Smart Innovation, Systems and Technologies 190

Roumen Kountchev · Aniket Mahanti · Shen Chong · Srikanta Patnaik · Margarita Favorskaya *Editors*

Advances in Wireless Communications and Applications

Smart Communications: Interactive Methods and Intelligent Algorithms, Proceedings of 3rd ICWCA 2019

KES International

Springer

Smart Innovation, Systems and Technologies

Volume 190

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/8767

Roumen Kountchev · Aniket Mahanti ·
Shen Chong · Srikanta Patnaik ·
Margarita Favorskaya
Editors

Advances in Wireless Communications and Applications

Smart Communications: Interactive Methods and Intelligent Algorithms, Proceedings of 3rd ICWCA 2019

Springer

Editors
Roumen Kountchev
Department of Radiocommunications
and Videotechnologies
Technical University
Sofia, Bulgaria

Shen Chong
College of Information Science
and Technology
Hainan University
Haikou, China

Margarita Favorskaya
University of Science and Technology
Moscow, Russia

Aniket Mahanti
Department of Computer Science
University of Auckland
Auckland, New Zealand

Srikanta Patnaik
School of Computer Science
and Engineering
SOA University
Bhubaneswar, Odisha, India

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-981-15-5696-8 ISBN 978-981-15-5697-5 (eBook)
https://doi.org/10.1007/978-981-15-5697-5

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

Preface

The fast development of contemporary computers, machine learning and artificial intelligence opened new horizons and possibilities for the accomplishment of the related methods, algorithms and network technologies. The major part of the development of these technologies takes the intelligent wireless communications, smart services and the new applications. The papers in Volume 2 of the proceedings were presented at the **3rd International Conference on Wireless Communications and Applications (ICWCA)** on December 21–23 (2019) in Haikou, Hainan, China. Focusing on applications of the latest smart theories and approaches, and the recent advances in the field, they cover topics, such as OFDM and multi-carrier techniques; smart antenna and space-time signal processing; MIMO, multi-user MIMO and massive MIMO; modulation, coding and diversity techniques; dynamic spectrum access and cognitive radio; interference management and radio resource allocation; equalization techniques; synchronization, estimation and detection techniques; and wireless multiple access (e.g. CDMA, OFDMA, NOMA).

Volume 1 "**Methods, Algorithms and Network Technologies**" of the proceedings comprises 28 chapters, which could be separated into the following three groups:

First group: Chapters 1–11. In these chapters are presented new methods and algorithms of the contemporary wireless communications, related to various approaches, such as: low complexity orthogonalization method for prolate spheroidal wave function signal; radio station distribution technology for transrapid system communication; node scheduling algorithm for solar-powered wireless sensor networks; universal software radio peripheral-based test bed of multi-agent reinforcement learning for dynamic spectrum anti-jamming; improved direction of arrival matrix method for two parallel uniform arrays; unscented Kalman filter for target tracking based on single observation station; handover strategy based on greedy algorithm in vehicle edge computing; safety analysis method considering cascading trips based on particle swarm optimization; method for restraining cascading trip based on node power and improved window decoding of spatially coupled low-density parity check codes; and common algorithms for corner detection to sonar image registration.

Second group: Chapters 12–19. In these chapters are presented the investigations of the authors on various aspects of the contemporary digital modulation systems, the wireless communication networks and their applications, such as: performance analysis of interleaver for polar coding in bit-interleaved coded modulation system; technology of wireless sensor network; Scholte wave for target detection; visible and infrared image fusion based on online convolutional dictionary learning with sparse matrix computation; optimization of maintenance force dispatch of artillery command information system based on mission success; application of an improved genetic algorithm in tourism route planning; classification of medical symptoms based on social sensors; and method and technology of emergency logistics intelligent system engineering.

Third group: Chapters 20–28. These chapters comprise investigations in various application areas of the wireless communications and computer technologies, such as: approaches for detecting rogue access points; stereo matching based on the fusion of sum of absolute differences and census algorithm; method of vehicle technology status based on big data analysis; bay number recognition based on deep convolutional recurrent neural network; deep reinforcement learning-based resource allocation for smart grid in radio access network; identification of e-nose wine varieties based on feature extraction and classification modeling; and region contraction-based sparse approach for magnetocardiography current source imaging and differential properties of metric projector over a class of closed convex cones.

All chapters of the proceedings were reviewed and passed the plagiarism check. The editors express their deep gratitude to IRNet for the excellent organization, coordination and support. Also, we thank the sponsors, the members of the Organizing Committee, the members of the International Program Committee of ICWCA 2019 and all the authors for their efforts for the preparation of the proceedings for publishing. We want to express our deepest gratitude to Springer excellent team for their support and expertise, which did the publishing of the book possible.

The book will be useful for researchers, university students (BSc and MSc), PhD students and lecturers, working in the area of the contemporary wireless communication networks and their applications.

Editors

Sofia, Bulgaria — Roumen Kountchev
Auckland, New Zealand — Aniket Mahanti
Haikou, China — Shen Chong
Bhubaneswar, India — Srikanta Patnaik
Krasnoyarsk, Russian Federation — Margarita Favorskaya

Contents

About the Editors

Prof. Dr. Roumen Kountchev, D.Sc., works at the Faculty of Telecommunications, Department of Radio Communications and Video Technologies – Technical University of Sofia, Bulgaria. He has published over 340 papers in journals and conference proceedings, 15 books, 46 book chapters, and 20 patents. He was a member of the Euro Mediterranean Academy of Arts and Sciences, President of the Bulgarian Association for Pattern Recognition, and a member of the editorial boards of the IJBST Journal Group, the International Journal of Reasoning-based Intelligent Systems and the International Journal of Broad Research in Artificial Intelligence and Neuroscience.

Dr. Aniket Mahanti is a Senior Lecturer (North America: Associate Professor) of Computer Science at the University of Auckland, New Zealand. He holds a Ph.D. in Computer Science from the University of Calgary. Dr. Aniket has published over 40 papers in respected peer-reviewed journals and books, and at conferences. He has more than 15 years of technical experience in information technology, computer networks, and Internet systems.

Prof. Dr. Shen Chong, D.Sc., is Deputy Director of Hainan University Scientific Research Office, China. He graduated from the School of Electronic Information of Wuhan University and the University of Strathclyde. He holds 23 patents and has published more than 100 SCI and EI indexed papers. He was the PI of various projects funded by the National Natural Science Foundation, Ministry of Education, and the International Science and Technology Cooperation. He was a member of youth standing committee of the China Electronics Association and the Chinese Association of Marine limnology, and is an IEEE VTS senior member.

Prof. Dr. Srikanta Patnaik is with the Department of Computer Science and Engineering, Faculty of Engineering and Technology, SOA University, Bhubaneswar, India. He published over 100 papers in international journals and conference proceedings, 2 textbooks, and 32 edited volumes and is the Editor-in-Chief of International Journal of Information and Communication Technology and

International Journal of Computational Vision and Robotics and Book Series: Modeling and Optimization in Science and Technology (Springer), Advances in Computer and Electrical Engineering, Advances in Medical Technologies and Clinical Practices (IGI-Global).

Prof. Dr. Margarita Favorskaya is the Head of the Department of Informatics and Computer Techniques at Reshetnev Siberian State University of Science and Technology, RF. She is an IPC member and has chaired invited sessions at over 30 international conferences. She is also a reviewer for journals such as Neurocomputing, Knowledge Engineering and Soft Data Paradigms, Pattern Recognition Letters, and Engineering Applications of Artificial Intelligence; an Associate Editor of the Intelligent Decision Technologies Journal, International Journal of Knowledge-Based Intelligent Engineering Systems, and the International Journal of Reasoning-based Intelligent Systems; and is a Springer book editor.

A Novel Low-Complexity Orthogonalization Method

Faping Lu, Lucheng Yang, Chuanhui Liu, Jiafang Kang, and Zheng Zhang

Abstract A novel low-complexity orthogonalization method was proposed for prolate spheroidal wave function (PSWF) signal, taking the parity symmetry of PSWF signal and the intimate connection between parity symmetry and orthogonality in mathematics into account, aiming at reducing the complexity of the existing orthogonalization method. According to the parity symmetry of signal, the signal processing of proposed method divides PSWFs signals into two group signals (one is odd signal and the other is even signal) and then orthogonalizes the two group signals, respectively. Both the mathematical deduction and simulation result show that the proposed method can effectively reduce the algorithm complexity at least 68.38% without compromising other system performance, comparing with conventional orthogonalization method.

1 Introduction

Prolate spheroidal wave function (PSWF) is a collection class of special functions which were defined by Bell Laboratories D. Slepian and O. Pollak in 1961 [1]. It owns complete double orthogonality, parity symmetry, spectrum controllable, good time–frequency energy aggregation, and other excellent properties in Hilbert space. Since its introduction, especially in recent years, scholars in related fields have carried out extensive and in-depth research on its solution and its nature [2–4]. Additionally,

F. Lu · L. Yang · C. Liu (✉) · J. Kang
Naval Aviation University, Yantai 264001, China
e-mail: lchgfy@163.com

Key Laboratory on Signal & Information Processing of Shandong Province, Yantai 264001, China

L. Yang
Unit 91033 of the Chinese People's Liberation Army, Qingdao, China

Z. Zhang
Unit 91980 of the Chinese People's Liberation Army, Yantai, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_1

PSWFs have been widely employed in signal processing [5, 6], satellite communication [7, 8], fiber communication [9, 10], mobile communication [11], quantum cryptography [12], image processing [13], and other fields. Among them, benefiting from the good basic properties of the PSWF signal, non-sinusoidal modulation based on PSWF signal is expected to become an effective scheme to improve the bandwidth efficiency and power efficiency of communication system. Unfortunately, the complete double orthogonality only exists in the same parameter PSWF signals, the signals are non-orthogonal when the parameters are different, and the spectrum overlaps of different PSWF signal.

In view of the above shortcomings, Zhao et al. proposed an orthogonalization method for PSWF signal based on Gram–Schmidt [14]. However, due to the large rounding error and unstable performance of Gram–Schmidt orthogonalization method, the performance of Gram–Schmidt orthogonalization method continues to decline with the increase of PSWF number. Zhang et al. introduced the Householder and the Givens transform and proposed two orthogonalization methods, solving the weakness of Gram–Schmidt [15]. The high complexity of the above-described method and the increasing computation overhead cause difficulties for the PSWF modulation engineering applications. Considering that the PSWF signal has good properties of parity symmetry in the time domain, is it reasonable to introduce the parity symmetry approach to reduce the orthogonalization algorithm complexity?

In view of the above question, this paper analyzes the relationship between parity symmetry and the orthogonality of PSWF signal, and on this basis, a low-complexity PSWF signal orthogonalization method was proposed by introducing the parity of PSWF signal and effectively reducing the algorithm complexity.

2 Introduction

In this section, firstly, we analyze the PSWF properties, its orthogonalization methods, and the problems they faced in engineering applications. Secondly, we analyze the intimate connection between the parity symmetry and orthogonality. On this basis, the low-complexity orthogonalization method is studied by introducing the idea of group processing.

2.1 The Definition of PSWF and Its Orthogonalization Methods

PSWF has many definitions and representation methods, and among them, its differential equation is expressed as [1]

$$\left(1-t^2\right)\frac{\mathrm{d}^2\psi_n(c,t)}{\mathrm{d}t^2}-2t\frac{\mathrm{d}\psi_n(c,t)}{\mathrm{d}t}+\left(\chi_n(c)-c^2t^2\right)\psi_n(c,t)=0 \tag{1}$$

where $c = TB$, B is the signal bandwidth, T is the signal time width, $\psi_n(c, t)$ is the n-th-order PSWF, $t \in [0, T]$, is the signal continuous time, $\chi_n(c)$ is the eigenvalue of the n-th order PSWF, meeting $0 < \chi_0(c) < \chi_1(c) < \chi_2(c) \cdots$.

The PSWF shown in (1) owns complete bi-orthogonality property, controllable spectrum, good parity symmetry in the time domain, and excellent properties of the best energy aggregation in time-frequency domain [1, 16]. Among the excellent properties, the complete orthogonal basis refers to the number of PSWF which satisfies the requirement to be mutually orthogonal, to be fixed to $2BT$. Bi-orthogonality property refers to the PSWFs, which possess the same width and bandwidth, and is bi-orthogonal in time domain when the order n is different.

Therefore, if we want to increase the information transmission rate, the time bandwidth product should be increased in order to obtain more PSWFs. But in the practical application, the increase of the time bandwidth product has the following problems: The duration and frequency bandwidth are not infinite; PSWFs do not have closed analytic solution, and using computer numerical solution or function fitting has the shortcomings of high complexity and poor computational accuracy for large time bandwidth product.

Aiming at solving the above problems, Zhao et al. proposed a method using molecular band selection of PSWF, divided wide frequency band range into multiple smaller sub-zone, and then orthogonalized the different sub-band PSWFs, solving the problem of large time bandwidth product problems in the practical application effectively [14]. At present, the PSWF orthogonalization methods are mainly three kinds, based on Gram–Schmidt, Householder, and Givens transform. Although they can effectively guarantee good orthogonality between different PSWFs after orthogonalization, the above orthogonalization method has higher complexity. It limits the practical applications of PSWFs, especially the power-constrained systems, such as the satellite communication systems.

On the basis of existing orthogonalization methods, how to improve the orthogonalization method according to the nature of PSWFs, reducing the complexity of the algorithm and promoting the practical engineering application of PSWF? To solve this problem, the low-complexity PSWF orthogonalization method is studied in the next sections.

2.2 Improved Orthogonalization Method of PSWF

According to the properties of the function multiplication, the multiplication of different parity symmetry functions has the following relationship

$$\begin{aligned} &\text{Odd function} \times \text{Odd function} = \text{Even function} \\ &\text{Odd function} \times \text{Even function} = \text{Odd function} \\ &\text{Even function} \times \text{Even function} = \text{Even function} \end{aligned} \tag{2}$$

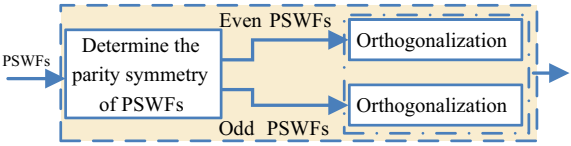


Fig. 1 Schematic diagram of PSWF generation

According to (2), the new functions which are the multiplication of functions with parity symmetry are odd or even symmetric about the central time in a period. This shows that there is an intimate connection between the parity symmetry and orthogonality of functions and can be interchanged under certain conditions. If Z_{odd} is a set made up of odd functions, Z_{even} is a set which is made up of even functions, $\forall f_{\text{odd}} \in \text{Z}_{\text{odd}}$,$\forall f_{\text{even}} \in Z_{\text{even}}$, so f_{odd} and f_{even} are orthogonal to each other.

Considering that the PSWF has good parity in the time domain, and the number of odd symmetric and even symmetric functions are equal, we suppose the number of PSWFs $\psi\{\varphi_i(t), i = 1, 2, \ldots, N\}$ required to be orthogonal is N, and ψ_{O} and ψ_{E} are a set made up of odd symmetric and even symmetric PSWFs, respectively. According to (2), the elements in set ψ_{O} and ψ_{E} are orthogonal to each other, and in other words, they have natural orthogonality.

At the same time, according to the basic principle of Gram–Schmidt, Householder transform, and Givens transform, we know that the above orthogonalization methods need linear computation. Moreover, the linear operations of even functions (or odd functions) do not change the parity of the function. Therefore, the even (or odd) PSWFs after orthogonalization are also even (or odd) functions. As for the even (or odd) PSWFs which are in ψ_{E}(or ψ_{O}) after orthogonalization, the orthogonal PSWFs made up the set ψ_{E1}(or ψ_{O1}) which are even (or odd) function; it means that $\psi_1 = \psi_{\text{O1}} \oplus \psi_{\text{E1}}$ is the set which is made up of the elements in ψ after orthogonalization.

The block diagram of the proposed method is shown in Fig. 1; the main steps of the improved orthogonalization method are as follows:

Step 1 Divide PSWFs into two groups according to its parity symmetry, the even PSWFs made up the set ψ_{E}, the odd PSWFs made up the set ψ_{O}.

Step 2 Orthogonalize the two group signals, respectively: The results are ψ_{E1} and ψ_{O1}.

Step 3 Combine ψ_{E1} with ψ_{O1}, set $\psi_1 = \psi_{\text{O1}} \oplus \psi_{\text{E1}}$, completing the orthogonalization of all PSWFs

2.3 Complexity Analysis of Improved Orthogonalization Method

The algorithm complexity mainly includes time complexity and space complexity, in which the time complexity of the algorithm refers to the calculation effort required by the algorithm, and it can directly reflect the complexity of the algorithm.

Table 1 Computation amount of orthogonalization methods

Orthogonalization methods		Computation amount	
		Multiplication	Addition
Gram–Schmidt	Original	$MN^2 - MN + 2M$	$MN^2/2 - MN/2 + M - N$
	Improved	$MN^2/2 - MN + 4M$	$MN^2/4 - N^3/6 + MN$
Householder	Original	$MN^2 - N^3/3$	$MN^2/2 - 2N^3/3 + MN$
	Improved	$MN^2/2 - N^3/12$	$MN^2/4 - N^3/6 + MN$
Givens	Original	$2MN^2 - 2N^3/3$	$MN^2 - N^3/3 + MN/2$
	Improved	$MN^2 - N^3/6$	$MN^2/2 - N^3/6 + MN/2$

We use multiplication and addition operations as numerical measures. Suppose the number of PSWFs is N, the sampling points is M, and $M > N$, the computation amounts of the different orthogonalization methods are shown in Table 1 [15, 17]. From the table, we see that the proposed method in this paper can effectively reduce the algorithm complexity, comparing with original orthogonalization methods. At the same time, the effect of the reduced complexity is more obvious with the increasing of the sampling points M and PSWF number, N.

3 Simulation Results

To show the performance of the proposed orthogonalization method more intuitively, and to verify the feasibility of the proposed method and the correctness of the theoretical analysis, we analyzed the performance in orthogonality performance, algorithm complexity, and power spectrum characteristic, comparing with the original orthogonalization methods.

3.1 Simulation Condition

PSWFs signal: Frequency band 3–8 GHz is divided into four sub-wave bands; the frequency spectrum overlapping degree is 50%; time bandwidth product $c = 4$ Hz s; sampling points $M = 300$. Each wave band takes the first two-order PSWF according to the energy from high to low.

Orthogonalization method: Householder transform. Because Givens transform algorithm has high complexity, the orthogonal performance of Gram–Schmidt goes down when PSWF number increases. Therefore, this part focuses on analyzing the performance of the Householder transform, showing the performance of the orthogonalization method.

Table 2 Dynamic range of cross-correlation of orthogonal PSWF

Sampling points M	Method	Minimum	Maximum
8	Householder	5.05E−17	1.57E−15
	Improved householder	−5.05E−17	−3.35E−12
16	Householder	5.73E−18	−7.80E−15
	Improved householder	1.86E−18	−4.02E−06
24	Householder	8.46E−18	−5.08E−14
	Improved householder	1.49E−17	5.58E−06

3.2 Analysis of Simulation Results

Orthogonal PSWF signal: The dynamic range of cross-correlation of PSWFs of Householder transform and the improved Householder transform is shown in Table 2 [15]. With the increase of PSWF number, the change of minimum cross-correlation value of Householder transform and the improved Householder transform is very small, almost unchanged; while the maximum cross-correlation values of improved Householder transform keep increasing.

Algorithm complexity: It shows that the performance of the improved Householder transform in maximum cross-correlation value declines. According to the algorithm results given in Tables 1 and 3 is obtained. From Table 3, we can see that the order of magnitude of the cross-correlation value of the improved Householder transform is $\leq 10^{-5}$, which means that the improved orthogonalization method still has good orthogonal properties.

Table 3 Computation amount of orthogonalization methods

Sampling points M	Orthogonalization methods	Computation amount		Reduced ratio (%)
		Original	Improved	
100	Schmidt	2.92E+05	8.35E+04	71.38
	Householder	2.55E+05	8.06E+04	68.38
	Givens	6.28E+05	1.37E+05	78.22
300	Schmidt	1.42E+06	2.71E+05	80.92
	Householder	1.02E+06	2.73E+05	73.09
	Givens	2.13E+06	4.52E+05	78.82
500	Schmidt	1.79E+06	4.59E+05	74.38
	Householder	1.78E+06	4.66E+05	73.77
	Givens	3.64E+06	7.67E+05	78.92

In order to illustrate the algorithm complexity of the different orthogonalization methods which vary with the number of PSWF M, and the number of sampling points N, the complexity of the algorithm is analyzed for $M = 100, 300, 500$, $N =$ 10–50. Figure 2 shows the complexity of the different orthogonalization methods, and Table 3 shows the complexity when $N = 50$.

The simulation results show that the improved and the original orthogonalization method complexity gradually increases with the increase of the number of PSWF M and the sampling points N. Meanwhile, comparing with the original orthogonalization method, the improved orthogonalization method can effectively reduce the algorithm complexity, and with the increase of the number of PSWF M and the sampling points N, the advantage of the improved orthogonalization method in reducing the algorithm complexity is gradually outstanding, e.g., the algorithm complexity reduced approximately at 68.38% when $M = 100$; reduced approximately 73.09% when $M = 300$, and reduced approximately 73.78% when $M = 500$.

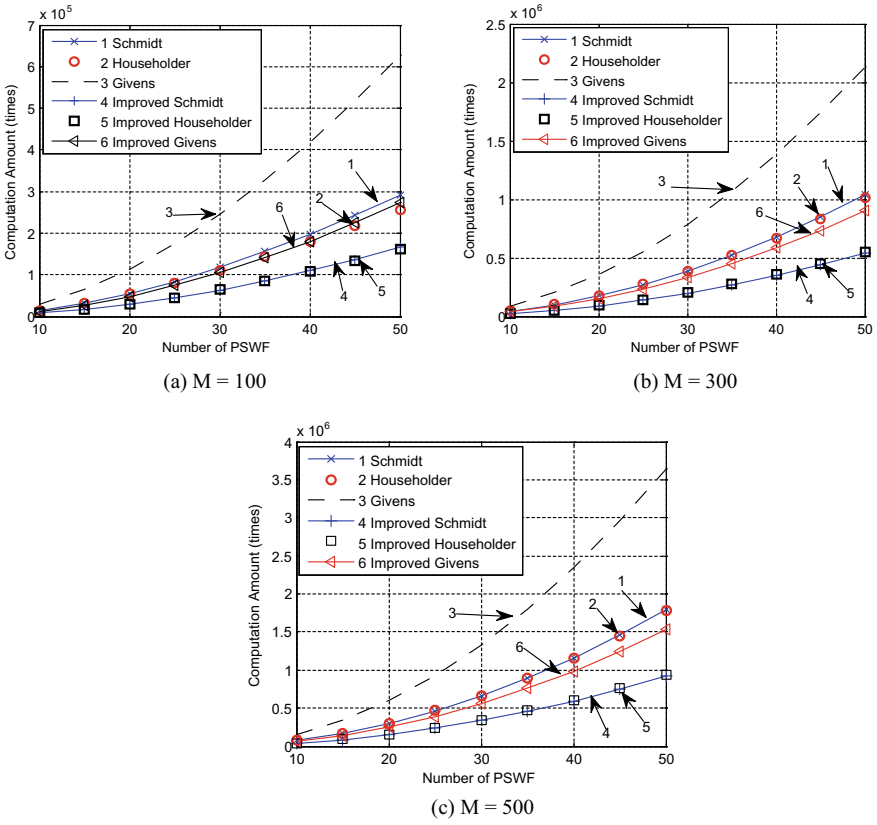
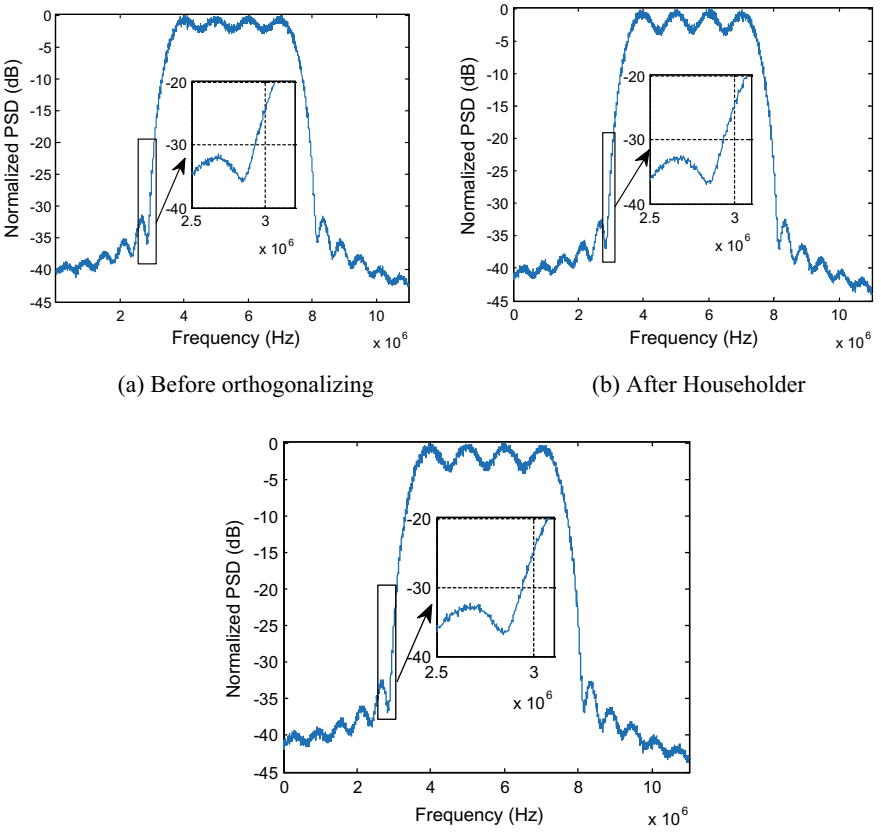


Fig. 2 Computation amount of different orthogonalization methods

According to the above analysis, the improved orthogonalization method can effectively reduce the complexity of the algorithm, which is consistent with the theoretical analysis.

Power spectrum density (PSD) characteristic: The PSD of PSWF modulation signal is shown in Fig. 3; the simulation results show that the PSWF modulation signal with high energy clustering in the frequency domain, and the out-of-band (OOB) quickly decay, illustrating that PSWFs have good frequency energy clustering. The Householder transform has almost no influence on the spectrum of PSWFs; the PSWFs still has high energy clustering after orthogonalization. Comparing with the original Householder transform orthogonalization method, the improved orthogonalization method, and its spectrum of PSWFs do not change significantly.

According to the above analysis, the improved orthogonalization method does not change the PSD property of PSWFs.



(a) Before orthogonalizing

(b) After Householder

(c) After improved method

Fig. 3 Computation amount of different orthogonalization methods

4 Conclusion

As for the existing PSWFs orthogonalization method for the problem of high complexity, an improved orthogonalization method was proposed based on the intimate connection between parity symmetry and orthogonality in mathematics. Both the mathematical deduction and simulation result show that the proposed method can effectively reduce the algorithm complexity at least 68.38% by appropriate sacrificing PSWFs maximum cross-correlation value, without changing the power spectrum of PSWFs and the minimum cross-correlation value, comparing with conventional orthogonalization method. Additionally, the reason why the improved orthogonalization method sacrifice the maximum cross-correlation value, is the parity of band pass PSWFs is relative. In other words, it is not strictly symmetric, the difference between symmetric points is not 0. Therefore, the performance of orthogonality will decline when the number is large. How to construct strictly odd or even symmetric PSWFs without destroying the other PSWFs properties, in order to reduce the maximum correlation value and enhance the performance of the improved orthogonalization method, will be the next study focus.

References

1. Slepian, D., Pollak, H.O.: Prolate spheroidal wave functions, Fourier analysis, and uncertainty-I. Bell Syst. Tech. J. **20**(1), 43–46 (1961)
2. Wang, H.X., Lu, F.P., Liu, C.H., Liu, X., Kang, J.F.: The study of time-frequency characteristics of cross-terms between prolate spheroidal wave function signal. J. Electron. Inf. Technol. **39**(6), 1319–1325 (2017)
3. Hogan, J.A., Lakey, J.D.: On the numerical evaluation of bandpass prolates II. J. Fourier Anal. Appl. **23**(1), 1–16 (2016)
4. Dunster, T.M.: Asymptotics of prolate spheroidal wave functions. J. Class. Anal. **11**(1), 1–21 (2017)
5. Zhu, Z.H., Wakin, M.B.: Approximating sampled sinusoids and multiband signals using multiband modulated dpss dictionaries. J. Fourier Anal. Appl. **23**(6), 1–48 (2015)
6. Mahata, K., Hyder, M.M.: Frequency estimation from arbitrary time samples. IEEE Trans. Signal Process. **64**(21), 5634–5643 (2016)
7. Sanctis, M.D., Cianca, E., Rossi, T., Sacchi, C., Mucchi, L., Prasad, R.: Waveform design solutions for EHF broadband satellite communications. IEEE Commun. Mag. **53**(3), 18–23 (2015)
8. Chen, Z.N., Wang, H.X., Zhao, Z.Y., Liu, X.: Novel PSWF-based multidimensional constellation modulation for broadband satellite communication. Wirel. Pers. Commun. **88**(3), 493–524 (2016)
9. Djordjevic, I.B., Jovanovic, A.Z., Cvijetic, M., Zoran, P.: Multidimensional vector quantization-based signal constellation design enabling beyond 1 pb/s serial optical transport networks. IEEE Photon. J. **5**(4), 7901312 (2013)
10. Djordjevic, I.B., Jovanovic, A.Z., Peric, Z.H., Wang, T.: Multidimensional optical transport based on optimized vector-quantization-inspired signal constellation design. IEEE Trans. Commun. **62**(9), 3262–3273 (2014)
11. Shen, J., Wang, O., Yu, J.G.: Non-orthogonal Pulse Shape Modulation for Wireless Relay D2D Communication. In: 2017 International Conference on Computing Intelligence and Information System. IEEE, Nanjing, China (2017)

12. Molotkov, S.N., Potapova, T.A.: Wavefunctions of a prolate spheroid and multiplexing in relativistic quantum cryptography on orthogonal states. JETP Lett. **100**(9), 596–603 (2015)
13. Landa, B., Shkolnisky, Y.: Steerable principal components for space-frequency localized images. SIAM J. Imaging Sci. **10**(2), 508–534 (2017)
14. Zhao, Z.Y., Wang, H.X., Li, H.L., Mao, Z.Y.: Designing method of orthogonal pulse in time domain based on prolate spheroidal wave functions for nonsinusoidal wave communication. J. Electron. Inf. Technol. **31**(12), 2912–2916 (2009)
15. Zhang, Z., Wang, H.X., Zhao, Z.Y., Zhang, L., Liu, X.G., Zhang, C.L.: Design of orthogonal pulse for prolate spheroidal wave function based on householder transformation. Telecommun. Eng. **53**(11), 1452–1458 (2013)
16. Osipov, A., Rokhlin, V., Xiao, H.: Prolate Spheroidal Wave Functions Of Order Zero. Springer, Boston, US (2013)
17. Zhaonan, Chen, Hongxing, Wang, Peilin, Zhong, Xuejun, Zhang, Chuanhui, Liu: PAPR Suppression of PSWF pulse modulated signal based on givens rotation transform. J. Electron. Inf. Technol. **000**(006), 1406–1412 (2013)

Research on Radio Base Station Distribution Technology in Transrapid System Communication

Tingjun Li, Haining Yang, Kai Fang, Haiming Wang, Maiqi Duan, Na Li, and Yong Fan

Abstract The distribution and switching strategy of radio base station (RBS) are an important part in the communication system design of transrapid system. In this paper, the distribution and switching strategy of RBS are studied for high-speed maglev communication from the points of power coverage and Doppler frequency shift along the track. The power coverage and Doppler frequency shift along the maglev track are evaluated, the influence of the distance between RBS antenna and track center on power coverage and Doppler frequency shift is discussed, and the fluctuation of power coverage and Doppler frequency shift in the same data frame are analyzed. Finally, some advices about the distribution and switching strategy of RBS in transrapid communication system are given.

