizontal polarization mode, and 4.71 dB in vertical polarization mode, therefore vertical polarization is more accurate than horizontal polarization.

In horizontal polarization mode, when excluding the exceptional data of Observing Point 13, the RMSE will be lower by 0.39 dB to 6.13 dB.

By analyzing the experimental data, it is then can be concluded that:

- (1) The test curves are generally smooth, except for some deviations from several observing points. The measured loss value is monotonically decreasing with the increase of the distance, which conforms to the patterns of the maritime radio propagation. The results are therefore valid.
- (2) The overall trend of the prediction curves and the observed curves are consistent, and the predicted RMSE is below 6 dB, which is considered to be accurate. The prediction error for the nominal frequency of 600 MHz is the smallest, which verifies the applicability of the ITU-R P.1546 model to China tropical maritime region.

4 Suggestions on the Localization of ITU-R P.1546

4.1 Correction Based on Measured Meteorological Data

As the meteorological data such as atmospheric refractivity gradient data used in Sect. 3 of this paper are annual average provided by ITU, it is suggested that the data should be replaced with the local average atmospheric refractivity gradient in the summer. The comparison between the prediction errors calculated by ITU-R and local gradients is shown in Table 2.

Table 2 shows that the RMSE using Hainan summer statistics are slightly higher than that using ITU recommended gradient data, by 0.0003 dB and 0.0064 dB respectively for horizontal and vertical polarization modes. Therefore, it is suggested to use the actually measured refractivity data or statistics in the same region in analyzing the propagation [13].

Table 2. Prediction errors under different refractivity gradients

Type of polarization	Near-surface refractivity gradient	Mean square error of different frequencies (MHz)					Root mean square error
		100	160	600	830	1000	
Horizontal	-59.1 (ITU Statistics)	10.99	4.93	4.57	4.87	4.73	6.5155
	-65.00 (Hainan Summer Statistics)	11.03	4.90	4.59	4.85	4.70	6.5152
Vertical	-59.1 (ITU Statistics)	8.06	3.17	1.39	4.48	3.77	4.7131
	-65.00 (Hainan Summer Statistics)	8.08	3.14	1.39	4.44	3.73	4.7067

4.2 Introduction of Polarization Correction Factors

In the ITU-R P.1546 model, the influence of the polarization modes on radio propagation is not taken into consideration. According to the Sect. 3 of this paper, the modes of polarization can affect the path loss for a certain degree. Therefore, in the propagation analysis, it is suggested to introduce correction factors to locally correct the model. The correction factors are listed in Table 3.

Table 3. Field strength correction factors for different polarizations

Frequency (MHz)	100		160		600		830		1000	
Type of polarization	VP	HP	VP	HP	VP	HP	VP	HP	VP	HP
Correction factor (dB)	5.8	10.3	-1.5	-3.1	-0.5	2.0	-4.2	-3.5	-3.3	-3.5

4.3 A Correction of the ITU-R P.1546 Nominal Curve

The ITU-R P.1546 model shows the field strength versus distance curves for the frequencies of 100 MHz and 600 MHz in warm seas. According the observed data for the same frequencies and considering the correction factors such as the height of the transmitting and receiving antennas, and antenna height difference, an equivalent fitting curve ($h_1 = 20$ m, 50% time) is shown in Fig. 4.

Figure 4 shows that for both polarization modes for the frequency of 600 MHz, the equivalent fitting curves based on the observation data are consistent with the field strength versus distance curves provided in the ITU recommendation, while for the frequency of 100 MHz for both polarization modes, the equivalent data are generally smaller that nominal data. It is suggested that the ITU recommended nominal curve for the frequency of 100 MHz should be updated based on the actually measured results, or nominal field strength versus distance curve particular for China tropical maritime region should be accepted.

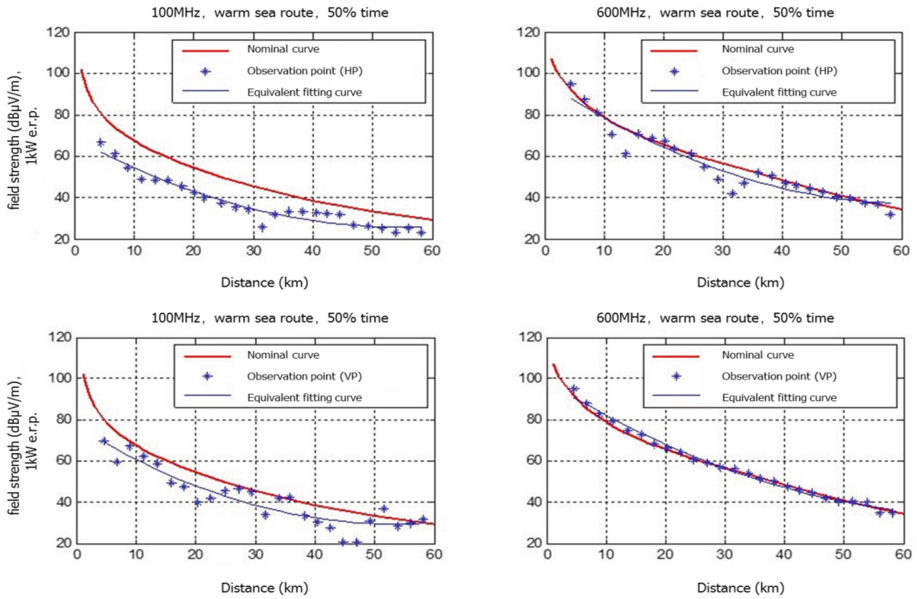


Fig. 4. Equivalent fitting curves for horizontal and vertical polarization modes

5 Conclusion

This study selects the ITU-R P.1546 model to carry out a research on the radio wave propagation in the tropical maritime areas. The research verifies the applicability of the model in the tropical South China Sea, and provides suggestion for localization. The conclusions are helpful for the study of the radio wave propagation characteristics over the tropical maritime environment, and established foundation for local application research of ITU-R P.1546 method in South China Sea. The results of this research are obtained through limited experiment data, thus provide a partial correction of the ITU-R P.1546 model in tropical maritime region. In the future when conditions permits, more broad verification experiments could be conducted in different seasons, using different system parameters and in different ways the experimental instruments are equipped.

References

1. Li, W., Yang, M., Dai, H.L.: Research progress and core factor analysis of tropical radio wave propagation. *Telecommun. Eng.* **9**(54), 1314–1320 (2014)
2. Xu, L., Cheng, X., Chen, Y., Chao, K., Liu, D., Xing, H.: Self-optimised coordinated traffic shifting scheme for LTE cellular systems. In: 1st EAI International Conference on Self-Organizing Networks, pp. 67–75. Springer press, Beijing (2015)

3. Xu, L., Chen, Y., Chai, K.K., Luan, Y., Liu, D.: Cooperative mobility load balancing in relay cellular networks. In: 2nd IEEE/CIC International Conference on Communication in China, pp. 141–146. IEEE press, Xi'an (2013)
4. Xu, L., Luan, Y., Cheng, X., Cao, X., Chao, K., Gao, J., Jia, Y., Wang, S.: WCDMA Data based LTE site selection scheme in LTE deployment. In: 1st International Conference on Signal and Information Processing, Networking and Computers, pp. 249–260. CRC Press, Taylor & Francis Group, Beijing (2015)
5. An, J.: Empirical analyses on maritime radio propagation. In: 59th Vehicular Technology Conference, pp. 176–180. IEEE press, Milan (2004)
6. Kim, K.B., Ali, M., Lee, J.H., Park, S.O.: Experimental study of propagation characteristic for maritime wireless communication. In: 2012 International Symposium on Antennas and Propagation, pp. 1481–1484. IEEE press, Nagoys (2012)
7. Recommendation ITU-R P.1546-4: Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3 000 MHz. ITU press, Geneva (2009)
8. Yang, K., Røste, T., Bekkadal, F., Ekman, T.: Channel characterization including path loss and Doppler effects with sea reflections for mobile radio propagation over sea at 2 GHz. In: 2010 International Conference on Wireless Communications & Signal Processing, pp. 1–6. IEEE press, Suzhou (2010)
9. Yang, K., Røste, T., Bekkadal, F., Ekman, T.: Channel characterization of mobile radio channel over sea at 2 GHz. In: 10th Mediterranean Microwave Symposium, pp. 389–392. IEEE press, Guzelyurt (2010)
10. Yang, K., Røste, T., Bekkadal, F., Husby, K., Trandem, O.: Long-distance propagation measurements of mobile radio channel over sea at 2 GHz. In: 2011 IEEE Vehicular Technology Conference, pp. 1–5. IEEE press, San Francisco (2011)
11. Wang, W., Hoerack, G., Jost, T., Raulefs, R., Walter, M., Fiebig, U.-C.: Propagation channel at 5.2 GHz in baltic sea with focus on scattering phenomena. In: 9th European Conference on Antennas and Propagation, pp. 1–5. IEEE press, Lisbon (2015)
12. Chen, B., Wang, J., Yang, H.M., Zhang, J., Sun, H.B.: Analysis and local application research on propagation prediction method for analogue FM broadcasting. *Chin. J. Radio Sci.* **2**(28), 378–384 (2013)
13. Li, L., Wu, Z.S., Lin, L.K., Zhao, Z.W., Zhang, S.B., Guo, X.M.: Study on the relativities of the tropospheric microwave trans-horizon propagation above ocean surface and the marine atmospheric environment characteristics. *J. Electron. Inf. Technol.* **1**(38), 209–215 (2016)

A Novel Architecture and Machine Learning Algorithm for Real Estate

Chen Cheng^(✉), Xinzhou Cheng, Mingqiang Yuan, Kun Chao,
Shiyu Zhou, Jie Gao, Lexi Xu, and Tao Zhang

China Unicom Network Technology Research Institute, Beijing 100048, China
chengc40@chinaunicom.cn

Abstract. The real estate industry is a hot topic and the factors of a house which affect the investment benefit is worth of research. This paper designs a novel machine learning assisted real estate industry investment guidance (MLRIG) architecture and a machine learning algorithm, aiming at researching the factors and their weight respectively of a house which have influence on its investment value. The MLRIG architecture is composed of 4 stages: Data collection, Data discretization, Data Mining Process and Factors weight output; the proposed machine learning algorithm, called QSFL-LR (Quantum-inspired Shuffled Frog Leaping Logistic Regression), combines Quantum-inspired Shuffled Frog algorithm with Logistic Regression to select the factors of a house which affect the investment value before data training, then output the weight of the factors respectively. Experiment shows the proposed QSFL-LR algorithm has better performance in accuracy and precision compared with traditional Logistic Regression, proving the superiority of QSFL-LR. The experiment also shows MLRIG architecture can guide both business companies and individuals to reduce investment risk in real estate industry.

Keywords: Machine learning · Intelligent algorithm · Logistic Regression
Big data

1 Introduction

In recent years, real estate industry has been a hot business for individuals and companies. How to judge the investment value of the residence through its factors is a hot research topic [1, 2]. As we know, a real estate industry cycle contains rising period and falling period, so choosing residence for investment benefit with some factors which can be more stable in the failing period is worth of research. However, this is a complex problem which is closely related to local economic development, monetary policy, regulatory property, housing quality, location, etc. [3]

The continuous development of machine learning and intelligent algorithm provide good conditions for the study of this problem. This paper presents a novel machine learning assisted real estate industry investment guidance (MLRIG) architecture which can guide both business companies and individuals to reduce investment risk.

The rest of the paper is organized as follows. In Sect. 2.1, we designs a novel machine learning assisted real estate industry investment guidance (MLRIG)

architecture; in Sect. 2.2, the Residence Portrait is proposed; in Sect. 2.3, the data is discretized for the preparation of machine learning algorithm. In Sect. 3.1, the Shuffled Frog Leaping Algorithm (QSFL) is proposed to choose factors for Logistic Regression; in Sect. 3.2, Quantum-inspired Shuffled Frog Leaping Logistic Regression (QSFL-LR) Algorithm is used to get the weight of each factor; in Sect. 3.3, the weight vector of factors is output. Finally, in Sect. 4, conclusions are presented.