1 Introduction

The high-speed maglev rail transit system is a system with extremely high safety requirements [1–4], which includes vehicles, track lines, maglev, guide, traction, operation control, train positioning and communication systems. Among them, the vehicle-ground radio communication system establishes a reliable and effective radio communication connection between train and ground fixed equipment which realizes operation control data, location data, diagnostic data, passenger information system data and necessary voice communication. Compared with the current wheel–rail high-speed railway system, the high-speed maglev requires higher real-time and reliability for the vehicle-ground communication system. The traditional vehicle-ground communication system cannot meet the operational requirements of the high-speed maglev system [5–8]. As an important part of the design of high-speed maglev

T. Li · H. Yang (✉) · M. Duan · N. Li · Y. Fan
School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China
e-mail: hnyang@uestc.edu.cn

K. Fang · H. Wang
Hunan CRRC Times Signal & Communication Co. Ltd., Changsha, Hunan 410005, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_2

vehicle-ground communication system, it is necessary to research the distribution and switching strategy of RBS.

2 Receive Power Along the Track

When designing the vehicle-ground communication base station distribution and switching strategy of high-speed maglev vehicle, the millimeter wave (mmW) coverage power of RBS along the track must be higher than the minimum received power required by the vehicle-ground communication system. The following is an analysis of the received power along the track.

2.1 Communication Equation

When the high-speed maglev train is moving, the antenna of the RBS transmits mmW signal, which is received by the antenna of the train mobile base station. The power can be calculated by the communication equation:

$$P_{\mathrm{r}} = \frac{P_{\mathrm{t}} G_{\mathrm{r}} G_{\mathrm{t}} \lambda^2}{(4\pi)^2 \left(d^2 + d_n^2\right)} \tag{1}$$

where P_{r} is the power received by the mobile antenna (MA) at a certain point in the track, P_{t} is the transmitting power of the RBS antenna, G_{r} and G_{t} are, respectively, the gain of receiving antenna and transmitting antenna, λ is the working wavelength of the mmW, d is the distance from the RBS antenna to a certain point in the track, d_n is the distance from the RBS antenna to the center of the train track. The positional relationship between the train and the RBS is shown in Fig. 1.

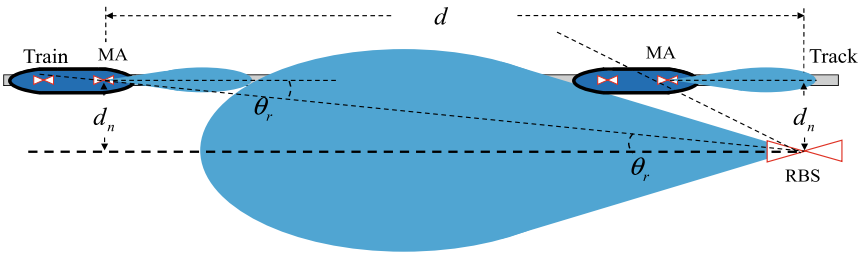


Fig. 1 Positional relationship between train and RBS

2.2 d_n Impact on Receiving Power

Given that the transmitting power of the RBS antenna is 10 dBm, the mmW operating frequency is 38 GHz, the RBS antenna is 2.5 m away from the train track center, the antenna gain of the RBS antenna is shown in Fig. 2, and then, the power received by the train during its movement from 0 to 1500 m away from the RBS is shown in Fig. 3.

It can be seen from Fig. 3. that when the train leaves the RBS to 15.8 m, the power received by the mobile antenna will rapidly increase to −28.7 dBm but will then decrease as the train moves away from the RBS. In the whole process, the lowest power received by the antenna of RBS is −62 dBm. The power of the RBS received by the track in the range from 0 to 15.8 m away from the RBS fluctuates greatly, which is not suitable for the base station switching. The distance between the antenna of the RBS and the center of the train track varies with the factors such as d_n. The following will analyze the impact of different d_n values on this area, and the simulation results are shown in Fig. 4.

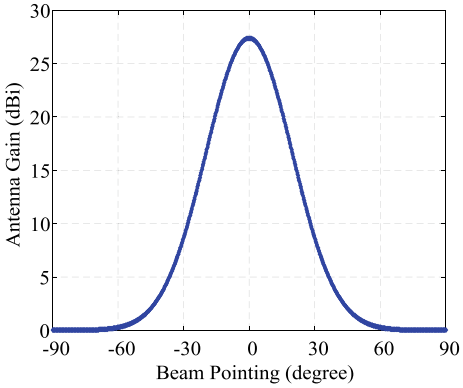


Fig. 2 Antenna gain

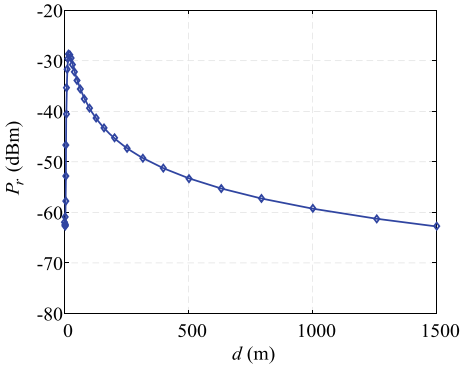


Fig. 3 Received power of the mobile antenna at $d_n =$ 2.5 m

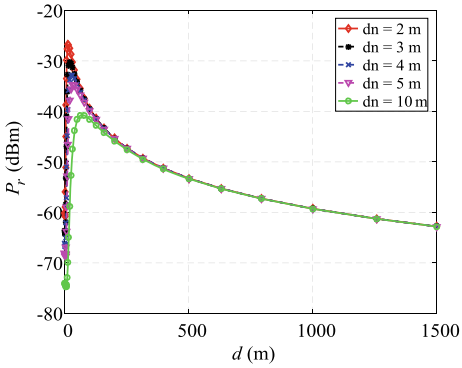


Fig. 4 Effect of different d_n values on train receiving power

It can be seen from Fig. 4 that as the d_n increases, the unsuitable switching area of the base station increases. When the value of d_n is 10 m, the area increases to 63 m. As the train continues to leave the RBS, the effect of d_n on the received power of the train is reduced, and the effect on the received power at the long-distance track is negligible.

2.3 Dual Base Station Communication Power Coverage

In order to ensure the reliability of communication, the high-speed maglev vehicle communication system adopts AB double RBS redundancy setting, and the A-RBS and B-RBS are interleaved and distributed along the track. Under normal circumstances, at least one base station of the AB double RBS communicates with the train. Take the 2 km track as an example to set the AB double RBS, and the received power along the track is as shown in Fig. 5, where red-star line is the coverage area of the

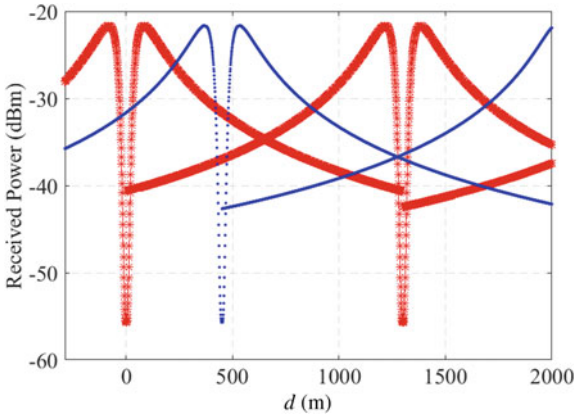


Fig. 5 2 km AB dual RBS communication power coverage

A-RBS and blue-dot line is the coverage area of the B-RBS. When switching the A-RBS at 0 m to the A-RBS at 1305 m, in addition to considering the unsuitable switching area of the A-RBS, it is also necessary to avoid the unsuitable switching area of the B-RBS at 448 m, in case that the B-RBS cannot communicate with the train after the switch fails.

2.4 Receive Power Fluctuation Analysis

The communication between high-speed maglev train and RBS is carried out by time division multiple access (TDMA). In order to ensure the reliability of base station switching, the system requires that the train receiving power in each TDMA frame fluctuates gently during the base station switching. Assuming that the transmission time required by each frame of TDMA is 1 ms, the following is an example of 600 km/h speed as an example to analyze the train receiving power fluctuations in the TDMA frame. The simulation results are shown in Fig. 6. It can be seen from Fig. 6 that the maximum fluctuation of the received power of the train at 600 km/h in TDMA does not exceed 1.2 dBm per frame, and after the train leaves the RBS 20 m, the maximum fluctuation of the received power is less than 0.1 dBm and gradually approaches 0. Based on the fluctuation of the received power, the switching area of the base station should be at least 20 m away from the base station.

3 Doppler Frequency Shift Analysis

In addition to receiving power along the track, the design of the mobile communication strategy for high-speed maglev vehicles must also consider the effects of Doppler frequency shift.

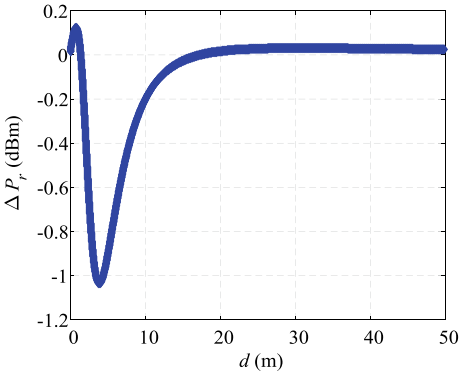


Fig. 6 TDMA frame train receiving power fluctuations

3.1 Doppler Frequency Shift at Different Speeds

When the train moves along the track at a certain speed, there is a difference between the transmitting frequency and receiving frequency, which is called the Doppler frequency shift. The formula for calculating the Doppler frequency shift is as follows:

$$f_{\mathrm{d}} = \frac{v}{c} f_c \cos\theta \tag{2}$$

where f_{d} is the Doppler frequency shift, v is the relative motion speed of the train and the RBS, c is the propagation speed of light, θ is the angle between the train motion direction and the RBS antenna, and f_c is the emission frequency of the RBS.

The Doppler frequency shift causes an increase in the bit error rate of the vehicle-ground communication. When designing the distribution strategy of RBS, the switching site of RBS should be placed in the area with moderate Doppler frequency shift fluctuation as far as possible to obtain the best communication quality.

The effect of Doppler frequency shift is analyzed by simulation. The antenna of the RBS is 2.5 m away from the center of the train track, the mmW operating frequency is 38 GHz, and the Doppler frequency shift generated by the train running from the left position 100 m away from the RBS (represented by negative value) to the right position 100 m at different speeds is shown in Fig. 7.

It can be seen from Fig. 7 that the Doppler frequency shift of the train at a speed of 600 km/h is about ±21 kHz when it is far away from the RBS (>50 m). When the train is close to the RBS (<20 m), the Doppler frequency fluctuation is large and should avoid switching base stations in this area.

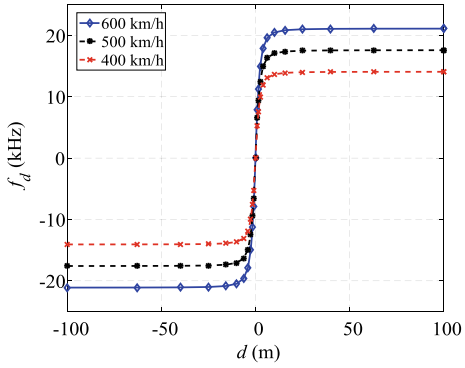


Fig. 7 Doppler frequency shift results at different vehicle speeds

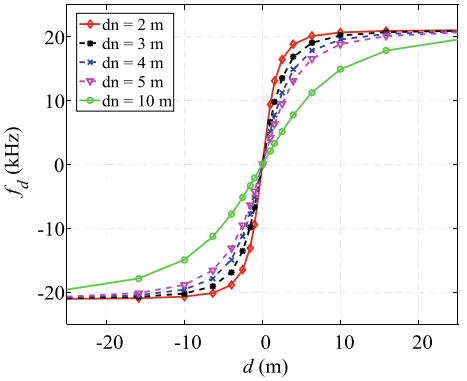


Fig. 8 Effect of different d_n values on Doppler frequency shift

3.2 The Impact of d_n on Doppler Frequency Shift

The RBS antenna distance from the center of the track also affects the Doppler frequency shift. Taking the train at a speed of 600 km/h as an example, the Doppler frequency shift generated by the antenna of the RBS at different position from the track center is shown in Fig. 8.

As is shown from Fig. 8, when the train is far away from the antenna of the RBS (>25 m), the Doppler frequency shift curve when d_n takes different values basically coincides. When the train position is relatively close to the antenna of the RBS (<10 m), there is a difference in the speed of Doppler frequency shift change when d_n takes different values, and the smaller the d_n is, the faster the Doppler frequency shift changes. When selecting the switching area of base station, the appropriate switching area should be selected according to the specific situation of the antenna distance from the track center of RBS.

3.3 Doppler Frequency Shift Fluctuation Analysis

Similar to the analysis of received power along the track, Doppler frequency shift fluctuation also exists in the same TDMA frame. When selecting the switching area of RBS, it should be avoided to select the area with large Doppler frequency fluctuation in the same TDMA frame. Assuming that the time required for TDMA transmission per frame is 1 ms, take 600 km/h speed as an example to analyze the Doppler frequency shift fluctuation within TDMA frames, and the simulation results are shown in Fig. 9.

As is shown from Fig. 9, the maximum fluctuation of Doppler frequency shift under the condition of 600 km/h speed in the same TDMA frame is about 1.4 kHz, and the maximum difference of received power is less than 0.1 kHz and gradually approaches 0 after the train leaves the RBS for 15 m.

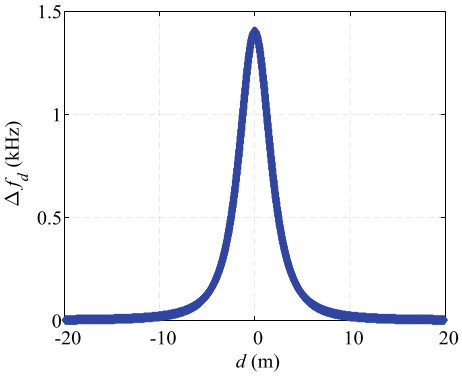


Fig. 9 TDMA intra-frame Doppler frequency shift fluctuations

4 Conclusion

There are many factors involved in the ground station distribution and switching of the ground communication system of the maglev vehicle, including the maximum train speed and the maximum operating distance of the RBS. From the perspective of receiving power and Doppler frequency shift, the switching area of RBS should meet the basic requirements of minimum mmW power coverage, receiving power and smooth Doppler frequency shift for double RBS independent communication, and the design should be made according to the track situation and economy.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB1200601, in part by the National Natural Science Foundation of China under Grants 61701088 and 61701093 and in part by the Sichuan Science and Technology Project of China under Grant 2019JDRC0061.

References

1. Wu, J., Fan, P.: A survey on high mobility wireless communications: challenges, opportunities and solutions. IEEE Access **4**, 450–476 (2016)
2. Pan, M., Lin, T., Chen, W.: An enhanced handover scheme for mobile relays in LTE—A high-speed rail networks. IEEE Trans. Veh. Technol. **64**(2), 743–756 (2015)
3. Fan, D., Zhong, Z., Wang, G.: Doppler shift estimation for high-speed railway wireless communication systems with large-scale linear antennas. In: 2015 International Workshop on High Mobility Wireless Communications (HMWC), pp. 96–100. IEEE, Xi'an (2015)
4. Wu, X.: Maglev Train. Shanghai Science and Technology Press, Shanghai (2003)
5. Wu, W.: Urban Rail Transit Signal and Communication System. China Railway Publishing House, Beijing (1998)
6. Chen, G.: Introduction to railway signals. China Railway Publishing House, Beijing (1995)
7. Wang, W.: Railway-specific communication. China Railway Publishing House, Beijing (1995)
8. Wu, Z., Zhou, J., Lin, B.: Research on wireless communication technology of subway. Mod. Urban Rail Transit. **3**, 19–23 (2010)

A Novel Node Scheduling Algorithm for Solar-Powered Wireless Sensor Networks

Wanguo Jiao, Xin Zhang, and Hao Wang

Abstract By utilizing solar energy, the wireless sensor network has the potential to survive longer. However, the solar energy is varying with the time and is susceptible to the surrounding. How to effectively manage the dynamic harvested energy is important for the solar-powered wireless sensor network. In this paper, we design an energy scheduling for the wireless sensor network which consists of fewer sensors powered by the solar energy. First, to effectively cover targets and reduce coverage cost, we design a new node deployment scheme which can reduce required sensors while guarantees target coverage. Then, based on this deployment, we propose a node scheduling algorithm to improve network lifetime. Through choosing relays and dynamically adjusting the role of the sensor with different energy levels, the harvested energy can be utilized effectively, and the network lifetime is prolonged. The simulation results verify that our proposed scheme can prolong the network lifetime while requires fewer sensors to achieve the coverage requirement.

1 Introduction

With the rapid development of the Internet of Things (IoTs), as the important part of the IoTs, the wireless sensor network (WSN) has become a hot research point in the field of the IoTs [1]. Due to the cost and the limitation of the device size, the sensor often uses the battery which has limited capacity. The battery is often difficult to be recharged or replaced. Hence, the lifetime is an important problem in the WSN and has aroused the widespread concern [2]. There are many works done to prolong the lifetime of the WSN [2–5]. Through energy management or route selection, the lifetime problem of the WSN has been effectively prolonged. However, due to the constraint of limited energy, the traditional WSN cannot be self-sustaining. By exploiting energy harvesting (EH) technology, the sensor can use the energy from ambient energy sources, such as solar, wind, vibration, and so on.

W. Jiao (✉) · X. Zhang · H. Wang
Nanjing Forestry University, Longpan Road 159, Nanjing 210037, China
e-mail: wgjiao@njfu.edu.cn

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_3

This makes possible the self-sustainability of the WSN. Recently, researchers have made some efforts on the lifetime of the WSN employing the energy harvesting technology which is referred to as the energy harvesting wireless sensor network (EHWSN) [6–10].

These available works have well studied the lifetime problem of EHWSNs. However, the coverage requirement of the network cannot be guaranteed [2, 6]. That is, the network coverage guarantees, and the lifetime optimization of the EHWSN with fixed-position sensors cannot be achieved simultaneously. Besides, to meet the requirement of lifetime, authors assume that the node density is large enough [10, 11]. The large node density means that huge number of sensors is required to cover a certain area, which results in high cost. The comprehensive network environment may not allow so many sensors to be deployed. In the solar-powered WSN, weather is an important factor affecting the energy harvesting of sensors. However, the influence of the extreme weather has not been considered in the current works. To address the above problems, we propose an algorithm to improve the lifetime and guarantee the coverage requirement in the solar-powered WSN with fixed sensor position.

In this paper, we consider the coverage guarantee and prolonging lifetime of an EHWSN which uses the solar energy source. First, we design a new scheme to deploy sensors to guarantee coverage requirement. This deployment scheme can reduce the number of sensors while guarantees the target coverage, which can be used in wide application scenarios. Second, considering more factors influencing the solar energy harvesting, we propose a node scheduling algorithm to utilize the harvested energy effectively and prolong the lifetime of the network. The proposed algorithm can effectively utilize the harvested energy to prolong the lifetime and even make the network work sustainable under certain conditions. Furthermore, the proposed algorithm brings multiple options of routing paths, and the robustness and expansibility of the network are also improved.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the network model including energy harvesting model and energy consuming model, and the proposed deployment scheme. Section 3 elaborates the proposed node scheduling algorithm. The experimental results are given in Sect. 4. At last, we conclude the whole paper in Sect. 5.

2 Network Model

In this paper, we use $\boldsymbol{G} = (\boldsymbol{S}, \boldsymbol{T}, \boldsymbol{L})$ to denote the whole network, where $\boldsymbol{S}$ is the set of all nodes, $\boldsymbol{T}$ is the set of all targets, and $\boldsymbol{L}$ represents the links of the network. Each node senses a circular area with a fixed radius which is denoted by r. According to [12], the transmission radius is twice the sensing radius in our work, while the communication range is at least twice the regular sensing range [9]. Under this condition, our algorithm makes the EHWSN survive longer and has better robustness.

We assume that nodes can harvest energy when they are sensing [7, 8, 13]. In this paper, the energy source of the sensor is the solar energy, and the solar energy collection model and the energy consumption model are introduced as follows.

2.1 Energy Model

We use the energy harvested by a solar panel per hour to represent the total energy that a node can collect in a certain period of time. According to solar radiation intensity and illumination time, the total amount of solar radiation in one sunny day can be calculated. In practical, the weather keeps changing, and it could be sunny, cloudy, or rainy, which affects the available energy. The influence of the weather can be expressed by the solar radiation intensity. Through studying the position and the movement of the sun, the variation of the solar radiation intensity can be modeled.

The sun is assumed to rotate $15^\circ\ \mathrm{h}^{-1}$, and ω denotes the solar hour angle. The value of ω can be used to express the time: $\omega < 0$ means sunrise, $\omega > 0$ denotes sunset, and $\omega = 0$ is the noon. According to the movement of the sun in a day, the relation between the solar altitude angle and other position parameters can be expressed as

$$\sin h = \sin\varphi \sin\delta + \cos\varphi \cos\delta \cos\omega \tag{1}$$

where h is solar altitude, φ is geographic latitude, and δ is obliquity of the ecliptic. When $h = 0$, we obtain the cosine expression of the sunrise and sunset point as

$$\cos\omega = -\tan\varphi \tan\delta \tag{2}$$

By using expressions of the sunrise and sunset point in (2), we calculate the time of sunrise and sunset, respectively. The times of the sunrise and sunset are

$$t_\mathrm{r} = 12 - \frac{|\omega|}{15^\circ} \text{ and } t_\mathrm{d} = 12 + \frac{|\omega|}{15^\circ} \tag{3}$$

The intensity of the radiation at night is 0. We assume that solar radiation intensity which is obtained by a solar panel on sunny days is $\alpha(t)$, and the intensity in cloudy days is a constant value K. Therefore, the total energy that a sensor can collect in f days is E^f, which can be expressed as

$$E^f = \eta * f * \lambda * \int_{t_\mathrm{r}}^{t_\mathrm{d}} \alpha(t)dt + K * (2\omega/15) * \eta * f * (1-\lambda) \tag{4}$$

where λ is the probability that a day is a sunny day, f is the number of days in which the network has worked, and η represents the efficiency of converting solar energy into electricity. Here, to simplify the design, we assume that if the day is not a sunny day, it is a cloudy day. As the intensity of the radiation is 0 in the rainy day, our model can be extended easily.

According to [9], the energy consumed by the network consists of three following parts: the energy consumed for sending data, receiving data, and collecting data. The energy consumed by sending X bits is $\mathrm{Tr}(X, d_0) = E_\mathrm{pro}X + E_\mathrm{amp}Xd_0$, where E_pro

denotes the energy consumed per bit in the information processing and E_{amp} denotes the energy used to signal amplification, and d_0 represents the distance between nodes. The energy consumed by receiving X bits is $\text{Re}(X) = E_{\text{pro}}X$. The energy consumed by sensing is $\text{Se}(Z) = E_{\text{pro}}Z + E_{\text{sen}}$, where Z represents the number of bits generated by a sensor during a day and E_{sen} is the fixed amount of energy consumed by a working sensor a day. In this paper, we assume that multiple targets in the sensing area of the same sensor do not affect the value of E_{sen}.

By using the harvested energy in (4) and the consumed energy, we can derive the remaining energy of a battery after working f days, which is expressed as

$$E_{\text{r}}^{f} = C_0 - \text{Tr}(X, d_0)f - \text{Re}(X)f - \text{Se}(Z)f + \eta f \lambda \int_{t_{\text{r}}}^{t_{\text{d}}} \alpha(t)\text{d}t + K\frac{2\omega}{15^\circ/h}\eta f(1-\lambda), \tag{5}$$

where C_0 is the capacity of the battery. Without loss of generality, we assume that the battery of each sensor has the same capacity and is full at the beginning [14].

2.2 The Deployment Scheme

To reduce required sensors while guarantee coverage requirement, we propose a node deployment scheme. In the scheme, sensors are deployed at fixed points regularly. To explain the deployment scheme, an example with 36 sensors is given in Fig. 1. As shown in Fig. 1, sensor nodes (blue dots) are fixed-position ones, and they are arranged in the following manner: In coordinates, it starts at the positive half axis of the X-axis, and in the Y-axis, the sensors are deployed on the line $Y = X$ and lines which are parallel to $Y = X$. According to the transmission radius, the distance between adjacent parallel lines is set as $\sqrt{2}r$. The distance between two adjacent sensor nodes on the same line also is $\sqrt{2}r$. By using this deploying scheme, 36 sensors can cover the whole network area. There are N targets randomly distributed, which are denoted by red dots in Fig. 1. In addition, S_i and T_j denote the sensor node and the target in the network, respectively, where $i = 1 \ldots |\boldsymbol{S}|$ and $j = 1 \ldots |\boldsymbol{N}|$. Furthermore, $\boldsymbol{T}_{\text{overlap}}$ is the set of targets which are in the overlapping part of the sensing area, while $\boldsymbol{T}_{\text{special}}$ is the set of targets in non-overlapping areas. Hence, $\boldsymbol{T} = \boldsymbol{T}_{\text{overlap}} \cup \boldsymbol{T}_{\text{special}}$. We further define some notations for the sensor nodes to differentiate the sensor node in different coverage areas. The set $\boldsymbol{S1}$ consists of sensor nodes which cover elements of $\boldsymbol{T}_{\text{special}}$, while sensor nodes which only cover $\boldsymbol{T}_{\text{overlap}}$ make up the set $\boldsymbol{S2}$. The rest sensor nodes which do not work or work as the relay node make up the set of $\boldsymbol{S3}$. Thus, $\boldsymbol{S} = \boldsymbol{S1} \cup \boldsymbol{S2} \cup \boldsymbol{S3}$. The sink is deployed outside the sensing area, and it has continuous power supply [6]. The deployment information is stored in the sink, that is, the sink has the location information of every sensor.

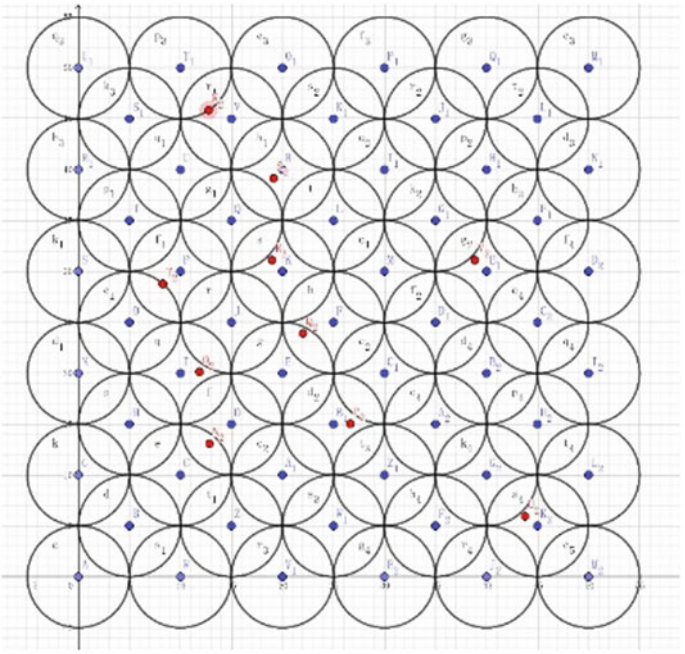


Fig. 1 An example of the proposed node deploy scheme

3 The Proposed Algorithm

In this section, we design a node scheduling algorithm to prolong the network lifetime while maintain the coverage requirement. Before introducing the detail of the proposed algorithm, we first provide the definition of the network lifetime. The network lifetime is the time that the network has worked before one target cannot be covered.

According to the above node deployment scheme, the sensor node has the fixed position while the target is randomly distributed. After the network works a period of time, the energy of some sensor nodes may not be sufficient. We set a threshold β to implement the node energy scheduling function. When the residual energy is smaller than β, the node scheduling algorithm is executed. The scheduling algorithm mainly consists of two parts: sensor selection and relay selection.

1. Sensor Selection

When a sensor covering target $T_j \in \boldsymbol{T}_{\text{overlap}}$ works for a while and its battery reaches the threshold, another sensor with sufficient energy which also covers T_j will switch to active mode to replace this sensor. The sensor with insufficient residual energy will switch to sleep mode until the moment another sensor replacing it reaches the residual energy threshold. The sensor node is charged by the solar energy until it is

full. In the scheduling algorithm, the energy coming from the battery or solar sources is not different. Due to the relative position to the target, some sensor nodes should not be chosen as relays. These sensor nodes construct a set ***S*1**. Thus, during node scheduling, $S_i \in \boldsymbol{S1}$ will never be chosen as the relay. As elements in ***S*1** must keep working, if they exhaust energy, the network will reach the lifetime. Therefore, we use the energy state of sensor nodes in ***S*1** to calculate the network lifetime.

2. Relay Selection

When the residual energy of a relay is less than or equal to the residual energy threshold, the network will select another sensor according to the residual energy state and the distance to replace this relay. The chosen sensor is from the sensor in the range of $2r$. After multiple replacements, some sensors can meet the requirement of distance but cannot satisfy the residual energy requirement. These sensors will be ignored in relay selection, and then, the network further selects a node that is farther away, but the remaining energy is higher than the threshold to work as a relay. With the size of the network increasing, the lifetime of relay becomes an important factor which limits the lifetime of WSN. This is because the relay node undertakes more transmitting tasks. Although there are some nodes replacing them, as the relay nodes with the fastest energy consumption rate, they have the shorter working time and the higher handover rate. Therefore, under some conditions, the network cannot find enough alternative nodes to replace other relays.