2 MLRIG Architecture

2.1 System Model

This paper designs a novel machine learning assisted real estate industry investment guidance (MLRIG) architecture to combine an algorithm of machine learning called logistic regression and an intelligence algorithm called Quantum-inspired Leap Frog algorithm to judge the weight of each factor which can keep the property price stable during the downturn in the real estate cycle.

Figure 1 is the overview of MLRIG architecture containing 4 parts: Data collection, Data discretization, Data Mining Process and Factors weight output. Data Mining Process is consist of Model Building, Data Preparation, Feature Engineering, Model Training, Model Evolution and Model Deploy.

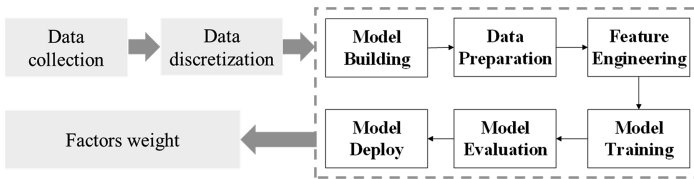


Fig. 1. Overview of MLRIG architecture

2.2 Residence Portrait

Residence Portrait is a list of factors containing 8 kinds of factors. Figure 2 illustrates the factors of the residence portrait.

- (a) P is the Price information, which includes total price, price per square meter respectively, represented as (1):

$$P = (P_1, P_2) \tag{1}$$

- (b) L is the Location information including administrative region and the number of circle respectively, represented as (2):

$$L = (L_1, L_2) \tag{2}$$

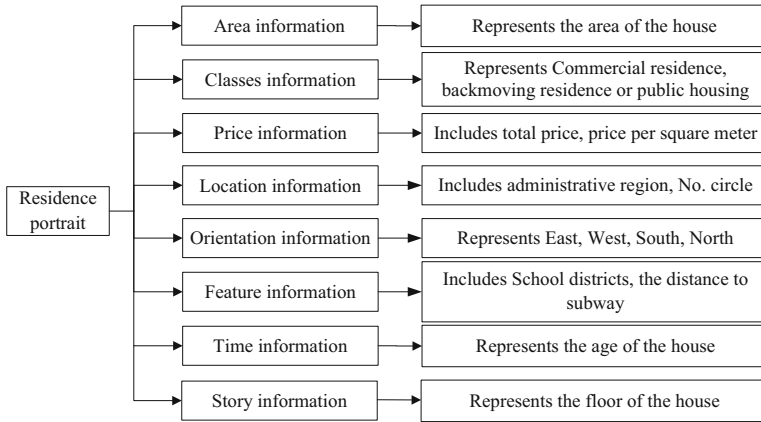


Fig. 2. Overview of MLRIG architecture

- (c) F is the Feature information, including school districts, the distance to subway, which is represented as (3):

$$F = (F_1, F_2) \quad (3)$$

- (d) A is the Area information, representing the area of the house. C is the Classes information, representing commercial residence, back moving residence or public housing. O is the Orientation information, represents East, West, South or North. T , the Time information, represents the age of the house. S represents the floor of the house.

The Residence Portrait is as (4):

$$R = (A, C, P, L, O, F, T, S) \quad (4)$$

2.3 Data Discretization

(1) Factors Discretization

In machine learning, many algorithms (such as logistic regression, decision tree, etc.) [4] are not suitable for continuous data. In order to improve the performance of the result, the factors need to be discretized [5]. There are three common methods to discrete the continuous variables: equidistant division, equal frequency division and statistic distribution [6]. In this paper, equal frequency division is used to discretize the continuous variables. For example, the total price (P_1) can be discretized as equal frequency division showing in Table 1:

Table 1. The example of factors discretization

Sub-factors	P_{1_0}	P_{1_1}	P_{1_2}	P_{1_3}	P_{1_4}
Value	0–32424	32425–54285	54286–74583	74584–98104	98105– $+\infty$

(2) Target Discretization

In the model of LR, the result is represented as a dichotomous variable ‘0’ or ‘1’ as (5):

$$x = \begin{cases} 1, & p > 0.5 \\ 0, & p \leq 0.5 \end{cases} \tag{5}$$

For the training data, the decrease magnitude of price per square meter is discretized as (6):

$$x = \begin{cases} 1, & \text{magnitude} > 21\% \\ 0, & \text{magnitude} \leq 21\% \end{cases} \tag{6}$$

where 21% is the average value of the decrease magnitude in the failing period of last real estate circle.

3 Using QSFL-LR to Determine the Weight of Features

3.1 Using Quantum-inspired Shuffled Frog Leaping to Choose Factors for Logistic Regression

In order to simplify the machine learning model and improve the performance, we need to select some factors from these 11 characters. This paper proposed an evolutionary algorithm, called QSFL (Quantum-inspired Shuffled Frog Leaping), which is fit for the Logistic Regression model to select some features [8].

The QSFL uses a pair of quantum bits to represent the location of the quantum frog. The i th frog is represented as (7):

$$\mathbf{d}_i = [\mathbf{d}_{i1} \quad \mathbf{d}_{i2} \quad \cdots \quad \mathbf{d}_{iM}] = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{iM} \\ y_{i1} & y_{i2} & \cdots & y_{iM} \end{bmatrix} \tag{7}$$

where $|x_{ij}|^2 + |y_{ij}|^2 = 1, (j = 1, 2, \dots, M), 0 \leq x_{ij} \leq 1, 0 \leq y_{ij} \leq 1$.

We use $U(\theta_{ij})$ to update \mathbf{d}_{ij} :

$$\mathbf{d}_{ij}^{new} = \text{abs}(U(\theta_{ij})\mathbf{d}_{ij}) = \text{abs}\left(\begin{bmatrix} \cos \theta_{ij} & -\sin \theta_{ij} \\ \sin \theta_{ij} & \cos \theta_{ij} \end{bmatrix} \mathbf{d}_{ij}\right) \tag{8}$$

simplified as (9):

$$x_{ij}^{new} = |x_{ij} \cos \theta_{ij} - \sqrt{1 - (x_{ij})^2} \sin \theta_{ij}| \tag{9}$$

Assuming an M -dimensionality space with p frogs, and each frog is an M -dimensionality vector as (10):

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iM}] \tag{10}$$

The step of QSFL is as following:

Step 1: Compute the fitness value of each frog, then sort them in descending order.

Step 2: Groups the p frogs into m groups, and each group has n frogs according to their fitness value. Set the local evolution times in each group as *generation*, and set the global evolution times as *GENERATION*.

Step 3: Update the worst fitness value of each grog as (11) and (12):

$$\theta_{wj} = a_1 * rand \cdot (x_{wj} - x_{bj}) + a_2 * rand \cdot (x_{wj} - x_{gj}) + a_3 * randn \cdot (x_{wj} - z_j) \tag{11}$$

$$x_{wj}^{new} = |x_{wj} \cos \theta_{wj} - \sqrt{1 - (x_{wj})^2} \sin \theta_{wj}| \tag{12}$$

where *rand* is Uniform distribution between 0 and 1, *randn* is Gaussian distribution, x_{bj} is the local best solution, and x_{gj} is the global best solution.

z is the average location of these p frog which is represented as (13):

$$\mathbf{z} = [z_1, z_2, \dots, z_M] \tag{13}$$

where $z_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, j = 1, 2, \dots, M$.

Step 4: $i_m = i_m + 1$. (14)

Step 5: If $i_{generation} < generation$, $i_{generation} = i_{generation} + 1$, then go to step 3; else $i_{GENERATION} = i_{GENERATION} + 1$, then go to step 2.

Step 6: If $i_{GENERATION} = GENERATION$, output the result.

The Griewank Function and Ackley Function are used to test the performance of the algorithm compared with BFOA (Bacterial Foraging Optimization algorithm) [9], PSO (Particle Swarm Optimization) and SFLA (Shuffled Frog Leaping Algorithm) [10], represented in Figs. 3 and 4, proving QSFL has higher accurate convergence.

In order to simplify the machine learning model and improve its performance, QSFL is used to choose factors of a house which affect the decrease magnitude of price per square meter a lot.

The coefficient vector of the 11 features is as (15):

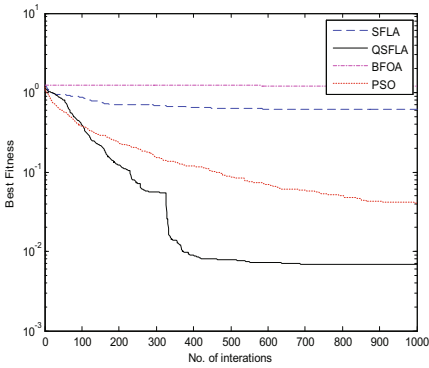


Fig. 3. The performance test using Griewank function

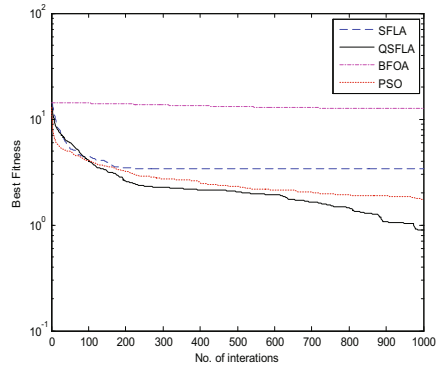


Fig. 4. The performance test using Ackley function

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}) \tag{15}$$

which is determined by the solution of QSFL as (5):

Hence, the appropriate factors whose coefficients are 1 will be selected as the factor used in next step.

3.2 Using Logistic Regression to Get the Weight of Residence Features

Logistic Regression (LR) is a kind of machine learning method to solve the problem of dichotomy. The probability of ‘1’ is computed as:

$$y = \theta^T \mathbf{x} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n \tag{16}$$

where θ is the weight of the factors, and \mathbf{x} is the vector of residence portrait selected in Sect. 3.1.

The model of LR is as (18) and (19):

$$P(Y = 1|x) = \frac{1}{1 + e^{-g(x)}} \tag{18}$$

$$P(Y = 0|x) = 1 - P(Y = 1|x) = 1 - \frac{e^{g(x)}}{1 + e^{g(x)}} = \frac{1}{1 + e^{g(x)}} \tag{19}$$

where

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \tag{20}$$

We use a logarithmic function to convert P to $(-\infty, +\infty)$ as (21):

$$\log\left(\frac{P}{1-P}\right) = \log(e^{g(x)}) = g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \tag{21}$$

where β is the learning object and x is the factors.

The algorithm of obtain β is as (22):

$$L(\beta) = \prod P(y_i = 1|x_i)^{y_i} (1 - P(y_i = 1|x_i))^{1-y_i} \tag{22}$$

which is a kind of steepest descent.

3.3 Factors Weight Output

In order to simplify the machine learning model and improve the performance, QSFL is used to choose factors of the residence which affect the decrease magnitude of price per square meter a lot.

The Residence Portrait is as (23):

$$R = (A, C, P, L, O, F, T, S) = (A, C, P_1, P_2, L_1, L_1, O, F_1, F_1, T, S) \tag{23}$$

The coefficient of each feature is as:

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}) \tag{24}$$

which is determined by the solution of QSFL as:

$$\beta_i = \begin{cases} 1, & x_i > 0.5 \\ 0, & x_i \leq 0.5 \end{cases} \tag{25}$$

Using the proposed QSFL algorithm, where global evolution time is 40 and the local evolution time is 1000, the result of factors selection is as:

$$R' = (A_1, P_2, T, L_2, F_1) \tag{26}$$

where 6 factors, including area, price per square meter, the age of the house, administrative region and school districts and the age of the house are selected as the factors in machine learning.

After factors selection, Logistic Regression is used to get the result of the weight of residence features respectively as table:

$$R' = (A_1, P_2, T, L_2, F_1) = (-0.3214, -0.436, 0.237, 0.197, 0.435) \tag{27}$$

Compared with LR without factors selection of QSFL, the vector of weight is as:

$$\begin{aligned} R &= (A, C, P_1, P_2, L_1, L_1, O, F_1, F_1, T, S) \\ &= (-0.3214, 0.0942, -0.5242, -0.5132, 0.1021, 0.0832, -0.1321, 0.3124, 0.3253, -0.1231, 0.2534) \end{aligned}$$

Usually, we use 2 indexes, called accuracy and precision to measure the performance of the LR model [12]. Table 2 is the result of machine learning using QSFL-LR and traditional LR, which shows that the accuracy and precision of QSFL-LR are higher than traditional LR, proving the superiority of QSFL-LR.