By using the above sensor selection and relay selection, we design the node scheduling algorithm. Let ***Ct*** denotes the coverage set. Firstly, by using the initial route path establishment algorithm and sensor selection, the set ***S*1** and set ***Ct*** are obtained. The set ***S*2** is determined accordingly. As the roles of ***S*1** and ***S*2** are different, then different operations are used. For the element of set ***S*1**, there is no replacing sensor node when the residual energy of node S_i reaches the residual energy threshold. Hence, it can only keep working until the energy runs out. When there is one element in set ***S*1** which does not have enough energy, the WSN is dead. If S_i is an element of set ***S*2**, there may be another sensor which can take the place of it to make the network keep working. The replace node selection criterion is defined as follows. Let $a = 1/d_{ij} + E_{\mathrm{r}}^i/C_0$, where d_{ij} denotes the distance between upper level node and the j-th element of the set constructed by possible chosen sensor nodes and E_{r}^i denotes the remaining energy in the battery of the considered sensor node at the end of i-th day. The node with the maximum value of a will be chosen as a relay. Then, the replaced node switches to sleep mode. All nodes can be charged in the daytime no matter whether they are working or not, and the charging rate is determined by the time and the weather, which is defined in (4). When the residual energy of the sensor node is smaller than the residual energy threshold, a new round of relay selection will be started. When the relay selection cannot be executed, the network reaches its lifetime.

4 Simulation Results

According to Fig. 1, 61 sensors are deployed in a 50 m × 50 m area. Referring to [9], we set network parameters as follows. The radius r is 5 meters, the constant value K equals to 28.8, and the efficiency of converting solar energy is 0.05. The capacity of the battery C_0 is 2000 mAh. Other parameters of consuming energy are set as E_{pro} = 50 nJ/bit, E_{amp} = 100 pJ/bit/m^2, and E_{sen} = 0.02 J. As the common overhead of the system is the main influence on the performance of the proposed algorithm, we do not consider the cost of the system in the simulation, such as the energy cost of the routing.

First, under different weather conditions, we observe the network lifetime varying with the target density. The simulation results are illustrated in Fig. 2, which indicates that the network lifetime decreases with the increasing of targets under the same weather condition. With the increasing of targets, the data traffic increases accordingly. Then, numbers of both sensing nodes and relay nodes increase simultaneously, which results in fewer opportunities for going to sleep mode. This decreasing obviously causes a short network lifetime. Moreover, from Fig. 2, it can be found that the network lifetime is proportional to the value of λ. The larger λ indicates higher probability of sunny day. That is, sensors have more charging opportunity and harvest more energy to support sensing and data transmission. However, with the increasing of targets, the difference among the network lifetimes under different λ becomes smaller. This phenomenon illustrates that the energy consumed rate is larger than harvested rate, even when the probability of the sunny day is larger. To guarantee the network sustainability, the effect way is increasing the ratio of the sensor to the target. To verify this consequence, we further observe the network lifetime varying with λ when the network has different numbers of targets. The results are given in Fig. 3.

From the results in Fig. 3, it can be found that larger λ can prolong the network lifetime greatly. When too many targets are in the sensing area, such as more than half of sensors, the increasing range of the lifetime is extremely limited. As shown in

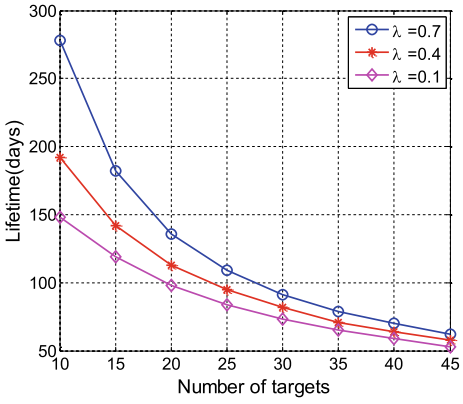


Fig. 2 Network lifetime versus the target density

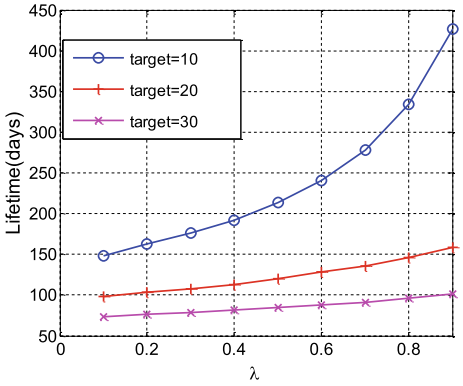


Fig. 3 Network lifetime versus the value of λ

Fig. 3, in the case of 20 targets, the impact of λ is very slight. Thirty or more targets will make this result be more obvious. When the sensing area is fixed, the WSN has more sensors, the longer it will survive. This is because that there are more choices for the sink to find a new relay.

In order to use the minimal number of nodes to cover the target agricultural field from plant to harvest, authors in [9] design an algorithm to schedule the node moving. As the moving of the node would consume extra energy which may be far more than the total of data transmission and sensing, to make the comparison fair, the energy consumed by moving is omitted during the simulation. According the deployment in [9], the number of required sensors can be calculated. It can be found that more sensors are required when covering the same area. Furthermore, we also compare the lifetime achieved by Eto et al. [9] with the network employing our proposed algorithm when $\lambda = 0.4$ and $r = 5$. The comparison results are given in Fig. 4.

In Fig. 4, the line labeled by "proposal" represents the lifetime of the proposed algorithm while the other line "comparison" represents the result of the algorithm in [9]. The comparison result shows that the network using our algorithm can survive

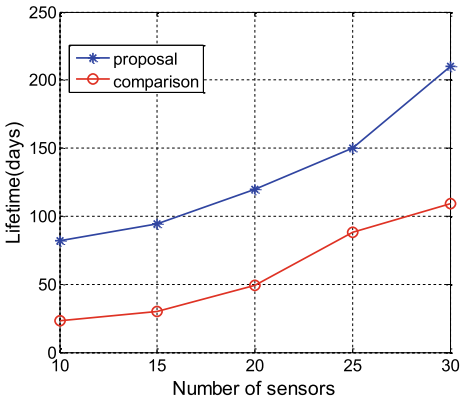


Fig. 4 Comparison of lifetime between two algorithms

longer when the number of sensors is the same. According to [9], the coverage area is divided into several square grids with the side length $r_s/\sqrt{2}$ where r_s is the sensing radius. As each grid needs at least one sensor to cover a square area with the same side length $2r(x-1)$, there are at least $[2(x-1)]^2 = 4x^2-8x+4$ sensor nodes required. To cover the same area, our algorithm only needs $(2x-1)(x-1)+x = 2x^2-2x+1$ sensor nodes. The required node ratio is

$$\rho = \frac{2x^2-2x+1}{4x^2-8x+4} = \frac{1}{2-\frac{4x-2}{2x^2-2x+1}} \tag{6}$$

From Eq. (6), it can be found that the ratio ρ tends to 0.5 when x tends to infinite. Actually, in the 50 m × 50 m area, the ratio is about 61.01%, while in the 70 m × 70 m area, the ratio is about 57.66%. This means that our algorithm requires fewer sensor nodes to cover the same area. Therefore, comparing with [9], the network using our algorithm can work longer when the number of deployed sensors is same, while covering the same area, our algorithm needs fewer sensor nodes.

5 Conclusion

In this paper, we studied the coverage requirement and network lifetime problem of a solar-powered wireless sensor network when the weather condition is considered. To guarantee the coverage requirement and prolong the network lifetime, we propose a node deployment scheme and a scheduling algorithm. The proposed algorithm considers some practical influence factors, such as the weather condition and the relative position of the node to the sun. The simulation results verify that our proposed algorithm can achieve longer lifetime and needs fewer sensor nodes. However, in this paper, we do not consider the system price problem. As the node scheduling algorithm is a central algorithm and is implemented by the sink node, a perfect control wireless channel is needed. This may increase the cost of the system. In the future, a simple and distributed algorithm will be designed.

Acknowledgements This work was supported by Natural Science Foundation of Jiangsu Province (No. BK20170935), National Natural Science Foundation of China (No. 61701241).

References

1. Nimbargi, S.R., Hadawale, S., Ghodke, G.: Tsunami alert & detection system using IoT: a survey. In: International Conference on Big Data, IoT and Data Science (BID), pp. 182–184, Pune, India (2017)
2. Du, J., Wang, K., Liu, H., Guo, D.: Maximizing the lifetime of k-discrete barrier coverage using mobile sensors. IEEE Sens. J. **13**(12), 4690–4701 (2013)

3. Chen, C.P., Mukhopadhyay, S., Chuang, C.L., Liu, M.Y., Jiang, J.A.: Efficient coverage and connectivity preservation with load balance for wireless sensor networks. IEEE Sens. J. **15**(1), 48–62 (2015)
4. Tashtarian, F., Hossein, M., Sohraby, K., Effati, S.: On maximizing the lifetime of wireless sensor networks in event-driven applications with mobile sinks. IEEE Trans. Veh. Technol. **64**(7), 3177–3189 (2015)
5. Nishi, G., Nishant, K., Satbir, J.: Coverage problem in wireless sensor network: a survey. In: IEEE International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 1742–1749, Paralakhemundi, India (2017)
6. Jonathan, D., Hongchi, S.: Maximizing lifetime for k-barrier coverage in energy harvesting wireless sensor networks. In: IEEE Global Communications Conference, pp. 300–304, Austin, USA (2015)
7. Yang, C., Chin, K.W.: Novel algorithms for complete targets coverage in energy harvesting wireless sensor networks. IEEE Commun. Lett. **18**(1), 118–121 (2014)
8. Yang, C., Chin, K.W.: On complete target coverage in wireless sensor networks with random recharging rates. IEEE Wirel. Commun. Lett. **4**(1), 50–53 (2015)
9. Eto, M., Katsuma, R., Tamai, M., Yasumoto, K.: Efficient coverage of agricultural field with mobile sensors by predicting solar power generation. In: 29th International Conference on Advanced Information Networking and Applications, pp. 62–69, Gwangiu (2015)
10. Cardei, M., Thai, M.T., Li, Y., Wu, W.: Energy-efficient target coverage in wireless sensor networks. In: 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), pp. 1976–1984, Miami, USA (2005)
11. Zhao, Q., Gurusamy, M.: Lifetime maximization for connected target coverage in wireless sensor networks. IEEE/ACM Trans. Netw. **16**(6), 1378–1391 (2008)
12. Wang, X., Xing, G., Zhang, Y., Lu, C.: Integrated coverage and connectivity configuration in wireless sensor networks. In ACM International Conference on Embedded Networked Sensor Systems, pp. 28–39 (2003)
13. Zhang, P., Xiao, G., Tan, H.A.: Preliminary study on lifetime maximization in clustered wireless sensor networks with energy harvesting nodes. In: 8th International Conference on Information, Communications and Signal Processing, Singapore (2011)
14. DeWitt, J., Patt, S., Shi, H.: Maximizing continuous barrier coverage in energy harvesting sensor networks. In: IEEE International Conference on Communications, pp. 172–177, Sydney (2014)

A USRP-Based Testbed of Multi-agent Reinforcement Learning for Dynamic Spectrum Anti-Jamming

Lijun Kong, Ximing Wang, Xufang Pei, Luliang Jia, Dianxiong Liu, Kailing Yao, and Yuhua Xu

Abstract In this article, we develop a demonstrated multi-agent dynamic spectrum anti-jamming (MDSA) system using LabVIEW software and USRP-based soft defined radio platform. In the system, we design four subsystems, i.e., wireless transmission subsystem, wideband spectrum sensing subsystem, autonomous decision subsystem, and jamming subsystem. A multi-agent collaborative Q-learning (MACQL) algorithm is adopted in the autonomous decision subsystem to avoid the jamming and the co-channel interference between the agents. The dynamic process of the experiment is illustrated by the screenshots of the software. By showing that the data are successfully received and the performance of the MACQL algorithm is better than the sensing-based method, the MDSA system is realized and the effectiveness of the MACQL algorithm is demonstrated.

1 Introduction

Due to the open nature, wireless communications are vulnerable to the jamming attacks. Useful signals can dodge the jamming by changing to the communication channel where situation is jamming-free, and thanks to the development of cognitive radio, the legitimate users can select the clear channels by sensing before transmission. But wherever there is hiding, there is chasing. The techniques that cope with the dynamic jamming are called dynamic spectrum anti-jamming.

Dynamic spectrum anti-jamming is a promising technology to protect legitimate wireless communications from jamming and improve the spectrum efficiency. Because of the jamming activities, the channel state varies between "idle" and "jammed" dynamically. Therefore, we can solve this anti-jamming problem through

L. Kong (✉) · X. Wang · X. Pei · D. Liu · K. Yao · Y. Xu
College of Communications Engineering, Army Engineering University, Nanjing, China
e-mail: 18763808175@163.com

L. Jia
School of Space Information, Space Engineering University, Beijing, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_4

optimizing channel selection. However, with the development of artificial intelligence, the intelligent jammer may learn the pattern of users' communication, which as a result, the traditional anti-jamming methods may no longer be effective. The research on intelligently anti-jamming techniques is in urgent need. There are some existing works, which can be classified into single-user scenario [1, 2] and multi-user scenario [3, 4], that model the problem of dynamic spectrum anti-jamming as Markov decision process (MDP) and resort to reinforcement learning to solve it. However, in fact, MDP is designed for single-user scenario and [3, 4] only employs reinforcement learning (Q-learning) independently without considering the influence of other users, which may result in the failure of algorithm's convergence [5]. Therefore, in [6], based on the classic Q-learning, a multi-agent collaborative Q-learning (MACQL) algorithm is proposed considering the "cooperation" and "competition" between multi-users, which were ignored in [3, 4]. An application of the MACQL algorithm into the unmanned aerial vehicle (UAV) anti-jamming network can be found in [7]. In [8], a collaborative dynamic spectrum anti-jamming communication framework for multi-user system is proposed, and a collaborative anti-jamming testbed is introduced. The details of the testbed are introduced in this paper.

Soft defined radio platform based on LabVIEW (Laboratory Virtual Instrument Engineering Workbench) [9] and USRP (Universal Software Radio Peripheral) [10] is widely used in customized design of wireless communication systems. In [11], the author built a real-time anti-jamming wireless communication system based on USRP platform, which adopted multiple-input multiple-output and interference cancelation technologies. A cooperative Q-learning method in [2] was proposed to solve the problem of hidden jamming in single-user scenario and was tested on the USRP platform. In [12], the authors designed a channel selection algorithm to protect control link from the jamming attacks and carried out the simulation on a real-life platform. In [13], a single-user reinforcement learning anti-jamming testbed based on the USRP is developed. However, there is no research to realize the dynamic spectrum anti-jamming system for multi-user scenarios based on USRP soft defined radio platform.

Therefore, in this article, we build a multi-agent dynamic spectrum anti-jamming system. Based on [13], the system consists of four subsystems, i.e., wireless transmission subsystem (WTS), wideband spectrum sensing (WBSS) subsystem, autonomous decision subsystem (ADS) and jamming system. Different from [13], the system in this paper contains multiple users. The algorithm is also designed for multi-user scenarios, i.e., MACQL algorithm. With the MACQL algorithm proposed in [6], we can realize the anti-jamming by collaboratively learning the pattern of jamming to make the best channel selection decision, which is demonstrated by the presented testbed.

The rest of the paper is organized as follows. In Sect. 2, the system model is introduced and the MACQL algorithm is reviewed. In Sect. 3, the system composition of the testbed and the implementation process of each subsystem are introduced. In Sect. 4, the simulation results are presented. Finally, the paper is concluded in Sect. 5.

2 System Model and MACQL Algorithm

In this section, we first introduce the system model. Then, in order to facilitate the understanding, we briefly review the proposed MACQL algorithm proposed in [6]. We consider a multi-agent wireless communications network as shown in Fig. 1, assuming there are N agents, one jammer and M available channels in the network. Denote the set of agents as $\mathcal{N} = \{1, 2, \ldots, N\}$, and the set of available channels as $\mathcal{M} = \{1, 2, \ldots, M\}$. In this paper, the agent/user is defined as a communication pair which consists of two nodes: a transmitter and a receiver (learning node). The receivers are in charge of doing the Q-learning to avoid the co-channel interference (CCI) and the jamming signal and completing the intelligent channel selection. Then, the outcome of the learning is transmitted by the receiver via a control link to inform the transmitter of which channel they are going to access in the next slot. To simulate the actual wireless spectrum environment, the jammer can generate different modes of jamming such as sweep, comb, and/or random [14].

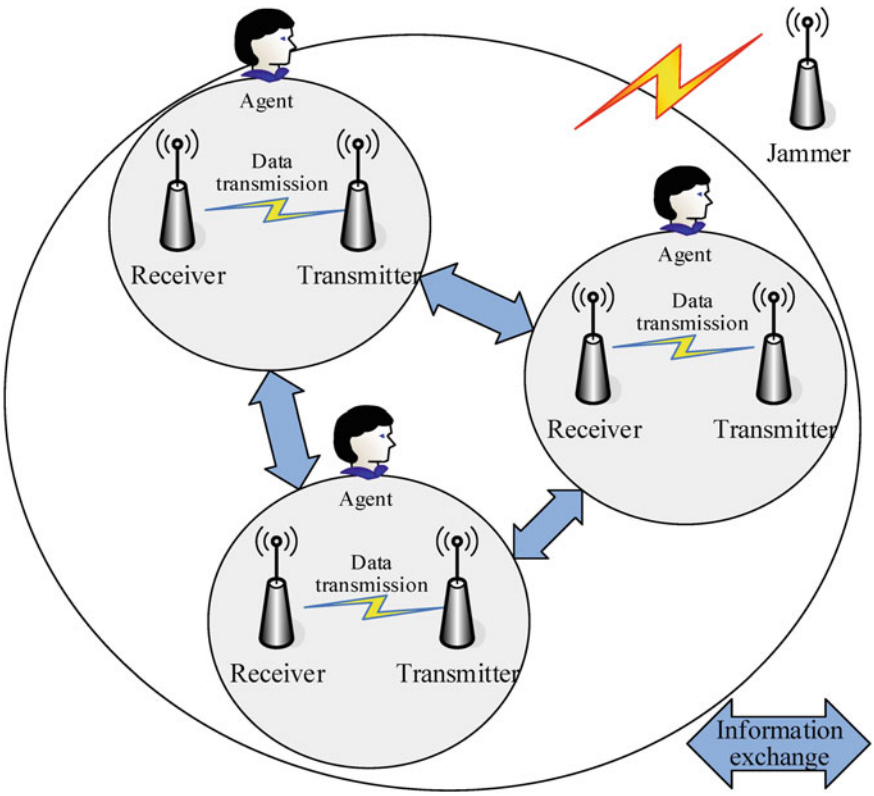


Fig. 1 Multi-agent anti-jamming system model

The authors in [6] resort to the Markov game, which is the extension of the MDP in multi-agent scenario, to model and analyze the system. The state is defined as $s = \{a, f_j\}$, where $a = (a_1, \ldots, a_N)$ is a joint action profile and f_j is the jamming channel. Denote the selected channel of agent n as $a_n = f_n$, and then, the reward of agent n at time slot t is defined as:

$$r_n^t(s,a) = \begin{cases} 1, \text{ if } f_n \neq f_j \& f_n \neq f_m (m \in \mathcal{N}/n) \\ 0, \text{ otherwise.} \end{cases} \tag{1}$$

$$Q_n(s,a) = (1-\lambda) Q_n(s,a) + \lambda [r_n + \gamma V_n(s')], \tag{2}$$

$$V_n(s') = Q_n(s', a^*), \tag{3}$$

where $a^* = \arg\max_{a'} \sum_{n=1}^{N} Q_n(s', a')$, λ and γ are the learning rate and discounting factor, respectively.

In this article, the number of agents is $N = 2$. The principle of the MACQL algorithm is as follows. The agent 1 is in the state $S1_t$ and the agent 2 is in the state $S2_t$. They can obtain each other's evaluation function (Q-value) through information exchange and make joint actions $a1_t$, $a2_t$ according to the evaluation functions. The agents i can obtain the reward value ri_t of action ai_t and enter the new state si_{t+1} after the effect of the environment, where $i = 1, 2$. Then, they update their respective evaluation functions to complete a learning process. The process will be repeated until satisfying the conditions of termination. Figure 2 shows the theoretical simulation result by MATLAB, where the sensing-based method is randomly selecting a

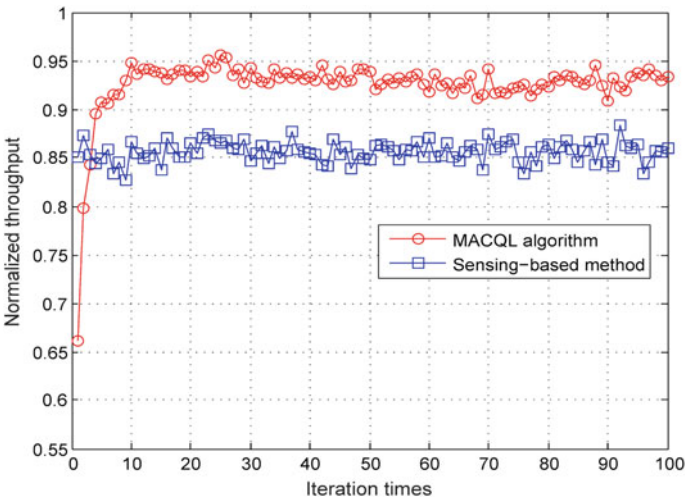


Fig. 2 Theoretical performance of multi-agent collaborative Q-learning algorithm compared with the sensing-based method

jamming-free channel after the WBSS. For more details of MACQL algorithm, the reader is referred to [6].

3 System Implementation

In order to build the testbed of dynamic spectrum anti-jamming system, LabVIEW is used as the software development environment to complete the digital signal processing. The wireless signal is transmitted and received using NI USRP2920 as the hardware platform. The schematic diagram of the transmitter, receiver and jammer and subsystems is shown in Fig. 3. The transmitter consists of one USRP which transmits the data to the receiver via the data channel $f_{\mathrm{DATA}} \in \mathcal{M}$ and receives ACK via the control link f_{ACK}. The receiver consists of two USRPs: one to receive data and send ACK and the other to sense the real-time spectrum signals. In the meanwhile, the jammer radiates the jamming signal via the USRP to disrupt the data

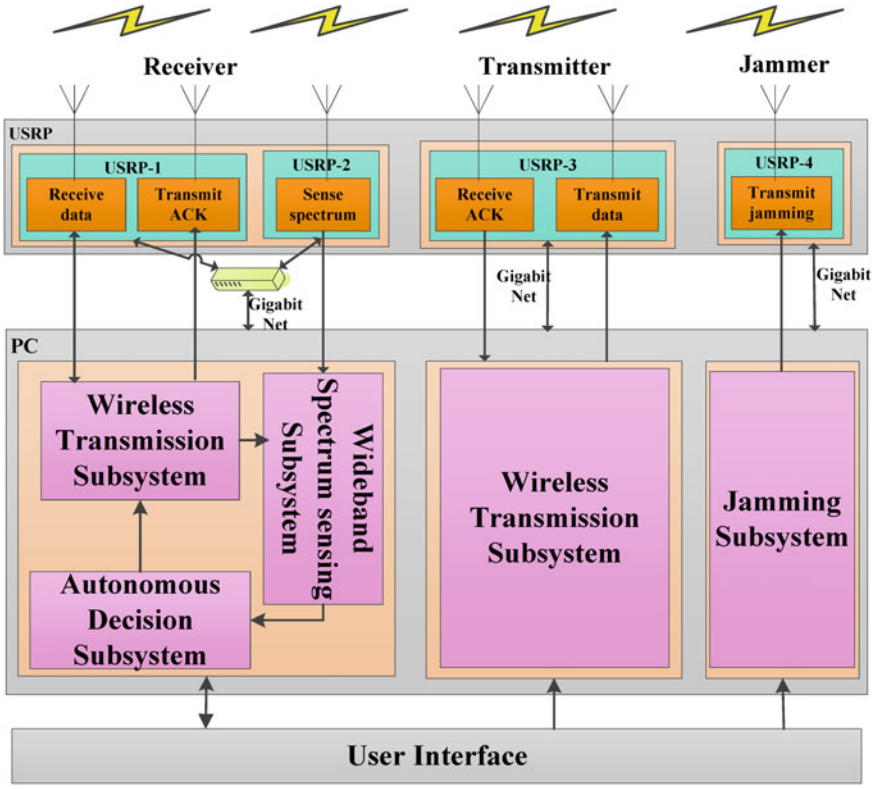


Fig. 3 Subsystems of USRP-based dynamic spectrum anti-jamming system

transmission. As shown in Fig. 3, the system consists of four subsystems: jamming subsystem, WTS, WBSS, and ADS. Next, we will introduce each subsystem in detail.

3.1 Jamming Subsystem and Wideband Spectrum Sensing Subsystem

In the jamming subsystem, we use the discrete-time Fourier inverse transform and the frequency-domain pulse signal to generate the time-domain jamming signal. We assume that the jammer releases the sweep jamming, i.e., the jammer only interferes one channel in each jamming slot T_{jam} and changes the channel in turn periodically. In the WBSS subsystem, we collect time-domain signals by USRP and obtain spectrum data by fast Fourier transform. However, due to the limited processing capability of USRP hardware, we cannot completely cover the entire spectrum only once. Therefore, we adopt the "piecewise-mosaic" design scheme [15] to realize fast WBSS. In each transmission slot T_{s}, the WBSS subsystem identifies the idle state of the channel and obtains the jamming channel.

3.2 Wireless Transmission Subsystem

In [16], we established a wireless transmission system based on ARQ protocol, but we did not have the concept of time slot in it. Considering the concept of time slot in dynamic spectrum access (DSA) system, we implement the classic go-back-N ARQ retransmission protocol. Receiving the data transmitted by the transmitter, the receiver feedbacks ACK through the control channel once in each transmission slot to realize the error control and the DSA between the transmitter and receiver. Here follows the performance of WTS.

The data to be transmitted is handled through the process of bit stream conversion, packet formation, shaping filtering, and modulation, and the transmitter sends it on data channel f_{DATA} through the USRP. At the same time, the receiver uses USRP to cyclically receive data at the same frequency and recovers the original data through the process of packet decomposition, demodulation, validity check, and data extraction. Then, the receiver calculates the lost packet sequence number once per data transmission slot T_{d} and sends it out on the control link f_{ACK} through the ACK frame. When the transmitter receives the ACK, it extracts the retransmission packet sequence number and sends the corresponding data packet; otherwise, it continues to send the data packet. This whole process is repeated until all packets are successfully transmitted.

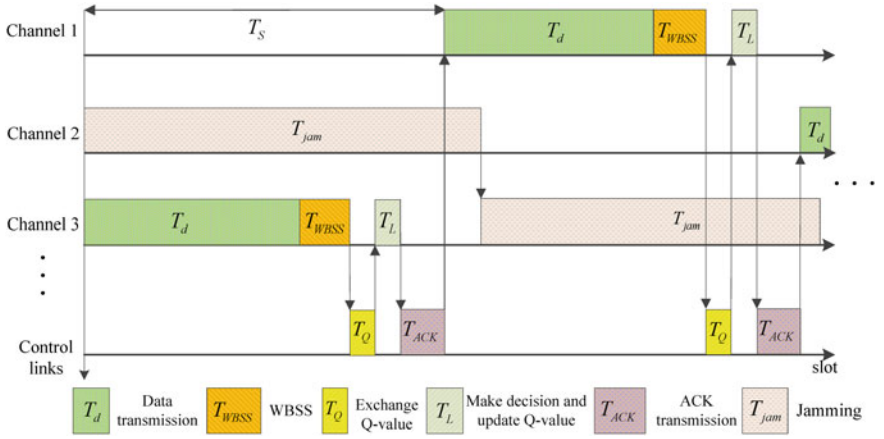


Fig. 4 Illustration of time slot of multi-agent collaborative Q-learning algorithm

3.3 Autonomous Decision Subsystem

Adopting DSA techniques, users can discover the idle spectrum through WBSS and select the optimal idle channel for communication. The easiest way of DSA is to randomly select from the idle channels, but we use the MACQL algorithm to cope with the multi-agent competition and make adversarial decisions toward the jammer, which can achieve better effectiveness and efficiency than the random method and avoid the problem of non-convergence caused by single-agent independent Q-learning.

The illustration of time slot of multi-agent collaborative Q-learning algorithm is shown in Fig. 4. In order to facilitate the demonstration, in the experiment, we choose the wire link to exchange the Q-value instead of wireless link.

4 Results and Discussion

According to the design principle of multi-agent dynamic spectrum anti-jamming system, we built a demonstrated testbed based on USRP soft defined radio platform and LabVIEW software. The system consists of five PCs and seven NI USRP2920, which simulate the jammer and the transmitter and receiver of two users. The wireless communication environment is set as follows: $M = 5$ available channels, $N = 2$ users (agents), one jammer that radiates sweep jamming. The time slot of the jamming signal is $T_{\text{jam}} = 4\,\text{s}$, the data transmission time of users is $T_{\text{d}} = 2\,\text{s}$, the wideband sensing time is $T_{\text{WBSS}} = 0.3\,\text{s}$, the ACK transmission time is $T_{\text{ACK}} = 0.6\,\text{s}$, and the transmission time slot is $T_{\text{s}} = T_{\text{d}} + T_{\text{WBSS}} + T_{\text{ACK}} = 2.9\,\text{s}$. The simulation environment settings are shown in Table 1 [8].

Table 1 Simulation parameters

Type	Connection mode	Function
Initial data frequency	$f_{t,1} = 800\,\text{MHz}$	$f_{t,2} = 804\,\text{MHz}$
ACK control channel frequency	$f_{\text{ACK},1} = 750\,\text{MHz}$	$f_{\text{ACK},2} = 915\,\text{MHz}$
Modulation	8PSK	QPSK
Filter type and parameters	Square root raised cosine rolloff-factor $a = 0.5$	Square root raised cosine rolloff-factor $a = 0.5$

In this system, each user's transmitter sends a picture to the corresponding receiver. The LabVIEW interfaces of user1's receiver is shown in Fig. 5. The first part represents the picture recovered from transmission. The second part shows the DSA sequences of user1 and user2 during communication, which can be acquired through cooperative Q-learning decision-making. The third part displays the normalized throughput to analyze the performance of this system. The fourth part shows the real-time spectrum energy sensed by the WBSS subsystem, by which the state of each channel can be obtained. As we can see, there are one jamming signal and two useful signals. The fifth part shows the channel state calculating from spectrum energy in the fourth part. The sixth part represents the constellation of the receiver transmitting the ACK frame.