Table 2. The result of machine learning using QSFL-LR and traditional LR

Algorithm	x	Accuracy	Precision
QSFL-LR	0	0.842	0.832
	1	0.761	0.781
Traditional LR	0	0.726	0.754
	1	0.713	0.723

4 Conclusion

This paper designs a novel machine learning assisted real estate industry investment guidance (MLRIG) architecture and a machine learning algorithm call QSFL-LR (Logistic Regression with Shuffled Frog Leaping Algorithm), aiming at researching the factors and their weight respectively of a house which have influence on its investment value. The MLRIG architecture is composed of 4 stages: Data collection, Data discretization, Data Mining Process and Factors weight output. Experiment shows the proposed QSFL-LR algorithm has better performance in accuracy and precision compared with traditional Logistic Regression, proving the superiority of QSFL-LR. The experiment also shows MLRIG architecture can guide both business companies and individuals to reduce investment risk in real estate industry.

References

1. Cao, J., Zhang, T., Zeng, Z., et al.: Multi-relay selection scheme based on quantum particle swarm optimization in relay networks. In: 15th Wireless Personal Multimedia Communications (WPMC), Beijing, pp. 648–652 (2012)
2. Cheng, C., et al.: Spectrum allocation based on data mining in heterogeneous cognitive wireless networks. In: 1st International Congress on Signal and Information Processing Networking and Computers, Beijing, pp. 315–324 (2015)
3. Xu, L., Cheng, X., Liu, Y., Chen, W., Luan, Y., Chao, K., Yuan, M., Xu, B.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: 16th International Conference on Communication Technology, pp. 806–812. IEEE Press, Hangzhou (2015)
4. Cheng, C., et al.: A novel cluster algorithm for telecom customer segmentation. In: IEEE International Symposium on Communications & Information Technologies, pp. 231–237 (2016)
5. Xu, L., Chen, Y., Chai, K.K., Luan, Y., Liu, D.: Cooperative mobility load balancing in relay cellular networks. In: 2nd International Conference on Communication in China, pp. 141–146. IEEE Press, Xi’An (2013)

6. Cheng, X., et al.: Big data assisted customer analysis and advertising architecture for real estate. In: 16th International Symposium on Communications and Information Technologies, pp. 263–267. IEEE press, Qingdao (2016)
7. Zhang, T., et al.: Joint relay selection and spectrum allocation scheme in cooperative relay networks. In: Proceedings of IEEE VTC, Seoul Korea, pp. 1–6 (2014)

Multi-index Evaluation Analysis of Region Network Development: A Cluster Empirical Study

Haina Ye¹(✉), Wensheng Li², Jian Guan¹, Xiaodong Cao¹,
Xinzhou Cheng¹, Mingqiang Yuan¹, and Kun Chao¹

¹ China Unicom Network Technology Research Institute, Beijing, China
yehn3@chinaunicom.cn

² China Information Technology Designing and Consulting Institute,
Beijing, China

Abstract. The evaluation of region development of network has been a widespread concern in recent years. However, network development of a region, due to its specific characteristics and application scenario, should have a tailor-made evaluation system. In this study, taking into account various factors in multiple fields, a multiple-index evaluation system is established. Then, a principal component analysis-based K-means clustering approach is proposed to address the analyzing problem with an acceptable complexity. A simulation experiment is implemented to verify the algorithm. The results can be used to compare the different areas telecommunication networks, and provide rational and effective suggestions for network planning and construction.

Keywords: Principal components analysis · Clustering algorithm
Regional evaluation

1 Introduction

With the growing popularity of communications, the evaluation of region development of network has been a widespread concern in recent years. The demand for providing better services and the relatively slowly increasing revenue in telecom operators puts a premium on an accurate evaluation of the network development. Previous work on development evaluation has mainly focused on economic or environmental scenarios [1–3]. However, network development of a region, due to its specific characteristics and application scenario, should have a tailor-made evaluation system.

In general, the evaluation analysis of network is utilized to support decision making and help better formulate region planning for the future [4–6]. In this aspect, clustering the group of the regions for horizontal comparing becomes a key issue. A wise classification can help decision makers to analyze, contrast and apply, which makes the development strategy or implement plan more efficient and scalable. Clustering algorithms have emerged as an alternative efficient learning tool to accurately analyze the huge volume of data generated in various applications [7]. Based on clustering analysis, the objective data can be identified from the dense areas and sparse areas, so that

the global distribution patterns and the data relationship can be found. In this way, a deeper knowledge and more profound understanding can be acquired from the data, which is helpful to give more rational suggestions for decision-making. However, the most common difficulties to have a perfect clustering results rely on three aspects: (1) how to choose the initial points; (2) how to choose the number of clusters; and (3) how to solve a high-dimensional problem.

In this work, rather than considering a certain characteristic or performance of the network, the network development is analyzed from a macro view. Various factors in multiple fields are taken into account to establish a multiple-index evaluation system. Then, a principal component analysis-based K-means clustering approach is proposed to address the analyzing problem with an acceptable complexity. The analyzing result can be used to provide rational and effective suggestions for network planning and construction. The contribution of this paper is several-fold.

- First, we consider the network development from a macro view. Multiple factors which have close influence in future network planning and construction are taken into account.
- Second, the multi-index evaluation system has a high dimensions, which is difficult for further analyzing. To address that, the principle component analysis (PCA) method is implemented to reduce the dimension of the proposed evaluation system.
- Third, addition to the multi-dimensional problem, another difficulty of the clustering is the choice of the number of cluster. In order to have a rational result, the Silhouette coefficient (SC) is counted for each option. In the simulation, we provide a reasonable grouping result.

The rest of this paper is organized as follows. In Sect. 2, focusing on the characteristics of network development, the multi-index evaluation system is given. Following in Sect. 3, the main algorithm and typical operation are shown. Simulation results are shown in Sect. 4. Finally, we present the conclusion in Sect. 5.

2 Multi-index Evaluation System and Data Source

In the index system, we consider multiple factors which have close influence in telecom network planning and construction. The proposed measurement analysis is detailed in Table 1.

The index system considers 5 types of factors and 14 parameters in total.

The population factor, the economic factor and the customer factor present the population basis, the economic development level and the telecommunication popularity of the region, respectively. The factors describe the background of the region, and to some extent, show the potential benefits of the network planning and construction in the region.

The climate factor has an underlying relationship with network planning and construction. For example in China, the climate of the north is colder than the south. In order to provide the winter protection, the external wall of the buildings in the north are

Table 1. Index system for network planning and construction

Factors	Parameters	Indexes
Population factor	Resident population	p1
	Urban population	p2
	City population density	p3
Economic factor	Gross regional product	p4
	Per capita disposable income nationwide	p5
	Per capita disposable income of urban households	p6
Customer factor	Number of mobile phone	p7
	Number of fixed line phone	p8
Climate factor	Geo-location	p9
Network factor	Number of base station	p10
	Number of small cell	p11
Service factor	Number of users	p12
	Voice traffic	p13
	Data traffic	p14

usually wilder than those in the south; thus, the penetration loss is higher, which could effects a lot in the network planning and construction [8].

The network factor and the service factor indicate the capability of a telecom operator. The former one shows the existing level of the network construction, and the later one describes the penetration rate of an operator.

In this work, due to the limitation of the data source, the objective is the provinces of China. The data are collected from the national bureau of statistics of China [9], the statistics of China Unicom in 2015.

3 The PCA-Based K-means Clustering Approach

In this section, the PCA-based K-means clustering approach is proposed. There are mainly three steps in the approach: (1) preprocessing, (2) dimensionality reduction, and (3) clustering.

3.1 Preprocessing Data

In order to enhance the veracity of the data collected from the real word, the preprocessing method is implemented first. In this step, raw feature vectors are transformed into a representation for further processing. The main idea is to scale the data within a range. For each feature vector, the Max-Min normalization is utilized to transform the data:

$$y_{i,j} = \frac{x_{i,j} - \min_{\forall i} x_j}{\max_{\forall i} x_{i,j} - \min_{\forall i} x_{i,j}} \quad (1)$$

where i and j denote the row and column of the data matrix, representatively. $x_{i,j}$ and $y_{i,j}$ represents the value before and after the transformation, representatively.

After this step, the value of the feature is scaled in the range of $[0, 1]$, which avoids a various magnitude for different types of value.

3.2 Dimensionality Reduction

The main purpose of dimensionality reduction techniques in machine learning and statistics is to reduce the number of variables and dimensions of dataset, so that valuable information or features can be extracted, which improves the interpretability of the data and enhances the exploratory data analysis. The principal component analysis (PCA) is a widely used linear dimensionality reduction method. PCA addresses the problem to find a low-dimensional embedding of the data point that best preserves their variance in the high-dimensional original space [10].

Let Y denote the data matrix, Q represent the principal directions, and U be the projected data point in the new space. PCA finds U and Q by solving

$$\min_{U,Q} \|Y - UQ^T\|^2 \quad \text{s.t. } Q^T Q = I. \quad (2)$$

Then, on the new directions, the projected data points enable reflect the most amount of information of the original values. It not only eliminates the correlation among features, but also decreases the dimensions of the dataset.

3.3 Clustering

On the new principal directions, the clustering step is performed. PCA relates closely to K-means clustering naturally [11]. Thus, in this work, the K-means clustering is used to find the cluster result. The purpose of the K-means clustering is to partition the data points U into K sets $\mathbf{S} = \{S_1, S_2, \dots, S_K\}$, then, the sum of squares within a cluster (i.e. variance) is minimized:

$$\arg \min_{\mathbf{S}} \sum_{k=1}^K \sum_{U \in S_k} \|U - c_k\|^2, \quad (3)$$

where c_k is the center of the cluster S_k .

Via several times of iteration, the algorithm would achieve a local optimum. In order to have a quasi-optimal result, at the beginning of the clustering step, the initial center points would be set based on previous experiences.

4 Experiment Results

Base on the data introduced in Sect. 2, experimental process is promoted by the three stages as we specified in Sect. 3, and then aiming to the network planning and construction application, the provinces in China are analyzed and clustered.

In Fig. 1, the amount of variance explained by the principal components is analyzed. As we can see, the more components in the new directions explain more variance of the original data. However, with the number large enough, the growth of the number does not increase much of the information any more. The first three principal components occupy 87.24% of the total variance of the original features. Therefore, the first three principal components, which is considered to contain most of information in the original data, are selected as the new dimensions.

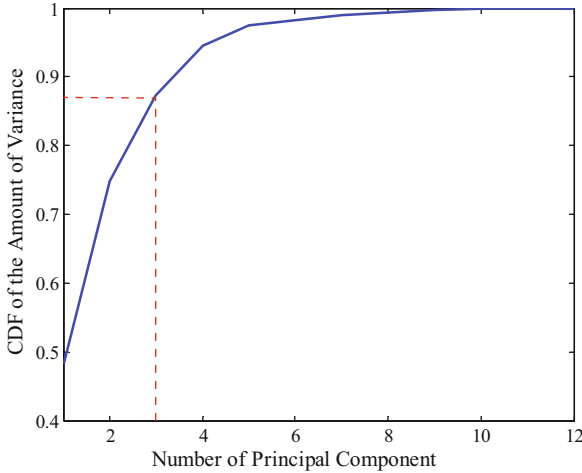


Fig. 1. CDF of the amount of variance explained by the principal components

Figure 2 presents the silhouette coefficient as a function of number of cluster. The silhouette coefficient is an index to describe the performance of a cluster [12]. The SC index used here is given as $SC = \frac{1}{N} \sum_i \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$, where $a(i)$ is the intra-cluster distance, and $b(i)$ is the nearest-cluster distance for each sample. A larger SC value means a better the clustering performance. As shown in the figure, when the number of cluster equals to five, the SC value is the largest. According to this, the given dataset is classified into five groups.