From Fig. 5, it can be seen that two users can cooperate with each other in making cooperative decisions to optimize channel selection and avoid jamming. Herein, we can draw the conclusion that the dynamic spectrum anti-jamming in the system is

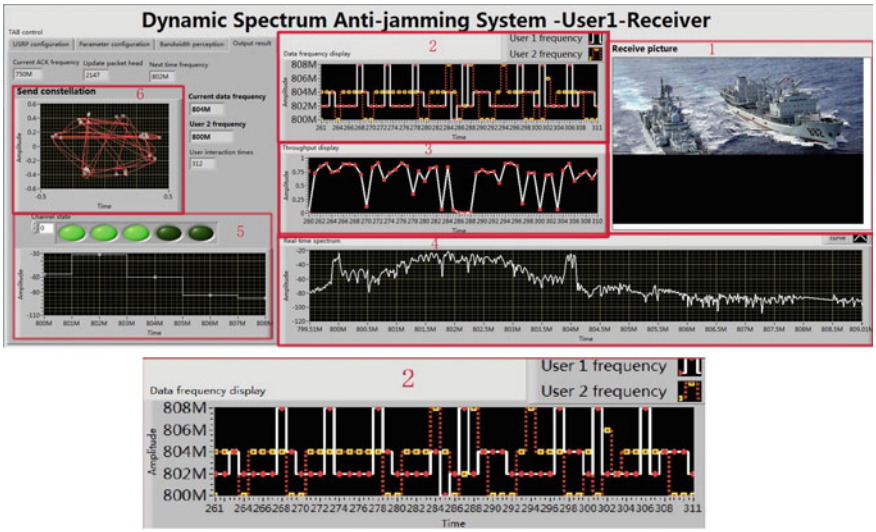


Fig. 5 Interface of user1's receiver

successfully realized, and the proposed algorithm in [6] can be applied into the USRP platform which has practical value.

5 Conclusion

In this paper, we built a multi-agent dynamic spectrum anti-jamming system based on the USRP-based soft defined radio platform and LabVIEW software. Based on the MACQL algorithm, we designed WTS, WBSS, ADS and jamming subsystem. Simulation and measured results showed that the MACQL algorithm had better anti-jamming ability and system throughput than the sensing-based method.

By far, the system can only serve two users since the computation complexity of the collaborative algorithm grows exponentially with the number of users. A more efficient collaborative algorithm for the multi-user system is needed.

Acknowledgements This work was supported by the Natural Science Foundation for Distinguished Young Scholars of Jiangsu Province under Grant No. BK20160034, the National Science Foundation of China under Grant No. 61771488, No. 61671473, No. 61631020.

References

1. Chen, C., Song, M., Xin, C., Backens, J.: A game-theoretical anti-jamming scheme for cognitive radio networks. IEEE Netw. **27**(3), 22–27 (2013)
2. Slimeni, F., Chtourou, Z., Scheers, B., Le Nir, V., Attia, R.: Cooperative Q-learning based channel selection for cognitive radio networks. Wirel. Netw. (2018)
3. Aref, M.A., Jayaweera, S.K., Machuzak, S.: Multi-agent reinforcement learning based cognitive anti-jamming. In: IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6 (2017)
4. Aref, M.A., Jayaweera, S.K.: A novel cognitive anti-jamming stochastic game. In: Cognitive Communications for Aerospace Applications Workshop (CCAA), pp. 1–4 (2017)
5. Vlassis, N.: A concise introduction to multiagent systems and distributed artificial intelligence. Synthesis Lect. Artif. Intell. Mach. Learn. **1**(1), 1–71 (2007)
6. Yao, F., Jia, L.: A collaborative multi-agent reinforcement learning anti-jamming algorithm in wireless networks. IEEE Wirel. Commun. Lett. **8**(4), 1024–1027 (2019)
7. Xu, Y., Ren, G., Chen, J., Zhang, X., Jia, L., Kong, L.: Interference-Aware cooperative anti-jamming distributed channel selection in UAV communication networks. Appl. Sci. **8**(10) (2018)
8. Wang, X., Wang, J., Xu, Y., Chen, J., Jia, L., Liu, X., Yang, Y.: Dynamic spectrum anti-jamming communications: challenges and opportunities. Accepted by IEEE Communications Magazine
9. NI: What Is LabVIEW. http://www.ni.com/en-us/shop/labview.html. Last accessed on 28 June 2019
10. NI: USRP Software Defined Radio Device. http://www.ni.com/en-us/shop/select/usrp-software-defined-radio-device. Last accessed on 28 June 2019
11. Zeng, H.: Demo abstract: An anti-jamming wireless communication system. In: IEEE INFOCOM 2018—IEEE Conference on Computer Communications Workshops, Honolulu, HI, pp. 1–2 (2018)

12. Sharma, N., Rawat, D., Bista, B., Shetty, S.: A testbed using USRP and LabView for dynamic spectrum access in cognitive radio networks. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, Gwangiu, pp. 735–740 (2015)
13. Kong, L., Xu, Y., Pei, X., Ke, M., Wang, X., Bai, W., Feng, Z.: A reinforcement learning approach for dynamic spectrum anti-jamming in fading environment. In: 18th IEEE International Conference on Communication Technology (ICCT'18) (2018)
14. Liu, X., Xu, Y., Jia, L., Wu, Q., Anpalagan. A.: Anti-jamming communications using spectrum waterfall: a deep reinforcement learning approach. IEEE Communi. Lett. **22**(5), 998–1001
15. Ge, J., Chen, X., Ge, L., Kong, L., Pei. X.: Fast wideband spectrum sensing method based on USRP. Commun. Technol. **50**(12), 2714–2721 (2017) (in Chinese)
16. Kong, L., Xu, Y., Chen, X., Zhao, L., Zhou, X.: Design and implementation of data transmission system based on USRP and selective retransmission protocol. Commun. Technol. **51**(06), 1259–1267 (2018). (in Chinese)

Improved DOA Matrix Method for Two-Parallel Uniform Linear Array

Yunxiang Li and Jianfeng Li

Abstract Direction-of-arrival (DOA) matrix method has been proposed for two-parallel uniform linear array (TPULA) due to its advantages in pairing match of 2D angle estimation. However, DOA matrix is constructed only based on sub-covariance matrices, so both the estimation ability and the maximum identifiable source number are limited. An improved DOA matrix method (DMM) is proposed in this paper, which extracts extended subarrays to construct the DOA matrix. Compared to conventional method, the proposed algorithm can not only improve the performance in 2D angle estimation but also increase the maximum identifiable source number. The simulations results verified the advantages of the new approach.

1 Introduction

Direction-of-arrival (DOA) estimation is a major issue for antenna arrays, which have extensive applications in various fields, e.g., radar detection and wireless communication [1–3]. Compared to one-dimensional (1D) DOA estimation that generally uses linear arrays, 2D DOA estimation is more useful in practical situation, and it has been intensively studied over the past decades [4, 5]. Many array geometries have been adopted for the 2D DOA estimation, e.g., the rectangular array [6], uniform circular array (UCA) [7, 8], L-shaped array [9, 10] and TPULA [11, 12]. TPULA has advantage in the pairing match of two-dimensional angle by decomposing the 2D problem into 1D but associated problem.

In [11], the DOA matrix method (DMM) was proposed to exploit the structure of TPULA, and it constructs a DOA matrix whose eigenvalues and the associated eigenvectors are corresponding to the two-dimensional angles, respectively.

Y. Li (✉) · J. Li
College of Computer and Information, Hohai University, Nanjing, China
e-mail: Xiang_lyx@163.com

J. Li
College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_5

Compared to other subspace-based methods, DOA matrix method requires neither peak search nor additional pairing match. Thereafter, DOA matrix method has been applied in many fields for multiple parameter estimation [12]. Meanwhile, propagator method (PM) [13] is well known for its low computational complexity since it needs neither singular value decomposition (SVD) of the received data nor eigenvalue decomposition (EVD) of the cross-covariance matrix. Reference [14] proposed a PM-based fast 2D DOA estimation algorithm for TPULA. Thereafter, an improved PM was proposed in [15], which has better estimation ability and larger aperture than the method in [14]. However, both the DMM and the PM for TPULA utilize the matrix constructed from the original subarrays, whose length limits the estimation performance and identifiable source number.

To settle this problem, we propose an improved DOA matrix method in this paper. Extended subarrays are extracted for the DOA matrix construction, which ensures the maximum identifiable source number and improves the estimation ability.

The rest of this paper is structured as follows. Section 2 develops the array model of TPULA. Section 3 reviews two conventional 2D DOA estimation method: the conventional DMM and the improved PM. Section 4 describes the improved DMM. Section 5 discusses the maximum identifiable source number for the traditional DMM and the proposed one, respectively. Section 6 presents numerical simulations to verify the advantages of the new approach. Section 7 gives the conclusion.

2 Array Model

Figure 1 shows the array geometry of TPULA. It is arranged in the x-y plane. Subarray 1 is arranged along Y-axis and subarray 2 is parallel to subarray 1 with half-wavelength distance, and both of them have M elements. Consider that there are K far-field sources impinging on the array with a 2D angle, and α and β are shown in Fig. 1. Then, the outputs are

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}_x(t) \tag{1}$$

$$\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{s}(t) + \boldsymbol{n}_y(t) \tag{2}$$

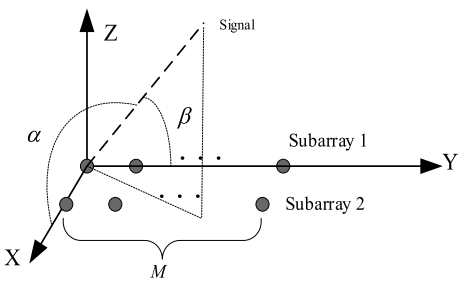


Fig. 1 Array geometry of TPULA

where $\boldsymbol{\Phi} = \mathrm{diag}\left(\mathrm{e}^{-\mathrm{j}\pi\ \cos\beta_1}, \ldots, \mathrm{e}^{-\mathrm{j}\pi\ \cos\beta_K}\right)$, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_K(t)]^{\mathrm{T}}$ is the source signal vector. $\boldsymbol{n}_x(t)$ and $\boldsymbol{n}_y(t)$ are additive white Gaussian noise vectors. $(.)^{\mathrm{T}}$ means transposition. $\boldsymbol{A}$ is the direction matrix of subarray 1 with column vector being

$$\boldsymbol{a}(\alpha_k) = \left[1, \mathrm{e}^{-\mathrm{j}\pi\ \cos\alpha_k}, \ldots \mathrm{e}^{-\mathrm{j}\pi(M-1)\cos\alpha_k}\right]^{\mathrm{T}}, \quad k = 1, \ldots, K \tag{3}$$

3 Conventional Method

3.1 Conventional DOA Matrix Method

The covariance matrix of $\boldsymbol{x}(t)$ can be obtained by

$$\boldsymbol{R}_{\boldsymbol{xx}} = E\left[\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)\right] = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^{\mathrm{H}} + \sigma^2 I_M \tag{4}$$

where $E[.]$ denotes the operation of expectation, $(.)^{\mathrm{H}}$ denotes conjugate transposition and the matrix $\boldsymbol{R}_s = \mathrm{diag}\left\{E\left[s_1(t)s_1^{\mathrm{H}}(t)\right], \ldots, E\left[s_K(t)s_K^{\mathrm{H}}(t)\right]\right\} \in R^{K\times K}$ is diagonal, representing the co-matrix of $\boldsymbol{s}(t)$.

Assume noise and signals are uncorrelated, the cross-correlation matrices of the outputs measured by subarray 1 and 2 are

$$\boldsymbol{R}_{\boldsymbol{yx}} = E\left[\boldsymbol{y}(t)\boldsymbol{x}^{\mathrm{H}}(t)\right] = \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{R}_s\boldsymbol{A}^{\mathrm{H}} \tag{5}$$

Perform EVD to $\boldsymbol{R}_{\boldsymbol{xx}}$, there are K signal eigenvalues and $M - K$ repeated noise eigenvalues. Then, we can remove the noise component from $\boldsymbol{R}_{\boldsymbol{xx}}$ as

$$\boldsymbol{C}_{\boldsymbol{xx}} = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{R}_{\boldsymbol{xx}} - \hat{\sigma}^2\boldsymbol{I}_M \tag{6}$$

where $\hat{\sigma}^2$ is the estimation of the noise variance, which can be obtained from the noise eigenvalues.

Then, DOA matrix $\boldsymbol{R} \in C^{M\times M}$ is defined as

$$\boldsymbol{R} = \boldsymbol{R}_{\boldsymbol{yx}}\boldsymbol{C}_{\boldsymbol{xx}}^{+} \tag{7}$$

where $\boldsymbol{C}_{\boldsymbol{xx}}^{+}$ is the pseudo-inversion of $\boldsymbol{C}_{\boldsymbol{xx}}$.

Assume $\boldsymbol{A}$ and $\boldsymbol{R}_s$ are full rank, then $\boldsymbol{\Phi}$ has K nonzero unequal diagonal entries. It can be verified from Eqs. (4)–(7) that the K eigenvalues and their corresponding eigenvectors of $\boldsymbol{R}$ are equal to the elements of $\boldsymbol{\Phi}$ and the column vectors of $\boldsymbol{A}$, respectively.

$$\boldsymbol{R}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{\Phi} \tag{8}$$

Perform EVD to $\boldsymbol{R}$ and then the estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{A}$ can be obtained through the eigenvalues and the eigenvectors of $\boldsymbol{R}$.

However, although DMM requires neither peak search nor pairing match, it has a major disadvantage: From Eqs. (5) to (7), we can see that the DOA matrix is constructed from original subarrays and the dimension is $M \times M$, so the estimation performance and identifiable source number are limited due to the limited aperture.

3.2 Propagator Method (PM)

The propagator method (PM) [13] is well known for its low computational complexity since it needs neither singular value decomposition (SVD) of the received data nor eigenvalue decomposition (EVD) of the cross-co-matrix. Wu proposed a fast DOA estimation algorithm which has lower complexity [14]. Thereafter, an improved PM-based method was proposed in [15], whose estimation performance was verified via several simulations.

Define $\boldsymbol{A}_{12} = \left[\boldsymbol{A}^{\mathrm{T}}, \boldsymbol{\Phi}\boldsymbol{A}^{\mathrm{T}}\right]^{\mathrm{T}}$ and partition $\boldsymbol{A}_{12}$ as

$$\boldsymbol{A}_{12} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{bmatrix} \tag{9}$$

where $\boldsymbol{A}_1 \in C^{K\times K}$, $\boldsymbol{A}_2 \in C^{(2M-K)\times K}$, then $\boldsymbol{A}_2$ is a linear transformation of $\boldsymbol{A}_1$. The propagator matrix $\boldsymbol{P} \in C^{K\times(2M-K)}$ is defined as

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}_1 = \boldsymbol{A}_2 \tag{10}$$

Let $\boldsymbol{R}_e = (1/J)\sum_{t=1}^{J} \boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)$ be the co-matrix of the received data, and the partition $\boldsymbol{R}_e$, as

$$\boldsymbol{R}_e = \left[\boldsymbol{R}_1 \ \boldsymbol{R}_2\right] \tag{11}$$

where $\boldsymbol{R}_1 \in C^{2M\times K}$, $\boldsymbol{R}_2 \in C^{2M\times(2M-K)}$. For the noise-free case, $\boldsymbol{R}_2 = \boldsymbol{R}_1\boldsymbol{P}$.

But, in fact, it is hardly possible to be noise-free, and the following minimization problem can be used to estimate $\boldsymbol{P}$:

$$J_{\mathrm{csm}}(\boldsymbol{P}) = \|\boldsymbol{R}_2 - \boldsymbol{R}_1\boldsymbol{P}\|_F^2 \tag{12}$$

where $\|\cdot\|$ is the Frobenius norm. Then, $\boldsymbol{P}$ is estimated by least squares (LS)

$$\hat{\boldsymbol{P}} = \left(\boldsymbol{R}_1^{\mathrm{H}}\boldsymbol{R}_1\right)^{-1}\boldsymbol{R}_1^{\mathrm{H}}\boldsymbol{R}_2 \tag{13}$$

where $(.)^{-1}$ denotes inversion. We define $\boldsymbol{P}_c = \begin{bmatrix} \boldsymbol{I}_{K\times K} \\ \hat{\boldsymbol{P}}^H \end{bmatrix}$, and then, $\boldsymbol{P}_c\boldsymbol{A}_1 = \boldsymbol{A}_{12}$ and the partition $\boldsymbol{P}_c$ is:

$$\boldsymbol{P}_c = \begin{bmatrix} \boldsymbol{P}_x \\ \boldsymbol{P}_y \end{bmatrix}, \boldsymbol{P}_x \in C^{M\times K}, \boldsymbol{P}_y \in C^{M\times K} \tag{14}$$

Based on Eqs. (8)–(13), it is derived that

$$\boldsymbol{P}_x\boldsymbol{A}_1 = \boldsymbol{A} \tag{15}$$

$$\boldsymbol{P}_y\boldsymbol{A}_1 = \boldsymbol{A}\boldsymbol{\Phi} \tag{16}$$

then

$$\boldsymbol{\Psi} = \boldsymbol{P}_x^+\boldsymbol{P}_y = \boldsymbol{A}_1\boldsymbol{\Phi}\boldsymbol{A}_1^{-1} \tag{17}$$

We perform EVD to $\boldsymbol{\Psi}$. The estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{A}_1$ can be obtained from the eigenvalues and the eigenvectors of $\boldsymbol{\Psi}$. Let us assume that $\hat{\xi}_k$ is the kth diagonal element of $\hat{\Phi}$.

Then, we can define $\boldsymbol{B}_1 = \begin{bmatrix} \boldsymbol{P}_1 \\ \boldsymbol{P}_3 \end{bmatrix}\boldsymbol{A}_1$, $\boldsymbol{B}_2 = \begin{bmatrix} \boldsymbol{P}_2 \\ \boldsymbol{P}_4 \end{bmatrix}\boldsymbol{A}_1$, where $\boldsymbol{P}_1$,$\boldsymbol{P}_2$ are the first to the $(M-1)$th rows of $\boldsymbol{P}_x$, $\boldsymbol{P}_y$; $\boldsymbol{P}_3$, $\boldsymbol{P}_4$ are the second to the Mth rows of $\boldsymbol{P}_x$, $\boldsymbol{P}_y$. Let us assume that $\hat{\gamma}_k$ is the kth diagonal entry of $\boldsymbol{B}_1^+\boldsymbol{B}_2$. The estimates of the 2D angle are

$$\hat{\alpha}_k = \cos^{-1}\left(\text{angle}\left(\hat{\xi}_k\right)/\pi\right) \tag{18}$$

$$\hat{\beta}_k = \cos^{-1}\left(\text{angle}\left(\hat{\gamma}_k\right)/\pi\right) \tag{19}$$

where angle(.) is to get the phase.

4 Improved DOA Matrix Method

The original subarrays shown in Eqs. (1)–(2) are corresponding to only half of the sensor number, which limits the utilized degrees of freedom (DOF). Both the DMM and the PM for TPULA utilize the matrix constructed from the original subarrays, whose length limits the estimation performance and identifiable source number. Due to the uniformity of the original subarrays, we can extract two extended subarrays with outputs being:

$$\boldsymbol{x}_a(t) = \begin{bmatrix} x_1(t) \\ y_1(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_x \\ \boldsymbol{A}_x \boldsymbol{\Phi} \end{bmatrix} \boldsymbol{s}(t) + \boldsymbol{n}_a(t) \tag{20}$$

$$\boldsymbol{x}_b(t) = \begin{bmatrix} x_2(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_x \\ \boldsymbol{A}_x \boldsymbol{\Phi} \end{bmatrix} \boldsymbol{\Phi}_x \boldsymbol{s}(t) + \boldsymbol{n}_b(t) \tag{21}$$

where $\boldsymbol{x}_i(t)$, $\boldsymbol{y}_i(t)$, $i = 1, 2$ are the first and last $(M-1)$ rows of $\boldsymbol{x}(t)$, $\boldsymbol{y}(t)$. $\boldsymbol{A}_x$ denotes the first $(M-1)$ rows of $\boldsymbol{A}$, $\boldsymbol{\Phi}_x = \text{diag}\left(\text{e}^{-\text{j}\pi \cos\alpha_1}, \ldots, \text{e}^{-\text{j}\pi \cos\alpha_K}\right)$. Then, the co-matrix of $\boldsymbol{x}_a(t)$ is

$$\boldsymbol{R}_a = E\left[\boldsymbol{x}_a(t)\boldsymbol{x}_a^{\text{H}}(t)\right] = \boldsymbol{A}_E \boldsymbol{R}_s \boldsymbol{A}_E^{\text{H}} + \sigma^2 \boldsymbol{I}_{2M-2} \tag{22}$$

where $\boldsymbol{A}_E = [\boldsymbol{A}_x^{\text{T}}, \boldsymbol{\Phi}\boldsymbol{A}_x^{\text{T}}]^{\text{T}}$ and $\boldsymbol{R}_s = E[\boldsymbol{s}(t)\boldsymbol{s}^{\text{H}}(t)]$. The cross-co-matrix between $\boldsymbol{x}_b(t)$ and $\boldsymbol{x}_a(t)$ can be obtained by

$$\boldsymbol{R}_{ba} = E\left[\boldsymbol{x}_b(t)\boldsymbol{x}_a^{\text{H}}(t)\right] = \boldsymbol{A}_E \boldsymbol{\Phi}_x \boldsymbol{R}_s \boldsymbol{A}_E^{\text{H}} \tag{23}$$

Then, the DOA matrix is constructed as follows

$$\boldsymbol{H} = \boldsymbol{R}_{ba}(\boldsymbol{R}_a - \sigma^2 \boldsymbol{I}_{2M-2})^{+} \tag{24}$$

Based on Eqs. (22)–(24), it is derived that

$$\boldsymbol{H}\boldsymbol{A}_E = \boldsymbol{A}_E \boldsymbol{\Phi}_x \tag{25}$$

We perform EVD to $\boldsymbol{H}$, and Eq. (25) indicates that the estimates of $\boldsymbol{\Phi}_x$ and $\boldsymbol{A}_E$ can be obtained through the eigenvalues and the eigenvectors of $\boldsymbol{H}$. Let us assume that $\hat{\mu}_k$ is the kth diagonal element of $\hat{\Phi}_x$. $\boldsymbol{A}_E = [\boldsymbol{a}_{e1}, \boldsymbol{a}_{e2}, \ldots \boldsymbol{a}_{eK}]$, and the partition $\boldsymbol{a}_{ek}$—as

$$\boldsymbol{a}_{ek} = \begin{bmatrix} \boldsymbol{a}_{k_1} \\ \boldsymbol{a}_{k_2} \end{bmatrix} \tag{26}$$

where $\boldsymbol{a}_{k_1}, \boldsymbol{a}_{k_2}$ are the first and last $(M-1)$ rows of $\boldsymbol{a}_{ek}$. Let us assume $\hat{\varpi}_k = \boldsymbol{a}_{k_1}^{+}\boldsymbol{a}_{k_2}$, the estimates of the 2D angle are

$$\hat{\alpha}_k = \cos^{-1}\left(\text{angle}\left(\hat{\mu}_k\right)/\pi\right) \tag{27}$$

$$\hat{\beta}_k = \cos^{-1}\left(\text{angle}\left(\hat{\varpi}_k\right)/\pi\right) \tag{28}$$

As now the DOA matrix is constructed from extended subarrays, improved estimation performance and increased identifiable source number can be expected.

5 Maximum Identifiable Source Number

In this part, we discuss the maximum identifiable source number for the traditional DMM and the proposed one, respectively. The conventional DOA matrix method for TPULA is constructed from the original subarrays, whose length limits the identifiable source number. To solve the problem, the improved DOA matrix was proposed, which extracts two extended subarrays for the DOA matrix construction.

In the conventional DMM, as discussed in Sect. 3, the diagonal entries of $\boldsymbol{\Phi}$ and the column vectors of $\boldsymbol{A}$ can be obtained after we perform EVD to $\boldsymbol{R}$, where $\boldsymbol{A} = [\boldsymbol{a}(\alpha_1), \ldots, \boldsymbol{a}(\alpha_K)]$ is the direction matrix. To ensure $\boldsymbol{A}$ full column rank, at least M sources can be identified. In the improved DMM, as discussed in Sect. 4, we extract two extended subarrays, as shown in Eq. (23), and the diagonal elements of $\boldsymbol{\Phi}_x$ and the column vectors of $\boldsymbol{A}_E$ can be obtained after we perform EVD to $\boldsymbol{H}$, where $\boldsymbol{A}_E \in C^{(2M-2)\times K}$. To ensure $\boldsymbol{A}_E$ full column rank, the maximum identifiable source number is $2M - 2$. Obviously, when $M > 2$, compared to the conventional method, the proposed one can identify more sources.

6 Simulations

Several simulations are performed to verify the superiority of the new approach. Assume there are three uncorrelated sources with the following angles being $(\alpha_1, \beta_1) = (50°, 55°)$, $(\alpha_2, \beta_2) = (65°, 40°)$ and $(\alpha_3, \beta_3) = (80°, 75°)$, respectively. TPULA is configured with $M = 4$, and 200 samples are collected to estimate the covariance matrix.

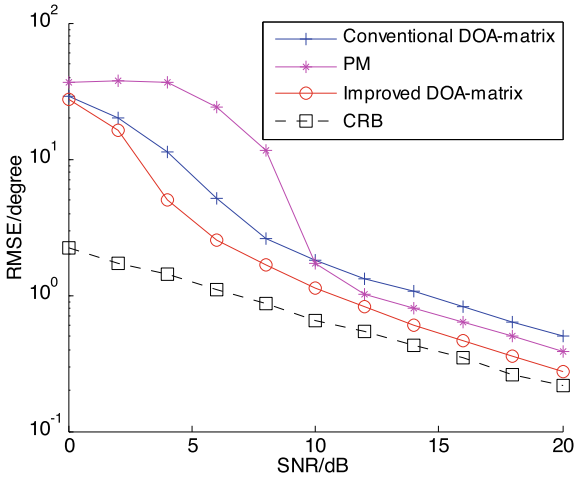


Fig. 2 Estimation comparison among the conventional DOA matrix, the improved PM and the proposed DOA matrix method as well as the CRB

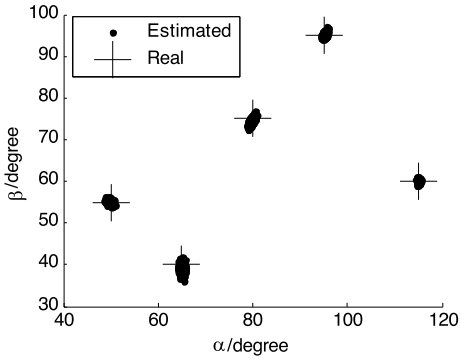


Fig. 3 Estimation results with $K = 5$ sources for the proposed method

Figure 2 gives the estimation performance comparison among the conventional DOA matrix, the improved PM and the proposed DOA matrix method as well as the Cramer Rao bound (CRB) [16], where the RMSE is used for the measurement [12]. It is shown that the proposed method outperforms the conventional method and the PM.

Figure 3 gives the estimation results of the improved DMM with $K = 5$ sources, which make conventional methods fail. Therefore, the proposed method can increase identifiable source number.

7 Conclusion

In this paper, we propose an improved DMM, which extracts two extended subarrays for construction. Compared to two conventional 2D DOA estimation methods, the conventional DMM and improved propagator method, the new approach has two following advantages: (1) It has better performance in 2D DOA estimation; (2) it can increase the maximum identifiable source number. These advantages have been verified by the simulations.

Acknowledgements This work is supported by National Science Foundation of China (61601167), the Fundamental Research Funds for the Central Universities (NT2019013).

References

1. Wang, X., Wang, L., Li, X., Bi, G.: Nuclear norm minimization framework for DOA estimation in mimo radar. Sig. Process. **135**, 147–152 (2017)
2. Wen, F., Zhang, Z., Wang, K., et al.: Angle estimation and mutual coupling self-calibration for ULA-based bistatic MIMO radar. Sig. Process. **144**, 61–67 (2018)
3. Li, J., Jiang, D., Zhang, X.: DOA estimation based on combined unitary ESPRIT for coprime MIMO radar. IEEE Commun. Lett. **21**(1), 96–99 (2017)

4. Wu, Q., Ding, G., Wang, J., Yao, Y.: Spatial-temporal opportunity detection for spectrum-heterogeneous cognitive radio networks: two-dimensional sensing. IEEE Trans. Wireless Commun. **12**(2), 516–526 (2013)
5. Wan, L., Kong, X., Xia, F.: Joint range-Doppler-angle estimation for intelligent tracking of moving aerial targets. IEEE Internet Things J. **5**(3), 1625–1636 (2018)
6. Zoltowski, M.D., Haardt, M., Mathews, C.P.: Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT. IEEE Trans. Signal Process. **44**(2), 316–328 (1996)
7. Liao, B., Wu, Y.T., Chan, S.C.: A generalized algorithm for fast two-dimensional angle estimation of a single source with uniform circular arrays. IEEE Antennas Wirel. Propag. Lett. **11**, 984–986 (2012)
8. He, M., Yin, Y., Zhang, X.: UCA-ESPRIT algorithm for 2-D angle estimation. In: International Conference on Signal Processing (2002)
9. Liang, J., Liu, D.: Joint elevation and azimuth direction finding using L-shaped array. IEEE Trans. Antennas Propag. **58**(6), 2136–2141 (2010)
10. Li, J., Jiang, D.: Joint elevation and azimuth angles estimation for L-shaped array. IEEE Antennas Wirel. Propag. Lett. **16**, 453–456 (2017)
11. Yin, Q.Y., Newcomb, R.W., Zou, L.H.: Estimating 2-D angles of arrival via two parallel linear arrays. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2803–2806 (1989)
12. Cao, Y.: Joint estimation of angle and Doppler frequency for bistatic MIMO radar. Electron. Lett. **46**(2), 170–172 (2010)
13. Marcos, S., Marsal, A., Benidir, M.: The propagator method for source bearing estimation. Sig. Process. **42**(2), 121–138 (1995)
14. Wu, Y., Liao, G., So, H.C.: A fast algorithm for 2-D direction-of-arrival estimation. Sig. Process. **83**(8), 1827–1831 (2003)
15. Li, J., Zhang, X., Chen, H.: Improved two-dimensional DOA estimation algorithm for two-parallel uniform linear arrays using propagator method. Sig. Process. **92**, 3032–3038 (2012)
16. Stoica, P., Nehorai, A.: Performance study of conditional and unconditional direction-of-arrival estimation. IEEE Trans. Signal Process. **38**, 1783–1795 (1990)

UKF Applied in Target Tracking Based on Single Observation Station

Bo Ren and Huan Li

Abstract Unscented Kalman filter is used to tackle nonlinear tracking target moving with uniform velocity in straight line, based on a single observation station. Symmetrical set unscented transform is applied to design a target tracking system which includes a two-dimensional and a three-dimensional state-space model.