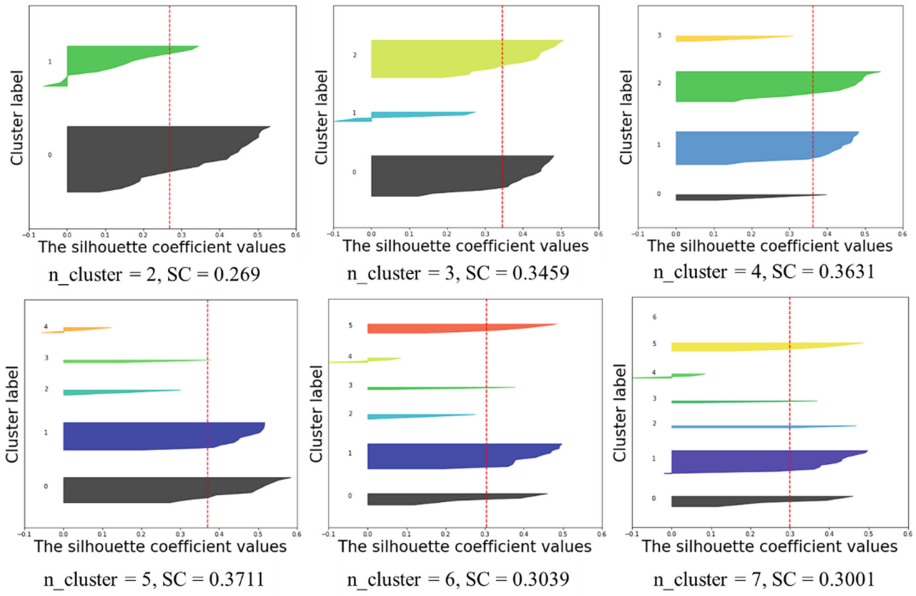


Fig. 2. The silhouette plot for the various clusters

Figure 3 and Table 2 illustrate the results of the PCA-based K-means clustering algorithm. In Fig. 3, each type of marker represents a cluster of provinces. The position of each point indicates each provinces in the direction of the principle components. The detailed classification of provinces in China is given in Table 2.

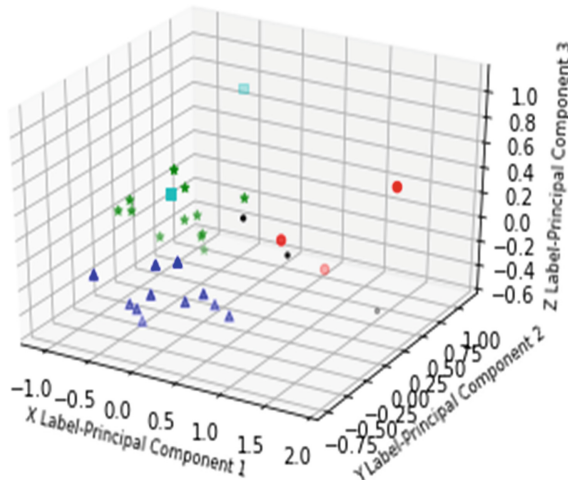


Fig. 3. Diagram of the clustering result

Table 2. The network development clustering result

Cluster name	Provinces
Cluster I	Beijing, Shanghai
Cluster II	Guangdong, Jiangsu, Zhejiang
Cluster III	Hebei, Henan, Shandong
Cluster IV	Anhui, Fujian, Guangxi, Guizhou, Hainan, Hubei, Hunan, Jiangxi, Sichuan, Yunnan, Chongqing
Cluster V	Gansu, Heilongjiang, Jilin, Liaoning, Neimenggu, Ningxia, Qinghai, Shanxi, Shan3xi, Tianjin, Xizang, Xinjiang

5 Conclusions

In this paper, a multi-index evaluation analysis of region network development is proposed based on machine learning techniques. The proposed evaluation system takes into account essential factors of the network development, such as the population factor, the economic factor, the customer factor, the climate factor, the network factor, and the service factor. Then, the PCA-based K-means clustering approach is proposed. The PCA stage is exploited for feature extraction, with the consideration of the silhouette coefficient of the dataset, the K-means stage is implemented for region classification. The simulation demonstrates that it is rational and effective to reflect the differences among regions. The results of the analysis can be used to compare the different areas telecommunication networks, and provide rational and effective suggestions for network planning and construction.

References

1. Zhang, W., Zhu, W., Zhang, H., et al.: Application of gray clustering method in the ecological classification of the cities in China. In: IEEE International Conference on Grey Systems and Intelligent Services, pp. 325–329 (2014)
2. Chen, Y., Tang, X., Shao, Y.: Cluster empirical study on the development of city circular economy of china based on factor analysis. *J. Ind. Eng. Manag.* **22**(4), 155–158 (2008)
3. Lao, X., Shen, T., Kong, Y.: The empirical research on China's city size distribution: an exploration based on the micro spatial data and city clustering algorithm. *J. Zhejiang Univ.* **45**(2), 120–132 (2015)
4. Chao, K., Wang, P., Xu, L., et al.: A novel big data based telecom user value evaluation method. In: 1st International Congress on Signal and Information Processing Networking and Computers, pp. 375–384, Beijing (2015)
5. Guan, J., Cao, L., Chen, W., et al.: A comprehensive method of evaluation for wireless network operation stability. In: IEEE International Symposium on Communications and Information Technologies, pp. 342–346. IEEE Press, Qingdao (2016)
6. Xu, L., Luan, Y., Cheng, X., et al.: Telecom big data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)
7. Fahad, A., Alshatri, N., Tari, Z., et al.: A survey of clustering algorithms for big data: taxonomy and empirical analysis. *IEEE Trans. Emerg. Top. Comput.* **2**(3), 267–279 (2014)

8. Xu, L., Luan, Y., Cheng, X., et al.: Self-optimised joint traffic offloading in heterogeneous cellular networks. In: 16th IEEE International Symposium on Communications and Information Technologies, pp. 263–267. IEEE Press, Qingdao (2016)
9. National data. <http://data.stats.gov.cn/>. Accessed June 2017
10. Shlens, J.: A tutorial on principal component analysis **51**(3), 219–226 (2014)
11. Allab, K., Labiod, L., Nadif, M.: A Semi-NMF-PCA unified framework for data clustering. *IEEE Trans. Knowl. Data Eng.* **29**(1), 2–16 (2017)
12. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**(1), 53–65 (1987)