1 Introduction

The technology of target tracking and locating is widely applied in kinds of works such as guidance and control, thus it takes a crucial role in national defense. However, with the increasing motility of the observed objects, it is a tough problem for observer to follow the tracks of extremely flexible targets in nonlinear state. Therefore, how to track the nonlinear and highly mobile system more steadily and precisely becomes an essential field in research.

Owing to the whole information applied into the system's model origins from time domain, Kalman filtering can not only estimate stationary and multidimensional stochastic processes, but also be recursive so that it is convenient to be calculated on computers in real time [1]. Nevertheless, the conventional Kalman filter is based on a linear mathematical model, and the highly maneuverable missiles, aircrafts, and other observed targets nowadays belong to nonlinear system to a large degree. So, the conventional Kalman filter is not that perfectly applicable. Some newly filters need to be studied to solve that problem [2].

A classical and traditional way to solve the nonlinear problems is the linear approximation of nonlinear functions and neglecting or approximating the higher-order terms. One of the most widely used is the extended Kalman filter (EKF). Unscented

B. Ren
College of Equipment Engineering, Shenyang Ligong University, Shenyang 110159, China

H. Li (✉)
School of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China
e-mail: lihuan9999@yeah.net

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_6

Kalman filter (UKF), a kind of Kalman filter, is based on unscented transform, whose principle totally differs from extended Kalman filter. UKF is another approximating method to deal with nonlinear distribution by sampling strategy, which is based on unscented transform (UT) and adopts Kalman linear filtering framework. Furthermore, the specific sampling form of UKF is deterministic sampling rather than particle filtering random sampling [3].

Nonlinear filtering UKF has been widely applied in kinds of fields. It was first utilized for navigation and tracking, such as missile reentry, autonomous robot positioning, ground vehicle navigation, aircraft attitude control, and image tracking. More recently, it has been applied in random signal processing, speech recognition and enhancement. In addition, the combination of UKF and particle filtering can achieve a better filtering effect through the complementation of both advantages and disadvantages of the two, and the application of UKF in the artificial neural network has also been concerned by scholars in a wide range [4].

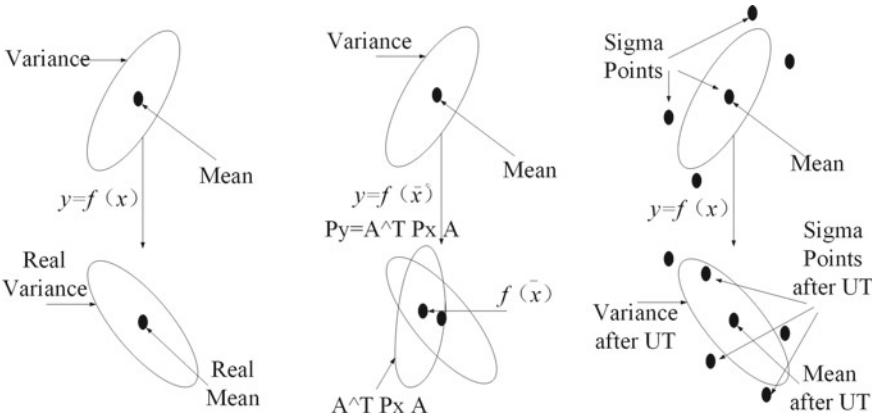
The nonlinear problem exists in tracking target moving with uniform velocity in a straight line. UKF shows an excellent performance in tracking the nonlinear system.

Two-dimensional and three-dimensional state-space models are established to describe the state of an observed object. In a further discussion of the tracking system's ability of anti-interference, noise is increased in the observed nonlinear system and the simulation results seem to be quite desirable with deviation within reasonable range during 10-min observing period. And the deviation presents a trend of oscillatory convergence.

2 Unscented Kalman Filter Algorithm

UKF is based on the UT transform and Kalman linear filtering framework. The sampling form is deterministic sampling rather than particle filtering random sampling. The number of particles (generally referred to as Sigma points) sampled by UKF is very small, and the number of specific particles is dependent on the selected sampling strategy. The most commonly used strategy is $2n + 1$ Sigma points symmetric sampling. One of the noises existing in nature is the white noise whose probability density function is symmetrical Gauss distribution.

The essence of unscented transform is: in the premise to ensure the sample's mean and covariance for x and Px, select a set of points (Sigma points), then apply nonlinear transform to each sampling point Sigma. We will get set y and Py through nonlinear conversion and they are statistics of the original set of Sigma points after transformation. It is easier to approximate the Gauss distribution by using a fixed number of parameters than approximate nonlinear function or transformation. And the principle is: to get some points in the original state of distribution according to some rules in the condition of mean and covariance of these points are corresponding equal to the mean and covariance of the original distribution; then, these points we selected will be substituted into nonlinear function to obtain the set of points of nonlinear function values; at last, we could figure out mean and covariance from the set of points after transformation. That is to say, without linearization or neglecting



(a) Actual distribution (b) after EKF (c) after UKF

Fig. 1 Distribution of a set of points

the higher-order terms, the nonlinear function could be resolved properly and gain a higher accuracy than EKF. The distribution of a set of points is shown in Fig. 1. The actual distribution and the distributions after EKF and UKF are (a), (b), and (c), respectively.

It is assumed that both the system noise and the observed noise are Gauss noise models. Therefore, the unscented Kalman filter in this paper selects the symmetric x Sigma Set UT, also known as the symmetric set unscented transformation, which obtains the distribution of the set of σ points symmetrical about mean. The basic principle of the unscented transformation of symmetric sets is as follows.

Let us assume a nonlinear transformation $y = f(x)$. The state vector x is an n-dimensional random variable, in which $\bar{x}$ and P are known as its mean and variance. Then, we can obtain the $2n + 1$ Sigma points X and the corresponding weight ω for the following UT to calculate the statistical characteristics of y.

1. Figure out $2n + 1$ Sigma points X (sampling points), where n refers to the dimension of states.

$$\begin{cases} X^{(0)} = \overline{X}, & i = 0 \\ X^{(i)} = \overline{X} + \left(\sqrt{(n+\lambda)P}\right)_i, & i = 1 \sim n \\ X^{(i)} = \overline{X} + \left(\sqrt{(n+\lambda)P}\right)_i, & i = n+1 \sim 2n \end{cases} \tag{1}$$

where $\left(\sqrt{P}\right)^{\mathrm{T}}\left(\sqrt{P}\right) = P$, $\left(\sqrt{P}_i\right)$ stands for column i of the square root of matrix.

2. Work out the corresponding weights of these sampling points.

$$\begin{cases} \omega_m^{(0)} = \frac{\lambda}{n+\lambda}, \\ \omega_c^{(0)} = \frac{\lambda}{n+\lambda} + \left(1 - \alpha^2 + \beta\right) \\ \omega_m^{(i)} = \omega_c^{(i)} = \frac{\lambda}{2(n+\lambda)}, \ i = 1 \sim 2n \end{cases} \tag{2}$$

where superscript i is the number of sampling point.

Parameter $\lambda = \alpha^2(n + \kappa) - n$ is a scaling parameter to reduce the total deviation of prediction.

α controls the distribution state of the sampling point; κ is the parameter to be determined, although there is no limit to the specific value, it is usually necessary to ensure that the matrix $(n + \lambda)P$ is a positive semidefinite matrix; $\beta \geq 0$ is a parameter to be determined, who can merge the motion difference of the higher-order terms in the equation, so that the influence of the higher-order term can be taken into consideration [5].

3 Mathematical Model of Observed Target

Assuming there is a target moving with uniform velocity in a straight line, we observe that project is based on a single observation station in a condition that the initial parameters (location, velocity) are known at first. Then, we try to work out the real-time state and trajectory of the observed one through UKF. Finally, figure out the deviation between the trajectories we get based on UKF and the actual one, then obtain the curve diagram of deviation we analyzed from the above steps.

Observed target's mathematical model consists of state equation and measurement equation. And after discretization, that model could be substituted into Kalman through symmetrical unscented transform. Then, we will get the trajectory of the target.

Aiming at the moving objects in the aforesaid assumption, we establish a two-dimensional model (in Fig. 2) and a three-dimensional model (in Fig. 3) which are corresponding to simulation of tank and airplane in reality as follows.

Operational equation of two-dimensional model:

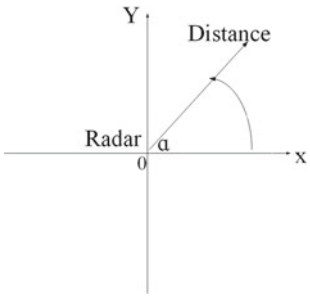


Fig. 2 Model in two-dimensional space

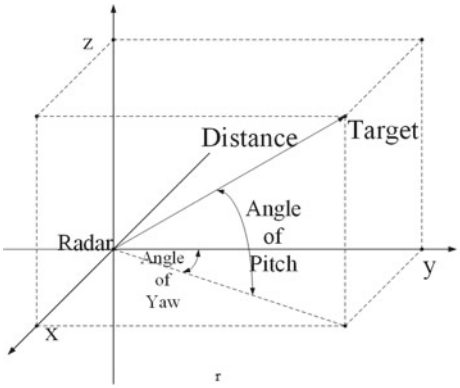


Fig. 3 Model in three-dimensional space

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ y_1(k+1) \\ y_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \\ y_1(k) \\ y_2(k) \end{bmatrix} + \begin{bmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & T^2/2 \\ 1 & T \end{bmatrix} \cdot \begin{bmatrix} a_x(k) \\ a_y(k) \end{bmatrix} \tag{3}$$

$$Z(k) = \begin{bmatrix} \sqrt{(x_1(k) - x_0)^2 + (y_1(k) - y_0)^2} + V(k) \\ \mathcal{Y} \end{bmatrix} \tag{4}$$

In (4), $\mathcal{Y}$ is angle of yaw, and here comes the discussion of the value of $\mathcal{Y}$:
Target in the first quadrant, $\mathcal{Y} = \left|\tan^{-1} \frac{y_1(t)}{x_1(t)}\right|$;
Target in the second quadrant, $\mathcal{Y} = \pi - \left|\tan^{-1} \frac{y_1(t)}{x_1(t)}\right|$;
Target in the third quadrant, $\mathcal{Y} = \pi + \left|\tan^{-1} \frac{y_1(t)}{x_1(t)}\right|$;
Target in the fourth quadrant, $\mathcal{Y} = -\left|\tan^{-1} \frac{y_1(t)}{x_1(t)}\right|$.
Operational equation of three-dimensional model:

$$\begin{bmatrix} x_1(k+1) \\ y_1(k+1) \\ z_1(k+1) \\ x_2(k+1) \\ y_2(k+1) \\ z_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ y_1(k) \\ z_1(k) \\ x_2(k) \\ y_2(k) \\ z_2(k) \end{bmatrix} + W(k) \tag{5}$$

$$Z(k) = \begin{bmatrix} \mathcal{P}(k) \\ \mathcal{Y}(k) \\ \sqrt{(x_1(k) - x_0)^2 + (y_1(k) - y_0)^2 + z_1(k) - z_0} + V(k) \end{bmatrix} \tag{6}$$

$\mathcal{P}$ in (6) is angle of pitch, $\mathcal{P} = \tan^{-1} \frac{z_1(t)}{\sqrt{(x_1(t))^2 + (y_1(t))^2}}$.

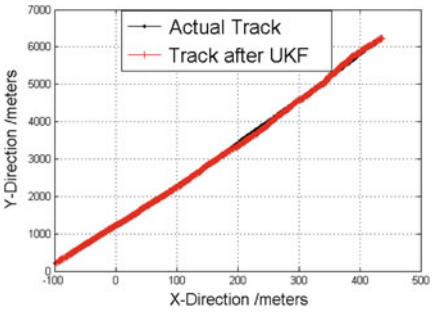
Obviously, both vector Z in (6) and (4) are made of nonlinear elements. We substitute these models into unscented transform (1), (2), then into linear Kalman filter Model.

4 Simulation Results and Analysis

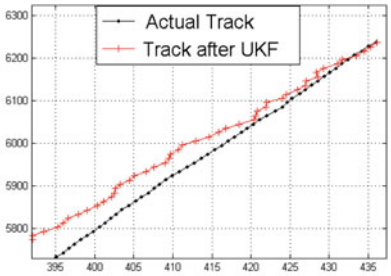
Some of the parameters: sampling period: 1 s; sampling time: 10 min; system noise in two-dimensional model: 10^{-2} m/s^2; system noise in three-dimensional model: 10^{-1} m/s^2.

Simulation results of two-dimensional model are shown in Fig. 4. We draw a conclusion from Fig. 4 that the two-dimensional model with UKF keeps an excellent performance and reaches the goal of tracking the target effectively. With the distance between moving target and observation station growing farther and farther, the tracking deviation keeps within 11 m and shows the trend of oscillatory convergence. It shows that the two-dimensional target tracking system has an excellent ability of anti-interference and achieves goal of fully observable.

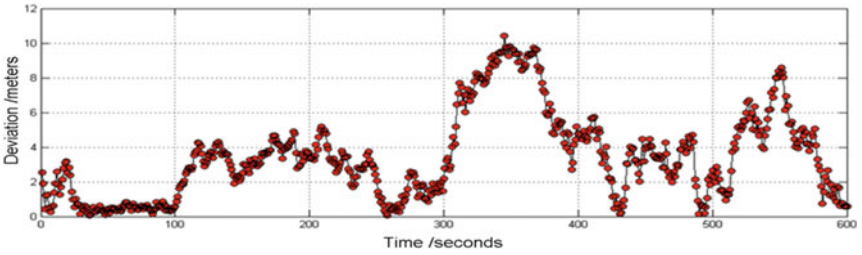
Simulation results of three-dimensional model are shown in Fig. 5. We could conclude from Fig. 5 that the three-dimensional model with UKF reaches the goal of



(a) Trajectory of target

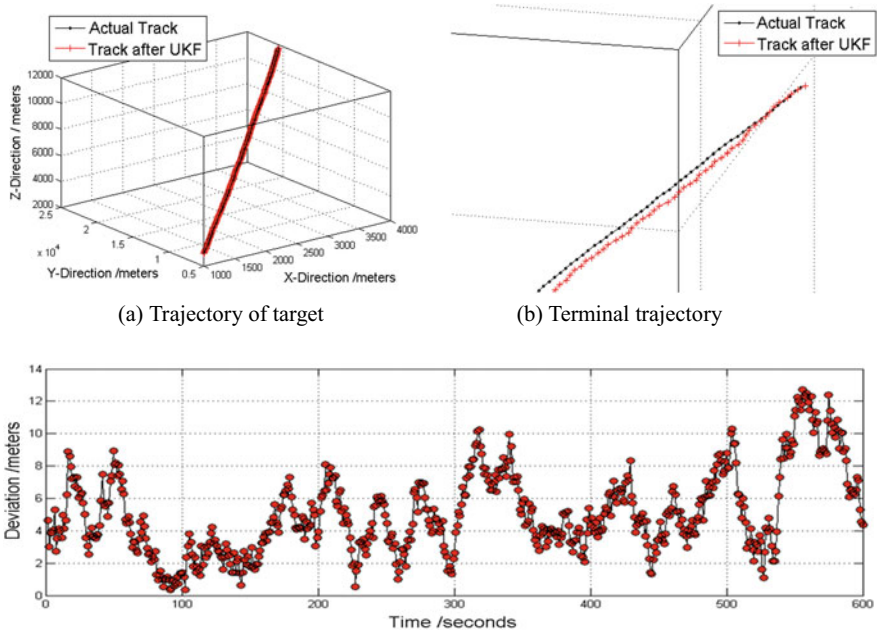


(b) Terminal trajectory



(c) Tracking deviation

Fig. 4 Simulation results of two-dimensional model



(a) Trajectory of target

(b) Terminal trajectory

(c) Tracking deviation

Fig. 5 Simulation results of three-dimensional model

tracking the target effectively. With the distance between moving target and observation station growing farther and farther, the tracking deviation keeps within 14 m and shows the trend of oscillatory convergence. All above shows that the three-dimensional target tracking system has an excellent ability of anti-interference and satisfies the requirement of the tracking target.

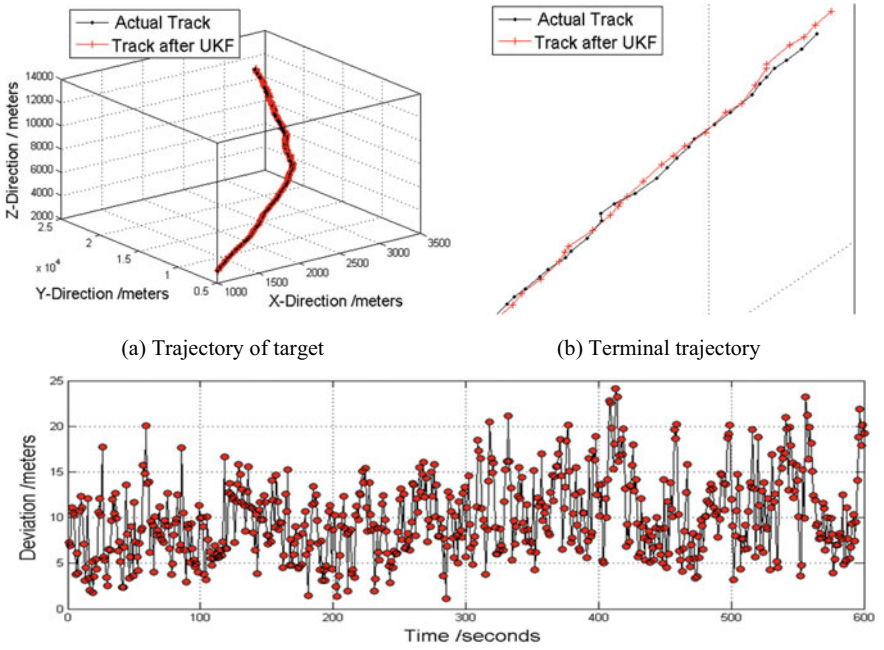
For a further discussion of verification for the filter's ability dealing with interference and nonlinear system, we increased the system noise $W(k)$ from 10^{-1} to 10 m/s^2 in the three-dimensional model.

Simulation results of the three-dimensional model with noise are shown in Fig. 6.

As is plotted in Fig. 6, the trajectory of the observed target becomes rather crooked under the influence of the system's noise. However, the result of target tracking seems quite desirable with deviation of no more than 25 m during 10-min observing period. And the deviation presents a trend of oscillatory convergence.

5 Conclusions

Unscented Kalman filter was developed from navigation and target tracking to deal with problems of the nonlinear system. The UKF is based on unscented transform



(a) Trajectory of target

(b) Terminal trajectory

(c) Tracking deviation

Fig. 6 Simulation results of three-dimensional model with noise

approximating nonlinear distribution by sampling method, which differs totally from EKF who linearizes the nonlinear system through neglecting or approximating the higher-order terms. The UKF based on symmetric set unscented transformation is applied to target tracking in condition of single observation station, and from analyses of simulation results, it has been proved to have an excellent performance in anti-interference to work out nonlinear system.

Compared with linear Kalman filter, the unscented Kalman filter has significant advantages. When using linear Kalman filter, the higher-order term is eliminated and only one term is reserved, which makes the error larger. The unscented Kalman filter is designed for nonlinear systems. The state equation or observation equation of the system is nonlinear, which does not eliminate the high term so that the filtering accuracy is improved.

Acknowledgements The authors acknowledge the financial support from Natural Science Foundation of Liaoning Province under Grant 2015010028-301.

References

1. Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. **53**, 343–367 (2003)
2. Azam, S., Ghisi, A., Mariani, S.: Parallelized sigma-point Kalman filtering for structural dynamics. Comput. Struct. **92–93**, 193–205 (2012)
3. Huang, X., Wang, Y.: The Principle and Application of Kalman Filter, pp. 1–5. Electronic Industry Press (2015)
4. Mariani, S., Ghisi, A.: Unscented Kalman filtering for nonlinear structural dynamics. Nonlinear Dyn. **49**, 131–150 (2007)
5. Alkaya, A.: Unscented Kalman filter performance for closed-loop nonlinear state estimation: a simulation case study. Electr. Eng. **96**(4), 299–308 (2014)

Research on Handover Strategy Based on Greedy Algorithm in Vehicle Edge Computing

Hongzhuan Zhao, Lina Zheng, Wenyong Li, Dan Zhou, and Wenhao Li

Abstract Due to the mobility of the vehicle, the communication link between the vehicle and the edge server changes dynamically in the vehicle edge computing, leading to the increase of task completion time, transmission energy consumption, and execution cost. In order to solve the problem, this paper studies the computational handover strategy in the vehicle edge computing environment. Considering the three indicators of completion time, transmission energy consumption, and execution cost, two handover strategies based on greedy weighted of simple additivity (SAW) algorithm and greedy expectation–maximization (EM) algorithm are proposed. The experimental results show that the two handoff strategies proposed in this paper can shorten the task completion time, reduce the energy consumption of vehicle transmission, and reduce the task execution cost.

1 Introduction

The Internet of vehicles [1] is a typical application of the Internet of things (IoT) to achieve the function of collecting vehicle information and performing task unloading. Vehicles are limited in resources due to limitations in their own batteries and computing power in the Internet of vehicles. The traditional solution is to migrate tasks to other resources with richer resources, such as the central cloud [2]. However, as more and more computer applications begin to be applied to the vehicle environment, moving the task closer to the edge of the vehicle becomes a better solution. This way of migrating tasks to a more resource-rich cloud or edge execution is called computation migration, and the mode of offloading vehicle tasks to the edge of the

H. Zhao · L. Zheng (✉) · W. Li · D. Zhou · W. Li
School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China
e-mail: lina211523@163.com

H. Zhao
Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_7

network is called vehicle edge computing [3]. In the vehicle edge computing environment, by deploying edge servers at the edge of the network, the edge layer provides computing, storage, and information processing for vehicles and other services.

Regarding the research on vehicle edge calculation, there have been many domestic and foreign achievements in recent years, mainly focusing on the seamless migration of services [4], resource scheduling [5], and so on. Among them, Gosain et al. [6] developed an edge computing vehicle networking platform. Kumar et al. [7] proposed a smart grid data management based on DTN network for vehicle edge computing. Feng et al. [8] proposed a decentralized automatic vehicle edge computing framework. These programs have their own limitations how to use idle resources in a dynamic vehicle environment, and how to build an efficient and common framework which still remains to be solved.

Considering that the movement of the vehicle will cause the communication link between the vehicle and the edge to change in a dynamic environment, how to find the task is best when the three main indicators of ensuring the task completion time, transmission energy consumption, and execution cost can be reduced. The available resources are the focus of this paper. This paper proposes two kinds of computational switching strategies based on greedy SAW and greedy EM. The corresponding quality of service (QoS) evaluation model is established according to the selected strategy to reduce the task completion time, transmission energy consumption, and execution cost.

The remaining part of this paper is organized as follows. In Sect. 2, we introduce handover strategy of multi-indicator; in Sect. 3, we recommend experiment simulation; in Sect. 4, we summarize experiment results and analysis. Finally, we conclude our work in Sect. 5.

2 Handover Strategy of Multi-indicator

2.1 Objective Functions and Constraints

The handover strategy in vehicle edge computing designed in this paper is actually a multi-index decision problem, which can be solved as a multi-index decision problem. In this study, the process of multi-index decision making is generally divided into three steps.

Step 1: Quantization of selected indicators and data preprocessing. In fact, the existence of missing values and outliers is not considered in the simulation experiment.

Step 2: Getting the index value and its corresponding weight. Three indexes were selected as the index parameters for decision making, which includes task completion time, task transmission energy consumption, and task execution cost. Next, we use SAW method and EM method to determine the index weight value.

Step 3: The optimal scheme is selected by the selected decision criteria. For Step 3, this experiment established two utility functions for the three indexes and made the decision of task switching under the condition that the utility function value was minimized.

We establish the basic model of switching strategy and the corresponding utility function is as follows:

$$f(x) = \sum_{n} w_n x_n, \quad \text{for } n = 1, 2, 3 \tag{1}$$

where the matrix x_1, x_2, x_3 of the index value at each time represents time, energy consumption, and cost; the corresponding weight w_1, w_2, w_3 of each index is determined according to the selected switching algorithm. The constraint condition is set as $w_1 + w_2 + w_3 = 1$.

The objective function gets as follows:

$$\min[f(x)] \tag{2}$$

2.2 Weight Determination

In this section, we propose the weight determination methods SAW and EM, both of which are based on the greedy algorithm. The greedy algorithm [9] is a simple and efficient method to find the optimal solution. In this article, after the task is uploaded, the greedy object is established: user experience QoS. The expression gets as follows:

$$\text{QoS} = q_1 \times T + q_2 \times P + q_3 \times C \tag{3}$$

where T represents the remaining completion time of the task, P represents the remaining transmission energy consumption of the task, and C represents the remaining execution cost of the task. The corresponding coefficients q_1, q_2, q_3 are determined by different switching strategies.

The task makes a migration decision when the task's QoS on the current server is not as good as the QoS for other servers, namely:

$$\text{QoS(now)} > \text{QoS(later)} + \text{QoS(migration)} \tag{4}$$

where QoS(migration) is the cost of task migration.

SAW The SAW method is a common method to solve the multi-indicator decision making. When solving the objective problem of reality, the result is more in line with the will of the decisionmakers.

Using the SAW method, the weight of each indicator is determined by the individual preference of the decisionmaker. $W = [w_1, w_2, \ldots, w_n]$ is the weight vector, where K is the total number of selected indicators, with $w_i \in [0, 1]$ and $w_1 + w_2 + \cdots + w_n = 1$. We set $A = \left[a_{ij}\right]_{m\times n}$ as the decision matrix.

Therefore, the expression of the SAW method is written as:

$$\text{SAWI} = w_1 x_{i1} + w_2 x_{i2} + \cdots + w_n x_{in} \tag{5}$$

It is assumed that the vehicle edge computing environment is composed of one acer station and three small base stations deployed with edge servers. The user terminal moves in this environment, and the following decision matrix is established for the candidate computing resources:

$$A = \begin{bmatrix} (A_1, C_1)\ (A_1, C_2)\ (A_1, C_3) \\ (A_2, C_1)\ (A_2, C_2)\ (A_2, C_3) \\ (A_3, C_1)\ (A_3, C_2)\ (A_3, C_3) \end{bmatrix} \tag{6}$$

where A_1, A_2, A_3 represent the small base station with edge server deployed. C_1, C_2, C_3 denote the completion time, transmission energy consumption, and execution cost of tasks on each resource, respectively.

The resource with the smallest decision coefficient is selected as the target resource.

EM The EM-based handover strategy uses the entropy value to determine the weight of each resource indicator. Compared with the SAW-based switching strategy, the EM-based switching strategy has certain objectivity. The core is to measure the utility value of information with the disorder of information. The lower the degree of disorder of information, the greater the utility value of the information.

The algorithm can be divided into the following three steps:

Step 1: Calculating index weight. EM is defined as follows:

$$H(X_i) = E[I(X_i)] = E\left[\log_2 \frac{1}{P(X_i)}\right] = -\sum_{i=1}^{m} P(X_i)\log_2 P(X_i) \tag{7}$$

where P represents the remaining transmission energy consumption of the task.

The entropy value of resource index is set as:

$$E(i) = -k \sum_{j=1}^{m} f_{ij} \log_2 f_{ij} \tag{8}$$

where k is an arbitrary constant, in this paper, it will be taken as $k = 1$. Besides, f_{ij} is the standardized resource index, where i and j denote the index number and the serial number of candidate resources, respectively.

The degree of difference is:

$$G(i) = 1 - E(i) = 1 + \sum_{i=1}^{m} f_{ij} \log_2 f_{ij} \tag{9}$$

Step 2: Calculating the composite index. The decision coefficient is defined as:

$$\mathrm{EM}_i = \sum_{j=1}^{n} b_{ij} \times \omega_j \tag{10}$$

Step 3: The handover decision is made according to the scheme with the smallest decision coefficient.

3 Experiment Simulation

In order to verify the performance and feasibility of the computational switching strategy based on greedy algorithm, this paper selects MATLAB 2016b tool to verify the experiment.

In this experiment, the problem of resource unavailability caused by insufficient storage space of edge server and cloud or other reasons is not considered, and the interference of other signals on communication is not considered. The task can be interrupted. Furthermore, the vehicle motion model uses the most common random waypoint model in the Internet of vehicles. The communication modes of $V2V$ in the vehicular ad hoc networks are not considered, and the five communication modes of $V2BS$, $V2AP$, $V2AP2BS$, $V2BS2RUS$, and $RUS2RUS$ are considered. The experimental parameter settings [10] are shown in Table 1.

4 Experimental Results and Analysis

We select the task average completion time $\tilde{T}$, the task average execution energy consumption $\tilde{E}$, and the task average execution cost $\tilde{C}$ to evaluate the method and the comparison method involved. The calculation formulas of these three performance indicators are given below:

$$\tilde{T} = T/N, \tilde{E} = E/N, \tilde{C} = C/N \tag{11}$$

where N is the number of tasks, T is the total time of task completion, E is the total energy consumption of task completion, and C is the total task execution cost.

Based on MATLAB 2016b experimental platform, it can be concluded that the vehicle can perform local task, unload from cloud server, and behave as the nearest

Table 1 Experimental parameters

Name	Number	Name	Number
Simulation interval (s)	1	Experiment number	1000
Acer station number	1	Center cloud number	1
Edge server number	3	Vehicle speed (m/s)	30–60
Small base station number	16	Task traffic (M)	[40 60]
Task computation (MIPS)	$1.8 \times [4\ 6] \times 10^6$	Small base station radius (m)	[100 150]
Central cloud computing capability (MIPS)	$4219 \times 5.5 \times 7$	Small base station to small base station bandwidth (Mbps)	5
Vehicle computing capability (MIPS)	$4219 \times 5.5 \times 2.5$	Vehicle to acer station bandwidth (Mbps)	1.28
Edge server computing capability (MIPS)	$4219 \times 5.5 \times [4\ 5\ 6]$	Vehicle to small base station bandwidth (Mbps)	10

Note 4219 is IPone 8 running points, and 5.5 is the ratio of running points to MIPS

edge server, and the edge server is with the most computing power, SAW switching strategy (SWA), and EM switching strategy (EM). The results are as follows:

From the average completion time of the task, the schemes that do not perform edge calculation in the six schemes have the most time. It can be seen that the entropy method and the simple weighting method have the best effect, and the execution time is 117 and 120 s, respectively. From the perspective of the average transmission energy consumption of the tasks, the energy consumption of the six schemes is 11.66, 12.80, 8.23, 12.38, 11.00, and 10.77 J. Among them, the highest execution cost is not edge calculation, and the lowest is the entropy method. From the perspective of the average execution cost of the task, the cost of the six schemes is 145.75, 109.31, 108.31, 93.69, 73.75, and 67.80 s, respectively. Among them, the highest execution cost is not edge calculation, and the lowest is the entropy method.