Author Index

A

An, Gang, 452

B

Bai, Hongxing, 452

C

Cao, Lijuan, 373, 460, 475

Cao, Xiaodong, 500

Celik, Torecan, 189

Chang, Chen, 172

Chao, Cong, 238

Chao, Kun, 356, 389, 475, 491, 500

Chen, Dianjun, 3, 172

Chen, Lei, 189

Chen, Liang, 164, 265

Chen, Ping, 181

Chen, Weiwei, 327, 404, 420, 429, 444, 468

Chen, Xiaofei, 483

Chen, Yaqin, 197, 213

Chen, Yue, 436

Cheng, Chen, 365, 373, 389, 420, 491

Cheng, Kan, 272

Cheng, Xinzhou, 53, 327, 342, 356, 373, 389, 404, 420, 460, 475, 491, 500

Cui, Zihao, 272

D

Ding, Jianzhao, 281

Dong, Runsha, 327

Dou, Chao, 483

Du, XuYang, 309

F

Fan, Xingyu, 444, 468

Fang, Yuan, 350

G

Gao, Jie, 342, 356, 373, 460, 491

Gao, Pan, 221

Gao, Yan, 36, 350

Gao, Yang, 265

Guan, Jian, 342, 356, 500

Guo, Bao, 36, 350

Guo, Fengrui, 139

Guo, Guanggen, 452

Guo, Hao, 319

H

He, Tao, 272

Hu, Jinxia, 103

Hu, Mantian (Mandy), 334

Hu, Shanqing, 139, 147, 156

Hu, Xiaochun, 36, 350

Huang, Hai, 11, 80, 95

Huang, Jialin, 483

Huet, Alexis, 334

J

Jia, Yuwei, 420, 429, 460, 475

Jian, Zhuoru, 11

Jiang, Yu, 257

Jing, Xiaojun, 11, 65, 72, 80, 87, 95, 103, 197, 205, 213

Jon, Kwanhak, 112

L

Li, Guo, 290, 298

Li, Jia, 11, 65, 72, 80, 87, 95, 103, 197, 205, 213

Li, Ling, 397

Li, Mingxin, 319

Li, Ran, 309

Li, Wensheng, 342, 500

Li, Xiang, 290, 309

Li, Xingming, 139, 147, 156

Li, Yi, 468

Liang, Jialong, 397

Liao, Hongxi, 319
 Liu, Dayang, 36
 Liu, Fengjing, 290, 298
 Liu, Hongjie, 27, 129
 Liu, Jian, 298
 Liu, Jixiang, 36
 Liu, Ning, 290
 Liu, Tongjuan, 129
 Liu, Wei, 229
 Liu, Wenting, 197, 213
 Liu, Yi, 36, 350
 Liu, Yu, 436
 Liu, Yunhe, 298
 Lu, Jun, 444, 468
 Luan, Yuting, 356
 Luo, Qiang, 249
 Luo, Rong, 229
 Lv, Zhiqiang, 397

M
 Mao, Xu, 412
 Mu, Junsheng, 65
 Mu, Mingjun, 389, 429, 475

N
 Ni, Linna, 257
 Nie, Jingwen, 436

O
 Ouyang, Ye, 334

Q
 Qi, Yongjia, 397
 Qi, Zhaoqun, 129
 Qin, Yanyan, 139

R
 Ran, Meng, 365
 Ren, Baisong, 452

S
 Sang, Lin, 27
 Shen, Ao, 36, 350
 Shen, Jinhua, 350
 Shi, Hao, 164
 Shi, Mengjie, 272
 Song, Chuntao, 460
 Song, Hong, 189
 Song, Lianbo, 483
 Su, Fei, 327, 412
 Sun, Chenchen, 65

Sun, Songlin, 19, 46, 53, 112, 121, 221
 Sun, Xuebin, 3, 172, 181
 Sun, Zhaocong, 72

T

Tan, Juanjuan, 319
 Tang, Tianbiao, 319
 Tian, Tao, 53
 Tong, Xiaoyun, 19

U

Ullah, Yasir, 46

W

Wang, Chenwei, 53, 172, 221
 Wang, Guozhi, 36, 350
 Wang, Jian, 483
 Wang, Jibin, 334
 Wang, Jinpeng, 281
 Wang, Luhan, 365
 Wang, Pengfei, 389
 Wang, Yongfeng, 342, 429
 Wang, Yue, 281
 Wang, Yuwei, 147, 156
 Wu, Rihan, 46
 Wu, Wendong, 452

X

Xia, Xiuyan, 365
 Xian, Kuitong, 412
 Xiaoran, Men, 238
 Xie, Mingyu, 281
 Xie, Tao, 36, 350
 Xiong, Jie, 381
 Xu, Bingyu, 436
 Xu, Kaicheng, 412
 Xu, Lexi, 356, 373, 389, 404, 420, 436, 452, 491
 Xu, Yafang, 483
 Xue, Fei, 129

Y

Yang, Cheng, 483
 Yang, Guowei, 249
 Yang, Jian, 189
 Yang, Jianyi, 27
 Yang, Shan, 327
 Yang, XiaoYong, 309
 Yang, Yinding, 249
 Yang, Yucang, 468
 Yao, Saibin, 397

Ye, Haina, 342, 389, 420, 500
Ye, Jingyi, 205
Yin, Liang, 27
You, Siqing, 27, 129
Yu, Jiacheng, 147, 156
Yu, Kai, 381
Yu, Kuai, 298
Yu, Yanli, 356, 381
Yuan, Mingqiang, 491, 500

Z

Zeng, Zhen, 221
Zhan, Yi, 281
Zhang, Heng, 404, 420
Zhang, Kang, 112

Zhang, Qiaoqiao, 3
Zhang, Tao, 420, 491
Zhang, Wenwen, 87
Zhang, Xin, 27
Zhang, Xinran, 121
Zhang, Yang, 36, 350
Zhang, Yangying, 95
Zhao, Pan, 181
Zhao, Xueqing, 356, 404
Zhao, Zhongxi, 452
Zhou, Hao-tian, 164, 265
Zhou, Shiyu, 365, 491
Zhou, Xuan, 80
Zhuang, Yin, 164, 249
Zou, Yongheng, 298