5 Conclusions and Future Work

In this paper, the multi-indicator switching in vehicle edge computing is studied. Two switching algorithms suitable for vehicle edge environment are proposed. Based on the MATLAB 2016b experimental platform, the performance average time of completion, the average transmission energy of the task, and the average execution cost of the task in the three kinds of task unloading scenarios of the vehicle's own execution, cloud execution and edge execution are obtained. The experimental results show that the proposed handover algorithm is reasonable and effective, which can shorten the task completion time, reduce the task execution cost, and reduce the task transmission energy consumption. In the future work, we can make more in-depth research from the weight determination of the handover decision and also refine the

mathematical model of multiple indicators to achieve the practical significance of the experimental results.

Acknowledgements This paper is funded by the National Natural Science Foundation of China (61803113), the Natural Science Foundation of China (61572146, U1501252), the Guangxi Natural Science Foundation (2018GXNSFAA138036), the Guangxi Natural Science Foundation (2016GXNSFAA380056), the Cooperative Education Program of Ministry of Education (201702117006), the Guangxi Key Laboratory of Trusted Software (kx201713), the Guangxi Science and Technology Base and Talent Special Project (GuikeAD18281015), the China Postdoctoral Science Foundation (2019M653313), and the scientific research fund project of Guilin University of Electronic Technology (UF17005Y).

References

1. Gerla, M., Kleinrock, L.: Vehicular networks and the future of the mobile internet. Comput. Netw. **55**(2), 457–469 (2011)
2. Zhang, J., Gu, Z., Zheng, C.: Survey of research progress on cloud computing. Appl. Res. Comput. **27**(02), 429–433 (2010)
3. Shi, W., Cao, J., Zhang, Q., Li, Y., Xu, L.: Edge computing: vision and challenges. IEEE Internet Things J. **3**(5), 637–646 (2016)
4. Cao, X.: Research on edge cloud computing architecture and data migration method. M.Eng. dissertation, Xidian University, Xi'an, China (2013)
5. Fan, C., Wang, S., Sun, Q., Zou, H., Yang, F.C.: IoV vertical handover research based on Bayesian decision. J. Commun. **34**(7), 34–41 (2013)
6. Gosain, A., Berman, M., Brinn, M., Mitchell, T., Li, C., Wang, Y.H., Jin, H., Hua, J., Zhang, H.W.: Enabling campus edge computing using GENI racks and mobile resources. In: Symposium on Edge Computing, Washington, DC, USA, pp. 41–50 (2016)
7. Kumar, N., Zeadally, S., Rodrigues, J.J.P.C.: Vehicular delay-tolerant networks for smart grid data management using mobile edge computing. IEEE Commun. Mag. **54**(10), 60–66 (2016)
8. Feng, J., Liu, Z., Wu, C., Ji, Y.: AVE: autonomous vehicular edge computing framework with ACO-based scheduling. IEEE Trans. Veh. Technol. **66**(12), 10660–10675 (2017)
9. DeVore, R., Temlyakov, V.: Some remarks on greedy algorithms. Adv. Comput. Math. **5**(1), 173–187 (1996)
10. Li, B., Pei, Y., Wu, H., Shen, B.: Heuristics to allocate high-performance cloudlets for computation offloading in mobile ad hoc clouds. J. Supercomput. **71**(8), 3009–3036 (2015)

Safety Analysis Method Considering Cascading Trips Based on Particle Swarm Optimization

Hui Qiong Deng, Xing-Ying Lin, Peng-Peng Wu, Qin-Bin Li, and Chao-Gang Li

Abstract View of preventing cascading trips, this paper mainly studies the security level model of how to effectively prevent cascading failures in the system is studied. First of all, from the point of view of relay protection, we consider the case of backup protection for current protection, by judging the electrical distance between the operating parameters of the branch and the relay protection fixed value after the initial fault of the system, it can be judged whether the branch will undergo cascading trips or not. Then, considering the critical state of cascading trips in branch grids, a safety index is proposed to measure the shortest distance between the critical state of power grids and the current operation state, and the specific calculation method is given. Finally, an optimization model is given, which aims to find the nearest state on the boundary from the current operation state. Aiming at this model, this paper presents a solution method based on PSO algorithm, which is verified by simulation through an example analysis.

1 Introduction

The development of society is inseparable from electricity, and the construction of power grid is becoming more and more powerful. The cascading trips of power grids can cause large blackouts, and the economic losses and social impacts caused by large blackouts are also very serious, so the research on cascading trips is becoming more and more urgent.

In order to prevent large blackouts caused by cascading failures, experts and scholars at home and abroad have done a lot of research on cascading failures from various perspectives. Literature [1] studies the network structure of power grid under

H. Q. Deng (✉) · X.-Y. Lin · P.-P. Wu · Q.-B. Li · C.-G. Li
School of Information Science and Engineering, Fujian University of Technology, Fujian 350118, China
e-mail: Denghuiqiong72@126.com; 2584880919@qq.com

Fujian Provincial University Engineering Research Center of Smart Grid Simulation Analysis and Integrated Control, Fujian 350118, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_8

small world network and its role in cascading failure propagation. Literature [2] uses decision tree to analyze cascading failure and obtains the set of key lines. Literature [3] analyzes the fragile lines in cascading failure of power system through Q-learning. Literature [4] studies the construction method of critical line model which affects the vulnerability of power grid from the perspective of vulnerability of cascading failures. Literatures [5–7] use vulnerability indicators such as average path length, degree, and median to identify critical lines.

Therefore, it is very vital to study the related issues of cascading failures in power grids. From the analysis of grid cascading failures, cascading failures are closely related to cascading trips. The early stages of cascading failures generally show cascading trips. In order to effectively prevent the occurrence of cascading failures, it is necessary to take measures at an early stage.

In this paper, according to the cascading trips phenomenon of power grid and the performance behavior of the branch cascading trips, the critical state of cascading trips in power grids is shown by mathematical expressions, and the grid operation status is expressed by the nodal injection power, the safety index of the grid trip is considered. Based on the relations between nodal injection power and grid safety level, a model is presented to indicate the safety level of the power system without cascading failure under the impact of initial failure, and the model is solved by particle swarm optimization. Taking the IEEE39 node system as an example, the rationality of the method is verified.

2 Description of Cascading Trip

2.1 Mathematical Expressions of Cascading Failures

It is known from the process of cascading trips that the cascading trips occur after the initial fault branch is disconnected, and the branch in the remaining branch has a load crossing the boundary to trigger the backup protection action. Therefore, when the initial fault had occurred, the difference between the measurement value of backup protection and the set value of the remaining lines other than the initial fault line can be compared to determine whether the backup protection is activated, thereby determining whether a cascading trips accident occurs.

The branch that causes the initial fault of the system is the L_i branch, and the remaining branch except the branch L_i is the branch L_j, and whether the branch L_j will have a cascading trip through Formula (1):

$$I_{j\cdot\mathrm{dist}} = \left|I_{j\cdot s}\right| - \left|I_j\right| \tag{1}$$

In Formula (1), I_j is the current value of the branch L_j after the power flow redistribution, the current setting value of backup protection on branch L_j is $I_{j\cdot s}$, and $I_{j\cdot\mathrm{dist}}$ is the electrical distance between $I_{j\cdot s}$ and I_j. When $I_{j\cdot\mathrm{dist}} > 0$, the branch L_j will

not trip, but when $I_{j\cdot\mathrm{dist}} < 0$, the branch L_j will trip. When $I_{j\cdot\mathrm{dist}} = 0$, the branch L_j is just in the critical state of cascading trip.

This paper only considers the case of the backup line protection in the case of current-type protection.

2.2 Critical State Considering Cascading Trip

Considering the critical state of cascading trips in power grids with all branches.

Assuming that after the initial fault branch is removed, the number of remaining branches in the power grid is l. According to Formula (1), all remaining branches are taken into account, and the matrix shown in Formula (2) can be given.

$$\boldsymbol{K} = \mathrm{diag}(I_{1\cdot\mathrm{dist}}, \ldots, I_{j\cdot\mathrm{dist}}, \ldots, I_{l\cdot\mathrm{dist}}) \tag{2}$$

$I_{j\cdot\mathrm{dist}}$ in Formula (2) represents the same meaning as in Formula (1), and K is a diagonal matrix.

According to the concrete manifestation of the interlocking tripping of the power grid, for the power grid operating under a certain state, after the original fault branch is removed, the power grid is transferred through the power flow, and if the current of all the remaining branches of the power grid meets Formula (3), the power grid is rigid.

$$\begin{cases} |\boldsymbol{K}| = 0 \\ I_{j\cdot\mathrm{dist}} \geq 0, \quad j = 1, 2, \ldots, l \end{cases} \tag{3}$$

Further, from the knowledge of power system analysis, it could been seen that the power flow formula of power system can be expressed by Formula (4) after the initial failure of branch L_i.

$$\boldsymbol{Y}\dot{\boldsymbol{U}} = \left(\boldsymbol{S}/\dot{\boldsymbol{U}}\right)^* \tag{4}$$

In Formula (4), after L_i outage of branch, $\mathbf{Y}$ represents the node admittance matrix of the grid, $\dot{\boldsymbol{U}}$ is the node voltage vector. For a given power grid and initial fault branch L_i, $\mathbf{Y}$ is a matrix whose elements are fixed values. Therefore, $\dot{\boldsymbol{U}}$ in Formula (4) principally depends on the nodal injection power $\mathbf{S}$. If the change of nodal injection power before and after L_i outage is neglected, $\mathbf{S}$ will remain unchanged before and after L_i outage. Then, $\dot{\boldsymbol{U}}$ is primary determined by the nodal injection power $\mathbf{S}$, and $\mathbf{S}$ is the injected power of grid nodes before branch outage.

From the power system analysis, it could be seen that in branch L_i, I_j in Formula (1) mainly depends on the node voltage after the initial failure. Therefore, combined with Formula (4), I_j mainly depends on the nodal injection power. It is further known that for a given initial fault, the critical operation state of the power grid after the

initial fault removal is mainly determined by the operation state before the initial fault removal, and nodal injection power of the power grid can determine the operation state of power grid before initial fault.

Thus, the safety of power grid can be described by injecting power into nodes of power grid.

3 Power Network Security Index Considering Cascading Trips

According to the above analysis, when considering cascading trips, when the power grid is running in a certain state, the cascading trips will not occur after the impact of initial faults. For larger distances between the operation boundary of the power grid and the operation, the safety of power grid is higher. Therefore, the security level of power grid can be measured in this paper: the shortest distance of injected power of the node between the current operation state of the power grid and critical state.

The distance between $\boldsymbol{S}'$ and $\mathbf{S}$ can be expressed by Formula (5), if $\boldsymbol{S}'$ is the power vector injected into the nodes under the current operation state s_1 and $\mathbf{S}$ is the power vector injected into the nodes under a certain boundary operation state s_2.

$$D(\boldsymbol{S}) = \left\| \boldsymbol{S}' - \boldsymbol{S} \right\| \tag{5}$$

In Formula (5), $\boldsymbol{D}(\boldsymbol{S})$ represents the norm of the difference between $\boldsymbol{S}'$ and $\mathbf{S}$.

Furthermore, the shortest distance between the nodal injected power in the current operation state and the nodal injected power in the critical state can be shown as follows:

$$F = \min D(\boldsymbol{S}) \tag{6}$$

From the above analysis, it could be seen that when the power grid operates under state s_2, if the power grid is not subject to the impact of initial fault, the greater the value of F, makes the power grid safer. Therefore, considering the case of cascading trips, F can be used as an indicator to characterize the power grid security. Before the initial fault occurs, the power flow constraint of steady-state operation should be satisfied when the grid operates under state s_2, that is, the power flow formula in steady-state operation can be expressed as shown in Formula (7).

$$\boldsymbol{h}^0(\boldsymbol{x}) = 0 \tag{7}$$

In Formula (7), before the initial fault occurs, $\boldsymbol{h}^0$ is the mapping relation of power flow, x is the state variable of power flow.

Prior to the initial failure, when the power grid operates under state s_2, it should also satisfy some inequality constraints, which can be expressed in the form of

Formula (8).

$$\begin{cases} Q_{Gi\cdot\min} \le Q_{Gi} \le Q_{Gi\cdot\max}, \ i = 1, \ldots, N_1 \\ P_{Gi\cdot\min} \le P_{Gi} \le P_{Gi\cdot\max}, \ i = 1, \ldots, N_1 \\ U_{k\cdot\min} \le U_k \le U_{k\cdot\max}, \ k = 1, 2, \ldots, N_2 \\ P_m \le P_{m\cdot\max}, \ m = 1, 2, \ldots, l \end{cases} \tag{8}$$

Among them, P_{Gi} and Q_{Gi} are, respectively, the active and reactive power of the I generator in the system; $P_{Gi\cdot\min}$ and $P_{Gi\cdot\max}$ are the lower and upper limits of the active power of the second generator; $Q_{Gi\cdot\min}$ and $Q_{Gi\cdot\max}$ are the lower and upper limits of the reactive power of the I generator, respectively; N_1 is the total number of generators in the grid; P_m is the active power of branch L_m transmission, $P_{m\cdot\max}$ is the active power limit of branch L_m transmission; U_k is the voltage of node K, $U_{k\cdot\min}$ is the lower voltage limit allowed by node K, $U_{k\cdot\max}$ is the upper voltage limit allowed by node K, and N_2 is the total number of grid nodes. Formula (8) is abbreviated and can be expressed as Formula (9).

$$\boldsymbol{g}^0(\boldsymbol{x}) \le 0 \tag{9}$$

When the initial fault occurs, the power flow constraint can also be satisfied under the action of the injected power of the node corresponding to the state s_2, which can be expressed as the formula shown in Formula (10).

$$\boldsymbol{h}^k(\boldsymbol{x}) = 0 \tag{10}$$

Among them, $\boldsymbol{h}^k$ is the corresponding power flow mapping relationship when the power flow redistribution occurs after the initial fault occurs in s_2 state.

The model for calculating $\boldsymbol{D}(\boldsymbol{S})$ shown in Formula (11) can be given by Formulas (3) and (6)–(10).

$$\begin{cases} \min D(\boldsymbol{S}) = \left\| \boldsymbol{S}' - \boldsymbol{S} \right\| \\ \text{s.t.} \ \ \boldsymbol{h}^0(\boldsymbol{x}) = 0 \\ \quad \boldsymbol{h}^k(\boldsymbol{x}) = 0 \\ \quad |\boldsymbol{K}| = 0 \\ \quad \boldsymbol{g}^0(\boldsymbol{x}) \le 0 \\ \quad I_{j\cdot\text{dist}} \ge 0, \quad j = 1, 2, \ldots, l \end{cases} \tag{11}$$

Formula (11) is an optimization model. The variable to be optimized is the node injection power corresponding to the operation state of the interlocking tripping boundary. Its goal is to find the nearest state on the boundary to the current operation state s_1 and the node injection power in this state.

4 Calculation Algorithm of Security Index

For the optimization model given in Formula (11), this paper will use particle swarm optimization to solve the problem. The main reason for using this algorithm is that it is easy to program, simple and fast, and is suitable for solving complex optimization problems which are difficult for classical optimization algorithms, and was used in many fields. For the basic particle swarm optimization, this paper carries out iterative calculation in the form of Formula (12).

$$\begin{cases} \boldsymbol{v}_i^{k+1} = w\boldsymbol{v}_i^k + c_1 r_1\left(\boldsymbol{p}_{\text{best}\cdot i} - \boldsymbol{x}_i^k\right) + c_2 r_2\left(\boldsymbol{g}_{\text{best}} - \boldsymbol{x}_i^k\right) \\ \boldsymbol{x}_i^{k+1} = \boldsymbol{x}_i^k + \boldsymbol{v}_i^k \end{cases} \tag{12}$$

In Formula (12), $\boldsymbol{x}_i^k$ and $\boldsymbol{v}_i^k$ are the particle i the kth of the iteration position and the iteration speed, respectively. $\boldsymbol{p}_{\text{best}\cdot i}$ is the optimal solution experienced by particle i itself; $\mathbf{g}_{\text{best}}$ is the optimal solution experienced by the whole particle swarm; w is the inertia coefficient, which decreases linearly from 0.9 to 0.1; c_1 and c_2 are the acceleration constants, whose values are generally 2; r_1 and r_2 are random numbers with uniform distribution in [0, 1] interval.

In order to cooperate with particle swarm optimization, the model of Formula (11) is simplified appropriately:

1. For power flow constraints in Formula (11), this paper will turn them into power flow calculation. In the process of iteration, if the power flow constraints are not satisfied, new particles will be generated.
2. For $|\boldsymbol{K}| = 0$ in Formula (11), it is recorded as $f(\boldsymbol{x}) = 0$ in this paper, and for $I_{j\cdot\text{dist}} \geq 0$ in Formula (11), it is recorded as $\boldsymbol{g}^1(\boldsymbol{x}) \leq 0$ in this paper. In this way, the constraints to be handled in Formula (11) can be expressed in Formula (13).

$$\begin{cases} f(\boldsymbol{x}) = 0 \\ g^0(\boldsymbol{x}) \leq 0 \\ g^1(\boldsymbol{x}) \leq 0 \end{cases} \tag{13}$$

Based on the above analysis, the penalty function shown in Formula (14) can be given by combining Formulas (11) and (13).

$$D' = D + \sum k\frac{1}{\alpha_k}\left[\min\left(0, -g^0(x)\right)\right]^2 + \sum k\frac{1}{\beta_k}\left[\min\left(0, -g^1(x)\right)\right]^2 + \frac{1}{\gamma}[f(x)]^2 \tag{14}$$

In Formula (14), α_k, β_k and γ are penalty factors.

A fitness function is established by Formula (14):

$$F = \frac{1}{D'} \tag{15}$$

Once the fitness F reaches the maximum value, which is the minimum value of the penalty function, we can find the set of values closest to the critical point.

According to Formulas (14)–(15), this paper combines particle swarm optimization algorithm to give the specific flow of the algorithm:

First, we need to check whether the initial state of the system meets the requirements: the particle x in the particle swarm algorithm represents the node injection power of the system. We finally find a set of node injection power by constantly adjusting the value of x to maximize the fitness value. As long as we avoid this set of values, then the next step is safe to run; v is the direction in which the particles run. Then, a new set of ground-state data is generated according to the original ground state data, and it is checked whether the set of ground-state data satisfies the requirements: if the requirements are met, the next step is performed, and if the requirements are not met, a set of data is regenerated. Then, calculate the fitness of the generated particles: firstly, randomly initialize the velocity of the particles, calculate the value of the particle fitness, obtain $\mathbf{P}_{\text{best}}$ and $\mathbf{g}_{\text{best}}$, and then update the velocity and position of the particle according to Formula (12) to obtain the local optimum for each time, solution, and global optimal value. Finally, after reaching the maximum number of iterations, the iteration is stopped and the optimal value is obtained. What we are looking for is the set of particles corresponding to the optimal value.

5 Example

This paper takes IEEE39 node system as an example to further verify the above methods. The wiring of the IEEE39 node system can be seen in Fig. 1. The system data of node information can be referred to [8].

The analysis of the examples in this paper was carried out in the MATLAB environment. The safety margin is the current state of the system and the shortest distance corresponding to the operating boundary of the system. The parameters of the particle swarm algorithm are set as follows: $c_1 = c_2 = 2$; w is taken as a linear law from 0.9 to 0.1. In Formula (14), α_k is taken as 0.005, β_k is taken as 0.005, and γ is taken as 0.01. The initial fault branch is set to be the branch between node 18 and node 19, namely branch $L_{18,19}$. The initial population size for this study is 50, and the number of iterations for setting the loop is 100. Assume that the backup protection of each branch configuration corresponding to the system in Fig. 1 is set to 3 kA, and the corresponding grid capacity is set to 100 MVA.

By simulating the example, the fitness value of Fig. 2 is obtained. In Fig. 2, the abscissa represents the number of particle iterations, and the ordinate represents the value of the optimal fitness for each iteration. It could be seen from Fig. 2 that after 100 cycles, the best fitness value is 0.2866 by comparison. From the results of the program operation, the optimal particle is the 30th particle. Therefore, in the example of this paper, when the initial fault occurs in the system: the branch between nodes 18 and 19 fails; when the outside world applies a disturbance: the artificial injection system powers the node; as long as we do not choose, the node corresponding to

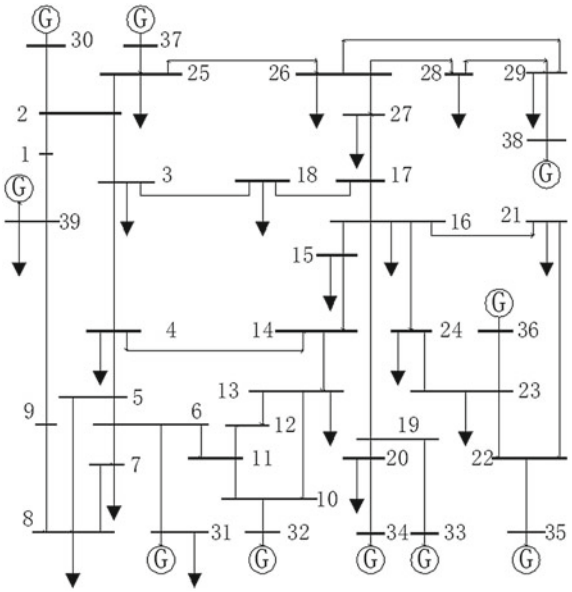


Fig. 1 Diagram of IEEE39-bus system

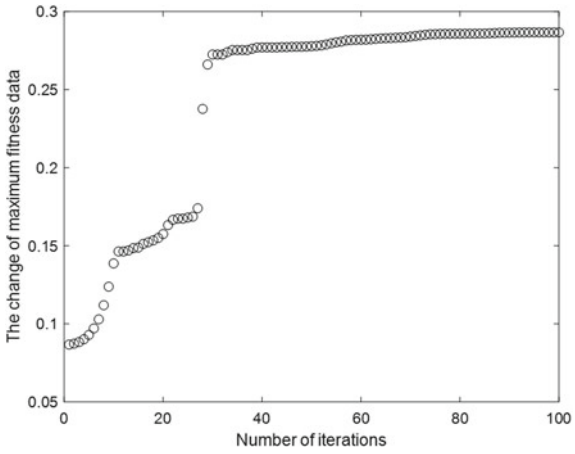


Fig. 2 Change of maximum fitness

the 30th particle injects power, and the system greatly reduces the probability of cascading failure.

From the results obtained from 100 cycles, we can get the best value in Table 1, which is the shortest distance $\boldsymbol{D}(\boldsymbol{S})$ of the nodal injection power between current

Table 1 Running results of PSO

Name	$\boldsymbol{D}(\boldsymbol{S})$	$\boldsymbol{D}'(\boldsymbol{S})$	F
Value	1.746	3.489	0.286

operating state and critical state. The maximum values of the model $\boldsymbol{D}'(\boldsymbol{S})$ and the fitness F of the safety index combined with the equality constraint and the inequality constraint in the normal operation state are shown in Table 1.

6 Conclusion

The results obtained from the above example analysis show that in the case of considering the backup protection of the line as current-type protection, we have successfully designed a model for considering the grid safety index of the cascading trips, combined with the particle swarm algorithm. Using MATLAB, the simulation was carried out in the environment to prove the practicability of the model. When an initial fault has occurred in the power grid, the power flow is reallocated in removing the faulty branch circumstances. The value of the injected power of the group of nodes most likely to have cascading trips can be found according to the method of this paper, as long as we avoid this. The choice of group value can effectively avoid the occurrence of cascading failures.

Acknowledgements This research was financially supported by Doctoral Research Foundation of Fujian University of Technology under the grant GY-Z13104, and Scientific Research and Development Foundation of Fujian University of Technology under the grant GY-Z17149.

References

1. Pan, Y.-F., Li, H.-Q., et al.: Cascading fault prediction of small world network. J. Power Syst. Autom. **25**(2), 148–153 (2013)
2. Zhong, Y.-X., Zhang, X.-M., Mei, S.-W.: Blackout risk analysis and control of power system integrated with wind farm. In: Control and Decision Conference 2016, Yinchuan, China, pp. 876–882 (2016)
3. Zhang, Z.-M., Huang, S.-W., Chen, Y.: Key branches identification for cascading failure based on q-learning algorithm. In: IEEE International Conference on Power System Technology 2016, Wollongong, Australia (2016)
4. Li, Y., Su, H.-L.: Research on key lines affecting the vulnerability of power system under N-K fault. J. Power Autom. Equip. **35**(3), 60–67 (2015)
5. Rosascasals, M., Valverde, S.: Topological vulnerability of the European power grid under errors and attacks. Int. J. Bifurc. Chaos **17**(7), 2465–2475 (2007)
6. Ding, M., Han, P.-P.: Vulnerability assessment algorithm for large power grid based on small world topology model. J. Power Syst. Autom. **30**(8), 7–10 (2006)
7. Dwivedi, A., Yu, X., Sokolowski, P.: Identifying vulnerable lines in a power network using complex network theory. In: IEEE International Symposium on Industrial Electronics, 2009. ISIE 2009. IEEE (2009)
8. Wang, H.: Prevention Strategy of Improving the Safety Level of Power Grid for Cascading Trips. Fujian University of Technology, China (2018)

A Method for Restraining Cascading Trip Based on Node Power

Hui-Qiong Deng, Peng-Peng Wu, Xing-Ying Lin, Qin-Bin Li, Chao-Gang Li, and Zhi-Han Shi

Abstract With the continuous development of the smart wireless communications, the importance of electric energy becomes increasingly evident. Intricately interconnected power grids will also suffer from the threats of blackouts caused by grid cascading failure while bringing convenience to the masses. The grid cascading trip is closely related to cascading failure. Generally, the cascading failure shows a cascading trip at early stage. Hence, the avoidance of occurrence of cascading trip at an early stage may restrain the occurrence of complex cascading failure, thereby avoiding the occurrence of the blackout. In this paper, the occurrence principle of the cascading trip is firstly analyzed from the perspective of relay protection. The output of the generator is taken as the optimization variable, and the optimization model for restraining the grid trip is presented. Finally, the example of IEEE14 node system is analyzed in this paper in combination with the particle swarm algorithm, which verifies the rationality of the proposed method.

1 Introduction

When the Second Industrial Revolution took place, the human society entered the electrical age, and all fields were increasingly dependent on electrical energy. The masses would feel inconvenient even if one-day power failure occurred in real life. In response to the development of social economy, China's power grid is developing toward the directions of large power grid, large unit, ultra-high voltage, automation and information technology. The connections between power grids are becoming

H.-Q. Deng (✉) · P.-P. Wu · X.-Y. Lin · Q.-B. Li · C.-G. Li
School of Information Science and Engineering, Fujian University of Technology, Fujian 350118, China
e-mail: 574760430@qq.com

Fujian Provincial University Engineering, Research Center of Smart Grid Simulation Analysis and Integrated Control, Fujian 350118, China

Z.-H. Shi
Xiamen University of Technology, Fujian 361024, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_9

increasingly closer. The problem following is that the possibility and risk of grid failure are constantly increasing, and the occurrence scope of failures is becoming increasingly wider.

The problems like the grid cascading trip and even cascading failure have been the focuses and concerns of many researchers thanks to the emphasis on the importance of avoidance of the occurrence of blackout. A type of interconnected grid cascading failure control measures based on heterogeneous cellular automata is proposed in [1]. The dispatchers may reasonably select interregional support coefficients according to grid operations in regions and effectively reduce the risk of power failures in regions. Normalized network power flow indicator acts as the predictive indicator of the intermediate link of fault chain. The fault chain model is established online, and it may be used to quantitatively analyze the cascading failure. The fault chain model is applied to the risk assessment on the system and to provide preventive control guidance for the cascading failure in [2]. The fault chain model is applied to the vulnerability assessment on the system branch, and the branch portion of the system which is relatively fragile and has a greater impact on the stability of the power system is analyzed through the evaluation results in [3]. The weather factors are integrated into the fault chain model, providing an important basis for the operation mode developer to identify dangerous links in the bad weather in [4]. The shortcoming involving overload protection of the line or the action of deterministic analog protection control is only considered in the fault chain model in [5]. The protection and control action evaluation indexes for key components are raised, and the action probability of key component protection control is combined to determine the follow-up event of fault chain of the cascading failure.

From the point of view of restraining chain trips, it is clear that earlier the development of blocking cascading failures further makes for the safety of the grid, alternatively, the failure of the second level is to be avoided to the utmost extent for the grid.

As far as the cascading trip event of the power grid is concerned, the action behavior of relay protection is an influential factor to large extent, so the protection action must be considered while exploring this problem. On the basis of considering the current-type backup protection, an optimization model is proposed in combination with the action of relay protection to restrain the grid from tripping to the utmost extent to avoid the cascading trip and improve the safety of the grid. The IEEE14 system is taken as an example to combine the particle swarm algorithm, and the method raised in this paper is verified.

2 Safety Consideration Index of Grid While Cascading Trip Occurs

When the initial failure occurs in the power grid and the initial failure is removed, the existence of occurrence of cascading trip mainly depends on if the electrical quantity

of branches in the power grid enters the action area of backup protection when the power flow the grid is redistributed. If the backup protection of branches of the power grid is current-type protection, the situation, that if the grid trips after the power flow is transferred, mainly depends on the specific configuration of current and backup protection of branches.

It is supposed that Branch L_{ij} between Node i and Node j in the power grid suffers from initial failure at certain moment. When Branch L_{ij} is cut off and the power flow of the power grid is redistributed, the possibility if any branch L_d will suffer from the cascading trip may be measured by Eq. (1) for remaining portions of the power grid:

$$I_{d\cdot\mathrm{dist}} = |I_{d\cdot\mathrm{set}}| - |I_d| \tag{1}$$

In Eq. (1), I_d represents the current operating parameter of Residual Branch L_d, $I_{d\cdot\mathrm{set}}$ denotes the current-type backup protection setting value of Residual Branch L_d, and $I_{d\cdot\mathrm{dist}}$ denotes quantity measuring the electrical distance between $I_{d\cdot\mathrm{set}}$ and I_d.

According to actual performance of the abovementioned formula and grid cascading trip, Branch L_d will suffer from cascading trip when $I_{d\cdot\mathrm{dist}} > 0$ is satisfied, and Branch L_d is to be cut off by the backup protection when $I_{d\cdot\mathrm{dist}} \leq 0$ is satisfied that is L_d will suffer from cascading trip.

There are a total of l branches of the grid except for initial failure branch from [6]; Index shown in Eq. (2) can be further obtained through Eq. (1).

$$M = \min(I_{d\cdot\mathrm{dist}}) \tag{2}$$

In terms of Eq. (2) and cascading trip phenomenon, one of the residual branches is at the boundary of cascading trip at least in addition to initial failure branch when $M = 0$ is satisfied, and the grid is in the boundary status of cascading trip.

From the above analysis, we can see that M may be used as an indicator to measure the safety of the grid. If the generator output of the power grid may be optimized and adjusted in the operation of the power grid, M is to be guaranteed and to be greater than zero in Eq. (2), the value of M is to be great to the utmost extent, and the cascading trip of the power grid is to be effectively prevented and written in mathematical form to have the objective function shown in Eq. (3).

$$F = \max(M) \tag{3}$$

3 Optimization Model for Restraining Grid Cascading Trip

For the objective function shown in Eq. (3), the variable to be optimized is the output of the generator unit, and a complete mathematical model may be formed by adding the objective function to the electrical constraint relation necessary for the grid.

The electrical constraint relations considered mainly involve the equality constraint and inequality constraint. Where, the equation constraint mainly involves the power flow constraint of the grid before the occurrence of initial failure and that after the occurrence of initial failure, and the former may be abbreviated into the form of Eq. (4).

$$\boldsymbol{h}^0(\boldsymbol{x}) = 0 \tag{4}$$

In Eq. (4), $\boldsymbol{h}^0$ represents the mapping relation corresponding to the grid power flow before the occurrence of initial failure, and $\boldsymbol{x}$ denotes the grid state variable when the power flow is solved. The power flow constraint of the grid after the initial failure may be abbreviated into the form of Eq. (5).

$$\boldsymbol{h}^1(\boldsymbol{x}) = 0 \tag{5}$$

In Eq. (5), $\boldsymbol{h}^1$ denotes the mapping relation corresponding to the power flow of the grid after the occurrence of initial failure.

In this optimization problem, the inequality constraint relation to be considered is mainly shown in Eq. (6).

$$\begin{cases} P_{Gi\cdot\min} \le P_{Gi} \le P_{Gi\cdot\max}, \quad i = 1, \ldots, N1 \\ Q_{Gi\cdot\min} \le Q_{Gi} \le Q_{Gi\cdot\max}, \quad i = 1, \ldots, N1 \\ P_m \le P_{m\cdot\max}, \quad m = 1, 2, \ldots, l \\ U_{k\cdot\min} \le U_k \le U_{k\cdot\max}, \quad k = 1, 2, \ldots, N2 \end{cases} \tag{6}$$

In Eq. (6), P_{Gi} and Q_{Gi} are, respectively, the active and reactive power of the generator i in the system; $P_{Gi\cdot\min}$ and $P_{Gi\cdot\max}$ are, respectively, the lower limit and upper limit of active output of the generator i in the system; $Q_{Gi\cdot\min}$ and $Q_{Gi\cdot\max}$ are, respectively, the lower and upper limits of reactive power of the generator i in the system; $N1$ is the total number of generators in the grid; P_m is the active power transmitted by the branch L_m; $P_{m\cdot\max}$ is the active power limit transmitted by the branch L_m; U_k is the voltage of node k; $U_{k\cdot\min}$ is the lower-voltage limit allowed by node k; $U_{k\cdot\max}$ is the upper-voltage limit allowed by node k; $N2$ is the total number of grid nodes. Write Eq. (6) in abbreviated form, which can be expressed in Eq. (7).

$$\mathbf{g}_0(\boldsymbol{x}) \le 0 \tag{7}$$

Additionally, according to the above analysis, M in Eq. (2) is also required to satisfy a certain constraint relation, namely $M > 0$, as m is also the function of state variable $\boldsymbol{x}$, so m is denoted as $f(\boldsymbol{x})$ here, the corresponding inequality constraint relation is denoted as the form shown in Eq. (8).

$$f(\boldsymbol{x}) \ge 0 \tag{8}$$

It is possible to form optimized and adjusted output of the power grid and to improve the goal of the grid in terms of the safety of cascading trip by combining the above formulas thoroughly from Formulas (3) to (8). The complete mode is shown in Eq. (9).

$$\begin{cases} F = \max(M) \\ s.t.\ h^0(x) = 0 \\ \quad h^1(x) = 0 \\ \quad g^0(x) \le 0 \\ \quad f(x) \ge 0 \end{cases} \tag{9}$$

4 Solution Thoughts for Optimization Model

For the optimization model given by Eq. (9), this paper will use particle swarm optimization to solve it, and the reason for adopting this algorithm is mainly to consider that this algorithm is easy to program and implement. For the basic particle swarm algorithm, this paper performs iterative calculation according to Eq. (10).

$$\begin{cases} v_i^{k+1} = w v_i^k + c_1 r_1 \left(P_{\text{best}\cdot i} - x_i^k\right) + c_2 r_2 \left(g_{\text{best}} - x_i^k\right) \\ x_i^{k+1} = x_i^k + v_i^k \end{cases} \tag{10}$$

In Eq. (10), x_i^k is the k-th iteration position of particle i; v_i^k is the iteration speed of particle i at the k-th time; The condition that needs to be met is $v_{\min} \le v_i^k \le v_{\max}$; $P_{\text{best}\cdot i}$ is the optimal solution experienced by particle i itself; g_{best} is the optimal solution experienced by the entire particle swarm; w is the inertia coefficient; c_1 and c_2 are the acceleration constant; r_1 and r_2 are uniformly distributed random number in the interval [0, 1].

For the optimization problem shown in Eq. (9), this paper firstly simplifies the model:

(a) For the power flow constraint in Eq. (9), in this paper, it will be processed by the power flow calculation. In the iterative calculation process, if the power flow constraint is not satisfied, a new particle will be generated.

(b) For the optimization model shown in Eq. (9), in order to facilitate processing with particle swarm optimization, this paper changes the objective function of Eq. (9) to the penalty function form shown in (11).

$$F' = F + \sum_1^k \frac{1}{\alpha_k} \left[\min(0, -g^0(x))\right]^2 + \frac{1}{\gamma} [f(x)]^2 \tag{11}$$

In Eq. (11), α_k and γ are punishment factors, and k is the number of inequalities contained in Eq. (6).

Using the penalty function given by Eq. (11), this paper combines the particle swarm algorithm to give the following specific algorithm flow.

Step 1 Initialize the particle swarm to have the initial particle satisfy the feasible solution of the problem and assign the initial velocity of the particle;
Step 2 Solve the fitness of the particle according to the fitness function of the model;
Step 3 Compare the fitness of current particle with that at optimal position in the individual history and update the historical optimal position of the individual if the current fitness is higher;
Step 4 Compare the fitness of current particle at optimal position in global history and update the globally historical optimal position if current fitness is higher;
Step 5 Calculate the next position and velocity of the particle according to the iterative formula of the particle swarm;
Step 6 Stop the iteration when the maximum number of iterations is hit.

5 Example

Using IEEE14 system, and example is analyzed based on the previous algorithm (Fig. 1):

It is not to select cutting off certain generator unit in the initial failure due to the influence of output distribution of studied unit on the safety level indicator M. Node 1 is set as the balance node, and the output nodes of the adjustable unit are Node 2, Node 3, Node 6 and Node 8.

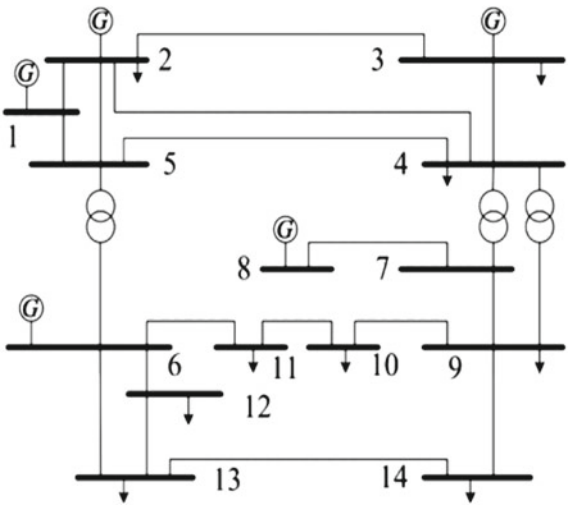


Fig. 1 Diagram of the example system

Table 1 Generator unit output distribution plan before failure

M	Node 2 output	Node 3 output	Node 6 output	Node 8 output
0.557	40 MW	10 MW	60 MW	70 MW

Set the initial failure to be the second branch. Line 2 is removed by the relay protection device when it suffers from the failure, and generator unit has not changed timely when the failure is to occur, and then the safety index *M* of the system is as follows.

From Table 1, when the initial fault is in the second branch, the safety index *M* of the system is only 0.557, which is less than 0.6. At this time, the whole system is in an unsafe state. If the system is subjected to external shocks again, the maximum probability of cascading trips will occur.

In order to further prevent the occurrence of cascading trips and increase the value of *M*, the system should timely adjust the output of generator units to ensure that their safety indicators are as large as possible.

After adjusting the output of the generator set, the index *M* of the system is shown in Fig. 2.

Meanwhile, the optimal outputs of the generator units are as follows (Table 2).

By adjusting the output of the generator, we find that the safety index has been raised from 0.557 to 0.86, which greatly reduces the risk of cascading trips, thus achieving the purpose of restraining cascading trips.

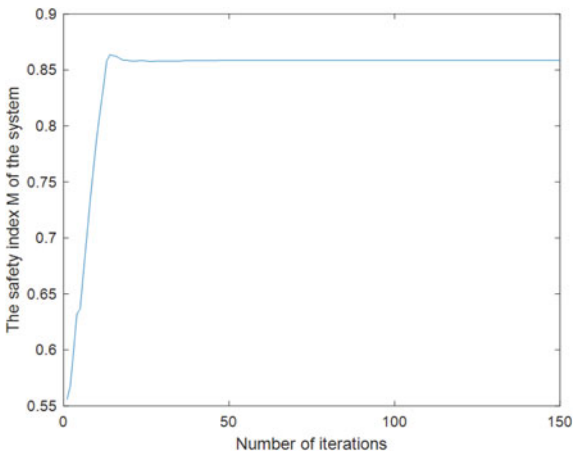


Fig. 2 Safety index after adjusting the output of generator units

Table 2 Generator unit output distribution plan after failure

M	Node 2 output	Node 3 output	Node 6 output	Node 8 output
0.86	55 MW	118 MW	8 MW	17 MW

6 Conclusion

According to the studies in this paper, it is possible to find out that the optimization on the output of the unit may effectively improve the safety level of the system to prevent the occurrence of the cascading trip fault and thus effectively restraining the occurrence of the cascading trip fault. The safety level evaluation indicator characterizes the electrical distance between the system and the occurrence of cascading trip fault. The greater the safety level indicator is, the less the system is likely to trip and the safer it is. It is found that the change in the output of the unit can effectively improve the safety level of the system through calculations, thus achieving the purpose of restraining the cascading trip of the power grid.

The conclusions drawn in this paper and the method of restraining the cascading trip proposed may have certain references for further research and actual operation of the grid.

Acknowledgements This research was financially supported by Doctoral Research Foundation of Fujian University of Technology under the grant GY-Z13104, and Scientific Research and Development Foundation of Fujian University of Technology under the grant GY-Z17149.

References

1. Yu, Q., Wang, Q., Cao, N.: A control measure for interconnection power grid cascading failure based on heterogeneous cellular automata. Power Syst. Protecti. Control **17**(01), 1–15 (2019)
2. Luo, Y., Tu, G.Y., Liu, P., Su, D.: Online determination method of fault chains of power system cascading failure available for preventive control. High Voltage Eng. **35**(10), 2446–2451 (2009)
3. Luo, Y., Tu, G, Y., Liu, P.: Online transmission vulnerability assessment method based on the fault chain risk index. Proc. CSEE **30**(25):44–50 (2010)
4. Liu, W. Y., Yang, N., Zhang, J. L., Sun, J. H., Yao, F., Wang, J. M.: Complex grid failure propagating chain model in consideration of adverse weather. Proc. CSEE **32**(07), 53–59 + 191 (2012)
5. Xuan, X.H., Zhou, Y., Song, X.F., Wu, H.X., Xue, F., Deng, H., Hua, W.: A novel searching method of fault chains for power system cascading outages based on quantitative analysis of dynamic interaction between system and components. Power Syst. Protect. Control **46**(02), 101–109 (2018)
6. Deng, H.Q.: Research on the safety margin of power network considering cascading tripping. J. Fujian Univ. Technol. **14**(03), 255–261 (2016)

Improved Window Decoding of Spatially Coupled LDPC Codes

Yamei Zhang, Lin Zhou, Shuying Zhang, Chen Chen, Yuqing Fu, and Yucheng He

Abstract In this paper, we analyze the performance of window decoding (WD) scheme for spatially coupled low-density parity-check (SC-LDPC) codes, and an improved WD scheme is proposed. Compared with the belief propagation (BP) decoding algorithm, window decoding has the advantages of low decoding latency, low decoding complexity, and small memory. But, the performance of WD is not as good as that of BP. It is found that the error bit number **err** of the target symbol does not decrease monotonously with the iteration progresses. This means that the target symbol likelihoods generated when decoding is terminated may not be the best choice to make a decision. To optimize the performance of WD, an improved WD is proposed. In this scheme, the achievable minimum of **err** is monitored, and the related likelihoods is stored to estimate the target symbol at the end of decoding. Simulation results show that the improved WD outperforms the conventional WD.

1 Introduction

Spatially coupled low-density parity-check (SC-LDPC) codes are first proposed by Felstrom and Zigangirov [1]. It is a kind of convolutional LDPC code [2]. Kudekar et al. have proved that SC-LDPC code has threshold saturation characteristic [3], that is, the belief propagation (BP) thresholds of SC-LDPC code can reach the maximum a posteriori (MAP) thresholds of corresponding regular LDPC code.

Using BP decoding long codes may cause high latency and need large memory. In [4], a window decoding (WD) method is proposed to solve this problem. The parity-check matrix $\boldsymbol{H}_{\text{sc}}$ of SC-LDPC code shows the structure of a nonzero diagonal band.

Y. Zhang · L. Zhou (✉) · S. Zhang · C. Chen · Y. Fu · Y. He
Xiamen Key Laboratory of Mobile Multimedia Communications, College of Information Science and Engineering, Huaqiao University, Xiamen, Fujian 361021, China
e-mail: linzhou@hqu.edu.cn

L. Zhou · Y. He
State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, Shaanxi, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_10

Hence, the BP decoding can be operated in a window. The window slides along the diagonal band of the parity matrix and the target symbols are estimated one by one with low decoding latency until the whole codeword is decoded. Limiting BP decoding to a window has the advantage of reducing decoding latency, complexity, and memory, but at the cost of decoding performance. Therefore, Shi Yuan Mo, Inayat, and Peng Kang propose to introduce supervision [5], effective decoding termination [6], and retain local reliable message [7], respectively, to improve the performance of WD. Then, the P-EXIT analysis is applied for SC-LDPC code with window decoding in [8] to reduce decoding complexity. The research of Shi Yuan Mo et al. shows that the average error probability of the target symbols they define does not monotonically decrease with the iteration.

Inspired by this phenomenon, we find that the number of error bits **err** of the target symbols does not monotonically decrease with the iteration progresses. This means that the log-likelihood ratio (LLR) generated when the window decoding terminates is not the best choice for making decisions. Therefore, we propose an improved window decoding (iWD) that monitors the minimum of **err** and store the related LLR. Then, the target symbols can be estimated according to the stored LLR at the end of the window decoding. Then, we analyze and simulate the performance of SC-LDPC code WD and iWD in the additive white Gaussian noise (AWGN) channel. The simulation results show that the performance of the iWD outperforms the conventional WD.

2 SC-LDPC Codes

An SC-LDPC code can be constructed by spatial coupling of protographs [9, 10]. Given a regular LDPC protograph, an SC-LDPC protograph can be obtained by coupling L replicas of a regular LDPC protograph through edge spreading, where L represents the coupling length. For example, a (J, K)-regular LDPC protograph is given in Fig. 1a, where J represents column weight and K represents row weight. Figure 1b is to make L copies of regular LDPC protograph. And they are indexed by t. Figure 1c shows the process of edge spreading. Defining w as the coupling width,

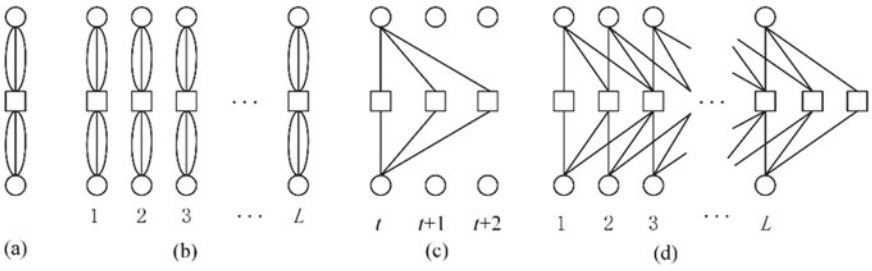


Fig. 1 **a** A (3, 6)-regular LDPC protograph, **b** L replicas of (3, 6)-regular LDPC protographs, **c** process of edge spreading with $w = 2$, and **d** protograph of the (3, 6) SC-LDPC codes with $w = 2$

then the edge of the variable nodes (VNs) at t is connected to the checking nodes (CNs) at $t, t+1, \ldots, t+w$. The protograph of the (3, 6) SC-LDPC codes is shown in Fig. 1d.

A base matrix $\boldsymbol{B} = [\boldsymbol{B}(i, j)]_{\mathrm{a}\times\mathrm{b}}$ can represent a regular LDPC protograph ($a = J/(\gcd(J, K)), b = K/(\gcd(J, K))$, $\gcd(J, K)$ denote the maximum common divisor of J and K). Figure 1a shows that the base matrix corresponding to the (3, 6)-regular LDPC protograph is $\boldsymbol{B} = \begin{bmatrix} 3 & 3 \end{bmatrix}$. The edge spreading divides the base matrix into component base matrices, and the component base matrices corresponding to Fig. 1c are $\boldsymbol{B}_i = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $i = \{0, 1, 2\}$.

Therefore, the base matrix of SC-LDPC code is

$$\boldsymbol{B}_{\mathrm{sc}} = \begin{bmatrix} \boldsymbol{B}_0 & & & \\ \boldsymbol{B}_1 & \boldsymbol{B}_0 & & \\ \vdots & \boldsymbol{B}_1 & \ddots & \\ \boldsymbol{B}_{\mathrm{w}} & \vdots & \ddots & \boldsymbol{B}_0 \\ & \boldsymbol{B}_{\mathrm{w}} & \ddots & \boldsymbol{B}_1 \\ & & \ddots & \vdots \\ & & & \boldsymbol{B}_{\mathrm{w}} \end{bmatrix}_{((L+w)a\times Lb)} \tag{1}$$

$\boldsymbol{H}_{\mathrm{sc}}$ can be obtained by further lifting $\boldsymbol{B}_{\mathrm{sc}}$ by the lifting factor M. Hence, the code rate of SC-LDPC code is

$$R_{\mathrm{sc}} = 1 - \frac{(L+w)Ma}{LMb} \tag{2}$$

3 Window Decoding and Improved WD

The parity matrix of SC-LDPC code presents a nonzero diagonal band structure. As shown in Fig. 2, the gray area is its nonzero diagonal band. Therefore, BP decoding can be constrained in a window, and the window size W needs to be satisfied with the condition $(w+1) \leq W \leq L$.

The general principle of WD is given below. After receiving partial codeword, the decoding window starts BP decoding through a partial Tanner graph of the codeword and then slides along the diagonal band to estimate the target symbols one by one and generate a lower decoding latency. The symbol set contained in the leftmost protograph in the decoding window is called the target symbol. Once the BER of the target symbol is 0 or the number of iterations reaches the maximum, the decoding window terminates and outputs the decoding result of the current target symbol. Then, the decoding window slides to the next position to decode the next target symbol,

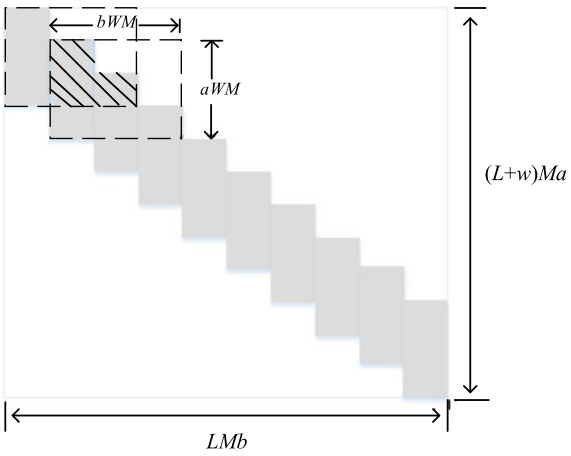


Fig. 2 WD of SC-LDPC codes, where the coupling width $w = 2$, the coupling length $L = 10$, and the window size $W = 3$

and the sliding process is shown in Fig. 2. Then, the above process is repeated until the entire codeword is decoded.

Assume that, the codeword is binary phase-shift keying (BPSK) modulated and sent to the AWGN channel. Let $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)$ and $r = (r_1, r_2, \ldots, r_N)$ represent the transmitted and accepted codeword, respectively, where $N = LMb$. The decoding process of WD is as follows.

Initialization: After receiving r_m from the channel in the mth window, the LLR of s_m can be initialized to

$$L_m = \ln\left(\frac{P(s_m = 0|r_m)}{P(s_m = 1|r_m)}\right) \tag{3}$$

Iterative Process: Only exchange and update the information of VNs and CNs in the current window. At lth iteration, the information transmitted by check node (CN) n to the connected variable node (VN) m is

$$R_{nm}^{l} = \ln \frac{1 + \prod_{k \in M_1(n)\backslash m} \tanh(Q_{nk}^{l-1}/2)}{1 - \prod_{k \in M_1(n)\backslash m} \tanh(Q_{nk}^{l-1}/2)} \tag{4}$$

Information passed by VN m to the connected CN n is

$$Q_{nm}^{l} = \begin{cases} L_m & l = 0 \\ L_m + \sum_{k \in M_2(m)\backslash n} R_{km}^{l} & l > 0 \end{cases} \tag{5}$$

$M_1(n)\backslash m$ represents the set of remaining VNs connected to CN n except the mth VN. Similarly, $M_2(m)\backslash n$ represents the set of remaining CNs connected to VN m except for the nth CN.

Decision And Window Sliding: After the current iteration, the LLR of VNs is calculated, and then, the hard decision is made based on the Q_{nm}^{l}. If $Q_{nm}^{l} \geq 0$, $s' = 0$, or $s' = 1$, otherwise. Let **err** denote the number of error bits of the target symbol. If the corresponding bit is decoded incorrectly, **err** will be increased by

$$\mathbf{err} = \mathbf{err} + 1 \tag{6}$$

If the BER of the target symbol is zero or the number of iterations reaches the maximum, the current window decoding will terminate and output the decoding result and then keep and propagate the information of the overlapped part (the shadow area in Fig. 2) from the current window to the next window and slide to the next position for decoding; otherwise, return to step 2 to continue the iteration.

Note that, **err** does not monotonically decrease with the iteration progresses. Therefore, improved WD is proposed. This scheme monitors the minimum of **err** and stores the related LLR. Let err* denote the minimum of error bits of the target symbol. err* will be initialized as err* $= 10^{10}$. As the iteration progresses, if **err** $<$ err*, err* will be updated by

$$\text{err}^* = \mathbf{err} \tag{7}$$

And the LLR of the target symbol will be stored. Once the BER of the target symbol is zero or the number of iterations reaches the maximum, the current decoding window will estimate the target symbol based on the stored LLR. Then, the current decoding window slides to the next position to decode the next target symbol.

4 Simulation Results and Analysis

This section simulates the (3, 6) SC-LDPC code, its corresponding base matrix $B = \begin{bmatrix} 3 & 3 \end{bmatrix}$, the coupling width $w = 2$, and the coupling length $L = 50$. The codeword is transmitted in the AWGN channel after BPSK modulation.

Figures 3 and 4 show the BER performance of the SC-LDPC code with the lifting factor $M = 10$ and $M = 50$, respectively. Hence, the codeword length is 1000 and 5000, respectively, and both code rates are 0.48. E_b/N_0 represents the signal-to-noise ratio. It is shown in each figure that iWD performance is better than WD performance at different window sizes. When the window size is 12 and the BER is 10^{-4}, the iWD of SC-LDPC with code length of 1000 has a gain of about 0.22 dB compared with that of WD. When the window size is 12 and the BER is 10^{-4}, the iWD of SC-LDPC with code length of 5000 has a gain of about 0.12 dB compared with that of WD. Theoretically, compared with WD, iWD only monitors the minimum error bit of the target symbol, and its decoding termination conditions have not changed. Therefore, the decoding complexity of iWD and WD is approximately the same.

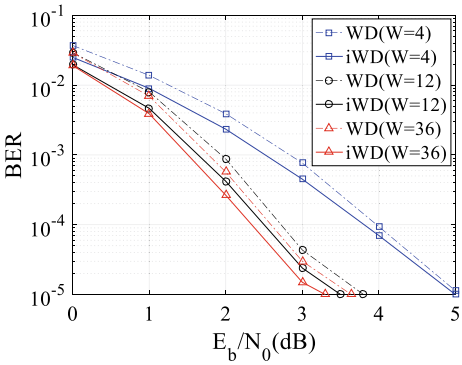


Fig. 3 BER curve of the SC-LDPC code with code length N= 1000

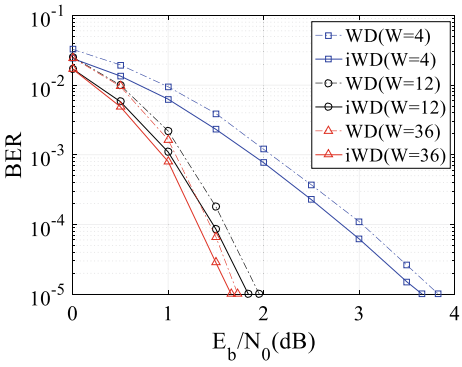


Fig. 4 BER curve of the SC-LDPC code with code length $N = 5000$

5 Conclusions

This paper introduces the window decoding of SC-LDPC codes and then proposes an improved scheme. The improved scheme monitors the minimum number of error bits of the target symbols and stores the related LLR. Then, the target symbols are estimated according to the stored LLR to make the more reliable decoding choices. The simulation results show that the improved scheme outperforms conventional WD.

Acknowledgements This paper was funded in part by the National Natural Science Foundation of China (61901182, 61302095), The Natural Science Foundation of Fujian Province of China (2018J01096, 2018J05105), Quanzhou City Science and Technology Program of China (2018C108R), and the Graduate Students Foundation of National Huaqiao University (18014082037).

References

1. Felstrom, A. J., Zigangirov, K. S.: Time-varying periodic convolutional codes with low-density parity-check matrix. In: Proceedings. 1998 IEEE International Symposium on Information Theory, Cambridge, MA, USA. IEEE Press (1998)
2. Kudekar, S., Richardson, T.J., Urbanke, R.L.: Threshold saturation via spatial coupling: why convolutional LDPC ensembles perform so well over the BEC. IEEE Trans. Inf. Theory **57**(2), 803–834 (2011)
3. Kudekar, S., Richardson, T., Urbanke, R.: Spatially coupled ensembles universally achieve capacity under belief propagation. IEEE Trans. Inf. Theory **59**(12), 7761–7813 (2013)
4. Iyengar, A.R., Papaleo, M., Siegel, P.H., Wolf, J.K., Vanelli-Coralli, A., Corazza, G.E.: Windowed decoding of protograph-based LDPC convolutional codes over erasure channels. IEEE Trans. Inf. Theory **58**(4), 2303–2320 (2012)
5. Mo, S., Chen, L.: Improved sliding window decoding of spatially coupled low-density parity-check codes. In: 2017 IEEE Information Theory Workshop (ITW), pp. 126–130. Kaohsiung, Taiwan. IEEE Press(2017)
6. Ali, I., Kim, J.H., Kim, S.H., Kwak, H.: Improving windowed decoding of SC LDPC codes by effective decoding termination, message reuse, and amplification. IEEE Access **6**, 9336–9346 (2018)
7. Peng, K., Yixuan, X., Lei, Y., Jinhong, Y.: Reliability-based windowed decoding for spatially coupled LDPC codes. IEEE Commun. Lett. **22**(7), 1322–1325 (2018)
8. Khittiwitchayakul, S., Phakphisut, W., Supnithi, P.: EXIT chart analysis for split-LDPC codes. In: 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), pp. 1–4. JeJu, Korea. IEEE Press (2019)
9. Thorpe, J.: Low-density parity-check (LDPC) codes constructed from protographs. Jet Propulsion Laboratory, INP Progress Report, pp. 42–154, Aug 2003
10. Divsalar, D., Dolinar, S., Jones, C., Andrews, K.: Capacity approaching protograph codes. IEEE J. Sel. Areas Commun. **27**(6), 876–888 (2009)

Application of the Several Common Algorithms for Corner Detection to Sonar Image Registration

Ruiying Luo and Haitao Guo

Abstract Sonar image registration plays a key role in intelligent perception of underwater target. Corner is an important feature of the image. Image registration algorithms based on corner detection have been widely used. Common corner detection algorithms include Harris algorithm, small univalue segment assimilating nucleus (SUSAN) algorithm, features from accelerated segment test (FAST) algorithm, scale-invariant feature transform (SIFT) algorithm, etc. This paper examines the application of these four corner detection algorithms in sonar image registration and gives the experimental results. At the same time, this paper compares these four corner detection algorithms from five aspects: the number of detected corners, corners logarithm participating in registration, corners logarithm of correct matching, matching rate, and matching time.

1 Introduction

With the development of science and technology, underwater target detection becomes more and more intelligent. Sonar is a powerful tool for underwater detection, it is widely used in underwater intelligent perception. Sonar image registration (or mosaic) is the basis of the intelligent perception of underwater target.

Single sonar image usually has small image size and limited coverage area. We often use image mosaic to seamlessly stitch multiple sonar images into a large image with high resolution and wide viewing angle [1]. Image registration is a key step in image mosaic. The quality of image registration directly determines the quality of image mosaic [2]. Image registration algorithms are mainly divided into three types: algorithms based on grayscale, algorithms based on transform domain, and

R. Luo
Inner Mongolia University, Hohhot 010021, China

H. Guo (✉)
Hainan Tropical Ocean University, Sanya 572022, China
e-mail: ghtpaper@126.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_11

algorithms based on feature [2, 3]. Sonar image registration usually adopt algorithms based on feature [3]. These algorithms mainly extract local stable feature in image. Common features include point feature, line feature, and regional feature [4]. Corner is easy to be detected and keeps stable in the case of image translation, rotation, scale transformation, and distortion, it has become an important tool for image feature expression [5]. The number and accuracy of detected corners are crucial to image registration. The registration algorithm based on corner detection includes two stages: corner detection and corner registration [6]. In recent years, the registration algorithm based on corner detection has attracted more and more attention. Common corner detection algorithms include: Harris algorithm [7], small univalue segment assimilating nucleus (SUSAN) algorithm [8], features from accelerated segment test (FAST) algorithm [9], and scale-invariant feature transform (SIFT) algorithm [10]. This paper examines the application of these algorithms in sonar image registration. At the same time, this paper analyzes and compares the four algorithms from subjective and objective perspectives.

2 Several Corner Detection Algorithms

In 1988, Harris and Stephens proposed a corner detection algorithm based on change in gray intensity, which is called Harris corner detection algorithm [7]. In 1997, Smith proposed SUSAN corner detection algorithm [8]. This algorithm is simple in theory and easy to implement. In 2010, Edward and Tom proposed FAST corner detection algorithm based on the SUSAN corner detection algorithm [9]. FAST algorithm is defined as: if a pixel is not similar to enough pixels in its surrounding neighborhood, then the pixel is called corner [11]. SIFT algorithm was proposed by Lowe in 1999 [12] and summarized in 2004 [10]. It is a corner detection algorithm based on scale space, which maintains invariable characteristic in image scaling, rotation, and even affine transformation. SIFT algorithm consists of four steps:scale-space extrema detection, keypoint localization, orientation assignment, keypoint descriptor, it is divided into two stages: corner detection and corner description, where scale-space extreme detection and keypoint localization belong to corner detection, the other steps belong to corner description.

3 Experimental Results and Analysis

3.1 Experimental Results and Analysis

This experiment aims to compare the performance of four corner detection algorithms in sonar image registration. In the following experiments, during corner detection stage, we respectively adopt the four corner detection algorithms mentioned above,

Table 1 Performance comparison of the four algorithms

Algorithms	n	a	b	r (%)	t (s)
Harris	39/34	3	3	100	5.85
SUSAN	250/214	27	10	37.0	13.58
FAST	42/27	4	3	75.0	6.12
SIFT	60/65	12	11	92.5	5.01

and the number of detected corners is represented by n. As given in Table 1, the number before the symbol "/" is the number of corners extracted from I_1 of Fig. 1b, and the number after the symbol "/" is the number of corners extracted from I_2 of Fig. 1b. In the corner registration stage, firstly, we uniformly adopt corner description method used in SIFT algorithm [9] to generate 128-dimensional corner descriptor for each corner, then the nearest neighbor sub-nearest neighbor algorithm [9] is used to confirm corners logarithm participating in registration, which is represented by a. Then, random sample consensus (RANSAC) algorithm [13] is used to calucate the geometric coordinate transformation model between the images to be registered, where corners logarithm conformed geometric coordinate transformation model are corners logarithm of correct matching, the others belong to corners logarithm of wrong matching. The number of corners logarithm of correct matching is represented by b, and the number of corners logarithm of wrong matching is represented by c, where a = b + c. Here, we use matching rate [14] to evaluate the performance of these algorithms, which is represented by r. The definition of matching rate is:

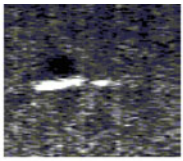
$$r = b/a \tag{1}$$

In general, the higher the matching rate, the better the image registration effect, and the better the comprehensive performance of the algorithm. At the same time, we record matching time, which is represented by t.

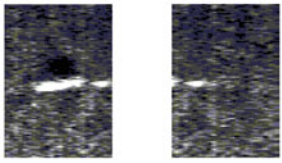
3.2 Parameter Setting

The following describes the parameters of the experiment during corner detection stage.

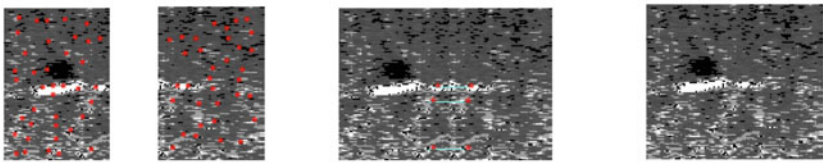
1. Harris algorithm: the σ of circular window is set to 1.6; the threshold of corner response function is set to $R = 20$; the constant k in the corner response function is set to 0.04 [7].
2. SUSAN algorithm: the brightness difference threshold is set to 45, n is the number of pixels in the USAN, and n is compared with a fixed threshold g, which is set to $3n_{\max}/4$, where $n_{\max}$ is the maximum value which n can take [8].



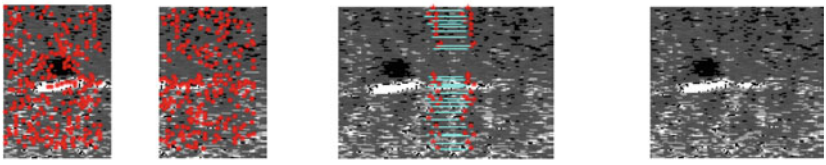
(a) Original sonar image



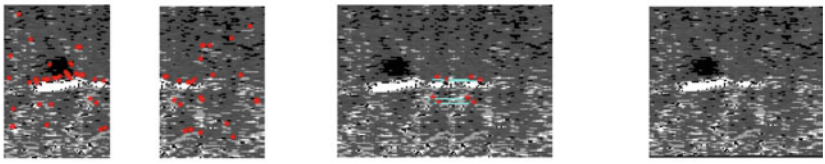
(b) Images to be registered (I_1 and I_2)



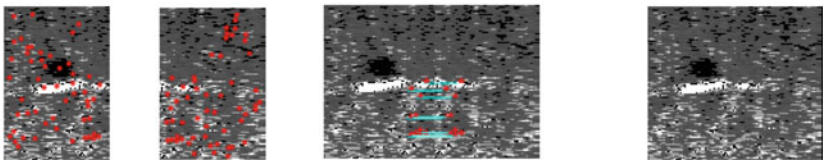
(c) Harris image registration



(d) SUSAN image registration



(e) FAST image registration



(f) SIFT image registration

Fig. 1 Four image registration algorithms

3. FAST algorithm: the threshold of t is set to 30, and the number of contiguous pixels n is set to $n = 12$ [9].
4. SIFT algorithm: the σ in variable-scale Gaussian $G(x, y, \sigma)$ is set to 1.6; the k in the difference of Gaussian function is set to 1.414; octave is set to 2; layer is set to 5; the threshold of $D(\hat{\mathbf{x}})$ is set to 0.03; the threshold of the ratio of principal curvatures is set to $r = 10$ [10].

In corner registration stage, we uniformly adopt these parameters. The orientation histogram has 36 bins covering the 360° range of orientations; when conduct orientation assignment, the σ of Gaussian-weighted circular window is set to 2.4; the highest peak is the main direction of the corner, any other local peak that is within 80% of the highest peak is auxiliary direction of the corner, then generate 128-dimention corner descriptior for every main direction and auxiliary direction. When caculate corner descriptor, the σ of Gaussian weighting function is set to 0.8; the ratio between the distance of the closest neighbor to that of the second-closest neighbor is set to 0.6.

3.3 Results and Analysis

Figure 1a is a sonar image. The two images in Fig. 1b are obtained through the Photoshop software screenshot of Fig. 1a. The two images are of same size and contain common area. In the process of collecting seabed images, sonar is inevitably affected by factors such as ocean waves [15]. There may be a certain degree of relative rotation between the sonar images to be registered, but the relative rotation angle is generally small, so this paper does not mention the case of relative rotation. The detected corners are marked with red dots, and the corners participating in registration are connected with blue lines. The registration results are as follows:

From the above experimental results, it can be seen that the four algorithms have completed image registration. Both SUSAN algorithm and SIFT algorithm detect a large number of corners. Although SUSAN algorithm detect more corners than SIFT algorithm, the number of corners logarithm of correct matching is obviously smaller than SIFT algorithm, so this illustrates that compared with SIFT algorithm, SUSAN algorithm has significantly more error detected corners. Both Harris algorithm and FAST algorithm detect relatively small number of corners, but the corner positioning accuracy of FAST algorithm is the highest of the four algorithms.

From objective perspective, in terms of matching rate, Harris algorithm and SIFT algorithm are significantly higher than the other two algorithms, but considering the logarithm of the corners participating in registration of Harris algorithm is obviously smaller than SIFT algorithm, the performance of SIFT algorithm is more stable than Harris algorithm. From the perspective of matching time, the time spent by SUSAN algorithm is obviously long, followed by the FAST algorithm, and the time spent by Harris algorithm and SIFT algorithm is not significantly different.

4 Conclusion

By setting up the experiments, this paper analyzes and compares the application of four corner detection algorithms in sonar image registration. Underwater intelligent services increasingly focus on perceived accuracy and speed. Sonar image registration is the basis of the underwater object intelligent perception. In this way, compared with Harris algorithm and SUSAN algorithm, SIFT algorithm and FAST algorithm are more suitable for sonar image registration. SIFT algorithm has high matching rate and is slightly faster than FAST algorithm; however, FAST algorithm has high positioning accuracy of corners. In practical application, we will choose the appropriate algorithm according to the specific situation.

Acknowledgements We would like to thank Professor Tian Tan for providing sonar images.

This work is supported by the Institute and Local Cooperation Foundation of Sanya in China (No. 2019YD08) and the National Science Foundation of China (No. 61661038).

References

1. Gao, J., Chen, L., Lan, X.: Image mosaic algorithm based on the FAST feature points extraction. Journal of Northwest University (Natural Science Edition) **46**(3), 351–357 (2016). (in Chinese)
2. Yang, C., Xu, X., Wang, J.: Research on importance of image mosaic technology. Computer Science **11**(43), 133–135 (2016). (in Chinese)
3. Li, P.: Research on feature matching method of side scan sonar image. Ph.D. Thesis, Harbin Engineering University, Harbin (2016).(in Chinese)
4. Lei, K., Liu, Y., Wang, Y., Xing, Z., Zhang, X.: A new corner points image registration algorithm based on TMS320C6416T. Electron. Device **29**(4), 1231–1234 (2006). (in Chinese)
5. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. Found. Trends Comput. Graph. Vis. **3**(3), 177–280 (2008)
6. Ge, Y., Yang, D., Lei, M.: Image registration based on well-distributed corners with sub-pixel localization. J. Electron. Inf. Technol. **32**(2), 427–432 (2010). (in Chinese)
7. Harris, C., Stephens, M. J.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151. IEEE (1988)
8. Smith, S.M., Brady, J.M.: SUSAN-A new approach to low-level image processing. Int. J. Comput. Vis. **23**(1), 45–78 (1997)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **2**(60), 91–110 (2004)
10. Rosten, E., Drummond, T.: Faster and better: a machine learning approach to corner detection. IEEE Trans. Pattern Anal. Mach. Intell. **32**(1), 105–119 (2010)
11. Liu, Y., Yu, H., Yang, W., Li, L.: SAR image registration using SAR-FAST corner detection. J. Electron. Inf. Technol. **39**(2), 430–436 (2017). (in Chinese)
12. Lowe, D. G.: Object recognition from local scale-invariant features. In: Proceedings of International Conference on Computer Vision, pp. 1150–1157. IEEE (1999)
13. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 726–740 (1981)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)

15. Li, H., Dong, Y., He, X., Xie, S., Luo, J.: A sonar image mosaicing algorithm based on improved SIFT for USV. In: Proceedings of 2014 IEEE International Conference on Mechatronics and Automation, pp. 1839–1843. IEEE (2014)

Performance Analysis of Interleaver for Polar Coding in Bit-Interleaved Coded Modulation System

Xu Shi, LIn Zhou, Yiduo Tang, Chen Chen, Yuqing Fu, and Yucheng He

Abstract A bit-interleaved polar-coded modulation (BIPCM) system was proposed in this paper. Considering the complexity of code construction, the polarization weight (PW) independent of the specific signal-to-noise ratio (SNR) was employed. For decoding, the successive cancellation (SC) decoding is to be known with a low complexity. Further, a unified description of BICM system was given, and the design of interleaver for polar code was analyzed in this paper. To evaluate the performance of interleaver for BIPCM system, the system with two existing interleaving algorithms were compared. Simulation is performed in the additive white Gaussian noise (AWGN) channel model with 16 quadrature amplitude modulation (QAM) signaling based on Gray mapping. It is shown that the system with random interleaver has a better performance of about 0.25–0.36 dB over the contrastive system. Moreover, the performance of equal interleaver scheme has a maximum gain of about 0.24 dB compared with the contrastive system at the low SNR.

1 Introduction

Polar code [1], proposed by Arikan in 2009, is a novel coding scheme that can theoretically realize the binary-input discrete memoryless channel (B-DMC) capacity under the SC decoding. With the characteristics of capacity-achieving performance, the polar code was successfully selected into the 5G standard and determined as the coding scheme of the 5G enhanced mobile broadband (eMBB) scenario control channel.

X. Shi · L. Zhou (✉) · Y. Tang · C. Chen · Y. Fu · Y. He
Xiamen Key Laboratory of Mobile Multimedia Communications, College of Information Science and Engineering, Huaqiao University, Xiamen 361021, Fujian, China
e-mail: linzhou@hqu.edu.cn

L. Zhou · Y. He
State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, Shaanxi, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_12

Generally, the polar-coded modulation scheme includes multilevel polar coded modulation [2] (MLPCM) and bit-interleaved polar-coded modulation (BIPCM) [3]. Although MLPCM system is an effective scheme to improve the bandwidth utilization, the system complexity is too high to be implemented. Therefore, a BIPCM system is discussed in this paper. For BIPCM, [4] proposed a channel mapping scheme based on equal-capacity partitioning by simply sorting the parallel subchannels and having a better performance compared with the existing schemes. Further, to evaluate the performance of BIPCM system, [5] adopted random interleaver to simulate the BER performance at different modulation orders. In this paper, polar code can be efficiently constructed with PW by calculating the polarization weight of bit channel. Then, we propose a common BIPCM scheme and simulate two different interleavers under that system. Simulation results show that the performance of this system with interleavers outperforms the contrastive system at low SNR.

2 The Basic Principle of Polar Code

2.1 Channel Polarization

The core of polar code is channel polarization, which consists of channel combining and channel splitting. When the code length N is large enough, there will be a polarization [6]. That is, these channels are either unreliable channels—noisy channel (channel capacity $I(W) = 0$), or reliable channels—noiseless channels ($I(W) = 1$). The encoding process is to transmit information bits by using a noiseless channel, and a noisy channel is employed to transmit frozen bits (generally 0).

2.2 Code Construction

In the process of 5G standardization, HUAWEI has proposed the coded construction method based on PW [7], which is based on the fixed row of generation matrix to estimate the reliability of the bit channel.

Let i denote the ith bit channel, whose binary sequence is $i \triangleq (a_{n-1}, a_{n-2}, \ldots a_0)$, $n = \log_2 N$. The polarization weight of a single bit channel i is defined as:

$$W_i = \sum_{j=0}^{n-1} a_j * 2^{\frac{j}{4}}, \quad i = 0, 1, \ldots, N-1 \tag{1}$$

By calculating (1), we can get the weight sequence $\boldsymbol{W}_0^{N-1}$ of bit channel. We write $\boldsymbol{Q}_0^{N-1}$ to denote the sequence obtained by sorting the $\boldsymbol{W}_0^{N-1}$, that is, $\boldsymbol{Q}_0^{N-1}$ represents the bit channel reliability rank of polar code, where $W_{Q_0} \leq W_{Q_1} \leq \cdots \leq W_{Q_{N-1}}$.

2.3 Polar Encoder

The polar code (N, K) is a linear block code with the code length N and K information bits. Polar coding is also expressed as:

$$\boldsymbol{x}_1^N = \boldsymbol{u}_1^N \boldsymbol{G}_N \tag{2}$$

$\boldsymbol{x}_1^N = (x_1, x_2, \ldots, x_N)$ and $\boldsymbol{u}_1^N = (u_1, u_2, \ldots, u_N)$ are the coded and source sequence, respectively. $\boldsymbol{G}_N$ is the generating matrix, which can be defined as:

$$\boldsymbol{G}_N = \boldsymbol{B}_N \boldsymbol{F}^{\otimes n} \tag{3}$$

where $\boldsymbol{B}_N$ is bit reversal operation and $\boldsymbol{F}^{\otimes n}$ represents the nth Kronecker product of $\boldsymbol{F}$ where $\boldsymbol{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

2.4 Polar Decoder

SC decoder can iteratively estimate $\hat{\boldsymbol{u}}_1^N = \left(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_N\right)$ from received sequence $\boldsymbol{y}_1^N$. For each bit u_i, if u_i is a frozen bit, directly decoding $\hat{u}_i = 0$; if u_i is a information bit, then calculating log-likelihood ratio (LLR) of the current bit [8]:

$$L_N^{(i)}\left(\boldsymbol{y}_1^N, \hat{\boldsymbol{u}}_1^{i-1}\right) = \log \frac{W_N^{(i)}(\boldsymbol{y}_1^N, \hat{\boldsymbol{u}}_1^{i-1} | u_i = 0)}{W_N^{(i)}(\boldsymbol{y}_1^N, \hat{\boldsymbol{u}}_1^{i-1} | u_i = 1)} \tag{4}$$

where (4) can be calculated recursively by the following two sub-formulas:

$$L_N^{(2i-1)}(\boldsymbol{y}_1^N, \hat{\boldsymbol{u}}_1^{2i-2}) = \frac{L_{N/2}^{(i)}\left(\boldsymbol{y}_1^{N/2}, \hat{\boldsymbol{u}}_{1,o}^{2i-2} \oplus \hat{\boldsymbol{u}}_{1,e}^{2i-2}\right) L_{N/2}^{(i)}\left(\boldsymbol{y}_{N/2+1}^N, \hat{\boldsymbol{u}}_{1,e}^{2i-2}\right) + 1}{L_{N/2}^{(i)}\left(\boldsymbol{y}_1^{N/2}, \hat{\boldsymbol{u}}_{1,o}^{2i-2} \oplus \hat{\boldsymbol{u}}_{1,e}^{2i-2}\right) + L_{N/2}^{(i)}\left(\boldsymbol{y}_{N/2+1}^N, \hat{\boldsymbol{u}}_{1,e}^{2i-2}\right)} \tag{5}$$

$$L_N^{(2i)}(\boldsymbol{y}_1^N, \hat{\boldsymbol{u}}_1^{2i-1}) = \left[L_{N/2}^{(i)}\left(\boldsymbol{y}_1^{N/2}, \hat{\boldsymbol{u}}_{1,o}^{2i-2} \oplus \hat{\boldsymbol{u}}_{1,e}^{2i-2}\right)\right]^{1-2\hat{\boldsymbol{u}}_{2i-1}} \cdot L_{N/2}^{(i)}\left(\boldsymbol{y}_{N/2+1}^N, \hat{\boldsymbol{u}}_{1,e}^{2i-2}\right) \tag{6}$$

In the above equation, $\hat{\boldsymbol{u}}_{1,o}^{2i-2}$ and $\hat{\boldsymbol{u}}_{1,e}^{2i-2}$ denote elements of the odd-numbered and even-numbered positions in the estimated sequence, respectively.

Under the AWGN channel model which obeys Gaussian distribution $N\left(0, \sigma^2\right)$, the initial value of the SC decoding algorithm is set to:

$$L_1^{(1)}(y_i) = \log \frac{W(y_i | u_i = 0)}{W(y_i | u_i = 1)} = \frac{2y_i}{\sigma^2} \tag{7}$$

Finally, the estimates are determined based on the LLR value:

$$\hat{u}_i = \begin{cases} 0 & L_N^{(i)}(\boldsymbol{y}_1^N, \hat{\boldsymbol{u}}_1^{i-1}) \geq 1, \\ 1 & L_N^{(i)}(\boldsymbol{y}_1^N, \hat{\boldsymbol{u}}_1^{i-1}) < 1. \end{cases} \tag{8}$$

3 Bit-Interleaved Polar Coded Modulation (BIPCM) System

The BIPCM system is a simple and practical code modulation scheme. It only needs a single encoder and decoder at the transmitter and receiver, respectively. In this paper, the interleaver [9] is proposed in order to minimize the correlation between coded bits. The 2^m-ary BIPCM system model is shown in Fig. 1, where the modulation order m is a positive integer.

Figure 1 gives a model, the source vector $\boldsymbol{u}_1^N$ consists of information bits and frozen bits, and $\boldsymbol{x}_1^N$ denotes coded sequence. Let vector $\boldsymbol{p}_1^N$ denotes output of interleaver. At the same time, the modulator adopts 2^m-QAM modulation based on Gray mapping. Let $\boldsymbol{t}_1^{N/m}$ and $\boldsymbol{r}_1^{N/m}$ denote the input and output of the channel, respectively. Let $\boldsymbol{q}_1^N$ denote the initial LLR sequence, and $\boldsymbol{y}_1^N$ denote the input of polar decoder. We write $\hat{\boldsymbol{u}}_1^N$ to denote a vector $\hat{\boldsymbol{u}}_1^N = (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_N)$ as the estimated bits.

The performance of BIPCM is found to rely heavily on the design of interleaver, and [5] adopted random interleaver in BIPCM system. For a coded vector $\boldsymbol{x}_1^N$ of length N, we generate N random numbers that take values in [0, 1], where the random number and the coded bits are in one-to-one correspondence. Then, the N random numbers is arranged in descending order, and the coded bits are sequentially read in accordance with the rearranged random number sequence.

Example 1 As an example, we construct a random sequence $\boldsymbol{c}_1^4$ of length 4, $\boldsymbol{c}_1^4 = \{c_1, c_2, c_3, c_4\} = \{0.65, 0.21, 0.76, 0.02\}$, when $\boldsymbol{c}_1^4$ is sorted in descending order $\{c_3, c_1, c_2, c_4\} = \{0.76, 0.65, 0.21, 0.02\}$, bit interleaver converts $\{x_1, x_2, x_3, x_4\}$ to $\{p_3, p_1, p_2, p_4\}$.

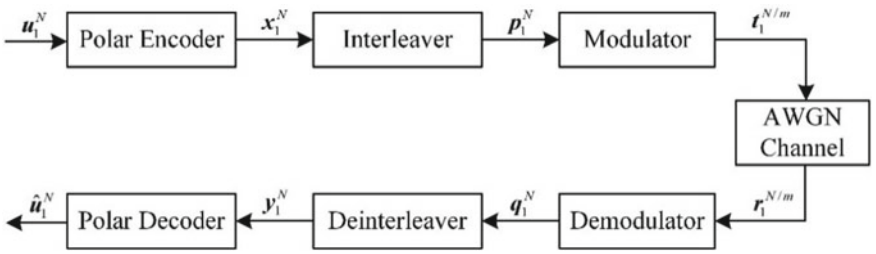


Fig. 1 BIPCM system model

Chen et al. [4] provided an empirically good channel mapping scheme, and in this paper, we applied this method to interleaver. First, we employ PW to calculate the weight of each bit channel, where the weight sequence and corresponding indices are $\boldsymbol{W}_0^{N-1}$ and, respectively. Next, we equalize the vector $\boldsymbol{W}_0^{N-1}$. After the first sorting, we divide $\boldsymbol{W}_0^{N-1}$ into $\boldsymbol{W'}_0^{N/2-1}$ and $\boldsymbol{W'}_{N/2}^{N-1}$, so that $\sum_{i=0}^{N/2-1} W_i$ and $\sum_{i=N/2}^{N-1} W_i$ are as equal as possible, and the corresponding indices are also changed from $\boldsymbol{Q}_0^{N-1}$ to $\boldsymbol{Q'}_0^{N-1}$. Then, we get N sets which only includes one weight and updated $\boldsymbol{Q'}_0^{N-1}$. Finally, let the coded bits are rearranged in the order of $\boldsymbol{Q'}_0^{N-1}$, and $\boldsymbol{p}_1^N$ is obtained as the output of interleaver.

Example 2 As an example, we calculate the polarization weight with code length $N = 4$, where $\boldsymbol{W}_0^3 = 0, 1, 1.18921, 2.18921\}$ and $\boldsymbol{Q}_0^3 = \{0, 1, 2, 3\}$. After the first sorting, we have $W_0'^1 = \{0,\ 2.18921\}$ and $\boldsymbol{W'^3}_2 = \{1, 1.18921\}$, and the corresponding indices are $Q_0'^3 = \{0, 3, 1, 2\}$. After the second sorting, we also have $W_0' = \{0\}$, $W_1' = \{2.18921\}$, $W_2' = \{1\}$, and $W_3' = \{1.18921\}$, and the corresponding indices are $Q_0'^3 = \{0, 3, 1, 2\}$. So, the coded sequence $\boldsymbol{x}_1^N$ are interleaved according to $\boldsymbol{Q'}_0^{N-1}$, which convert $\{x_0, x_1, x_2, x_3\}$ to $\{p_0, p_3, p_1, p_2\}$.

4 Simulation Results

In this section, for random interleaver, equal interleaver, and no interleaver (none) scheme, we adopt the same simulation assumptions: AWGN model, a fixed code rate of 0.5, and 16-QAM based on Gray mapping to simulate the performance of the interleaver on BIPCM system in SC decoding.

Figures 2 and 3 show the performance of the interleaver proposed in this paper when the code length is 512 and 1024, where E_b/N_0 represents the SNR.

As shown in Fig. 2, compared with the BIPCM system with no interleaver, the system with equal interleaver or random interleaver has a better BER performance at low SNR. When the BER $= 10^{-4}$ and code length $N = 512$, the scheme with random interleaver can provide performance gain of 0.25 dB over the system without interleaver. A similar simulation is shown in Fig. 3, where the BER $= 10^{-4}$ and code length is 1024. The results also show that the random interleaver has a better performance of about 0.36 dB improvement. In Figs. 2 and 3, the system with equal interleaver has a better performance at low SNR, compared to the system without interleaver. Especially, when compared with the equal interleaver, the random interleaver is a simple sorting operation on the reliability of bit channel, and its algorithm has low implementation complexity, so the BIPCM system has no significant increase in complexity.

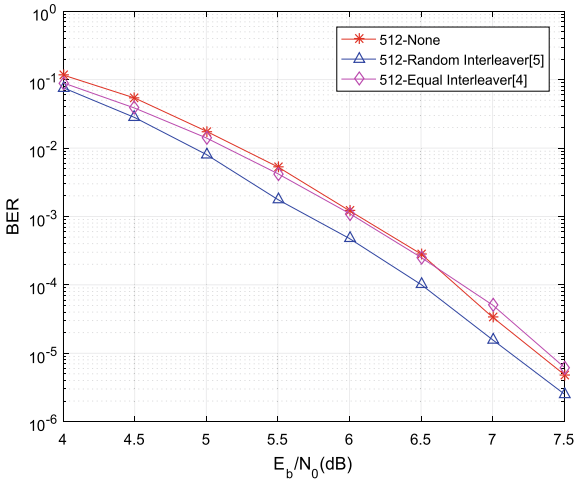


Fig. 2 Performance analysis of interleaver when the code length is 512

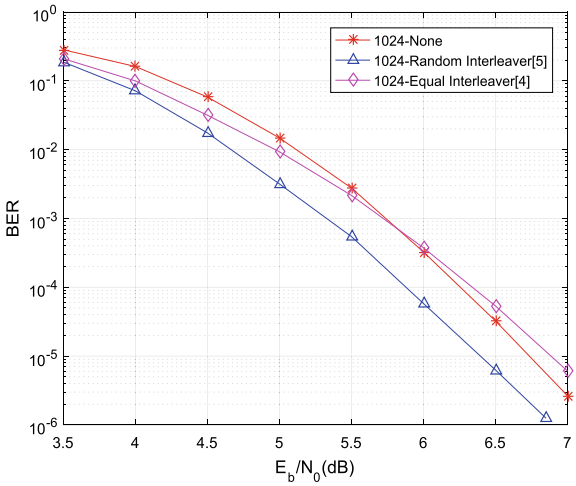


Fig. 3 Performance analysis of interleaver when the code length is 1024

5 Conclusion

For BICM system, polar code with excellent performance for short block lengths and high decoding rate is preferred in 5G communications. A BIPCM system that adopts SC decoding with moderate complexity was created. We also indicate that the polar code can be efficiently constructed by PW. Simulation results for BIPCM system show that random interleaver can provide an additional improvement in BER performance of about 0.2–0.36 dB against the scheme without interleaver. The results also indicate that the performance of BIPCM system with equal interleaver need to

be further improved, and the improvement of the system performance will be further studied in future.

Acknowledgements This paper was funded in part by the National Natural Science Foundation of China (61901182, 61302095), The Natural Science Foundation of Fujian Province of China (2018J01096, 2018J05105), Quanzhou City Science and Technology Program of China (2018C108R), and the Graduate Students Foundation of National Huaqiao University (18013082032).

References

1. Arikan, E.: Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Trans. Inf. Theory **55**(7), 3051–3073 (2009)
2. Zhang, Q.S., Liu, A.J., Pan, X.F., Zhang, Y.X.: Symbol-based belief propagation decoder for multilevel polar coded modulation. IEEE Commun. Lett. **21**(1), 24–27 (2017)
3. Mahdavifar, H., EI-Khamy, M., Lee, J., Kang, I.: Polar coding for bit-interleaved coded modulation. IEEE Trans. Veh. Technol. **65**(5):3115–3127 (2015)
4. Chen, K., Niu, K., Lin, J. R. (2013). An efficient design of bit-interleaved polar coded modulation. In: 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 693–697. London, UK, IEEE Press (2013)
5. Fan, T. T., Yang, W., Xu Changlong.: Performance of BICM system based on polar code in AWGN channel. J. Southeast Univ. (Natural Science Edition) **46**(1), 18–22 (2016)
6. Sharma, A., Salim, M.: Polar code: The channel code contender for 5G scenarios. In: 2017 International Conference on Computer, Communications and Electronics (Comptelix), pp. 676–682. Jaipur, India. IEEE Press (2017)
7. Details of the polar code design (R1-1611254). https://www.3gpp.org/. Last accessed 2019/11/15
8. Du, X. T., Xu, X. F.: The encode and decode theory of polar code and its performance simulating. In: IEEE Press. 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), pp. 24–27. Beijing, China (2016)
9. Fayyaz, U.U.: Symbol mapping design for bit-interleaved polar-coded modulation with iterative decoding. IEEE Commun. Lett. **23**(1), 32–35 (2019)

Research on Technology of Wireless Sensor Network

Zhang Huanan, Xing Suping, and Wang Jiannan

Abstract Wireless communication and embedded technology have made the wireless sensor network (WSN) possible. The sensors, which are deployed in an environment, are required to collect data from their surroundings, process the data, and send it to the sink. Traditional WSNs collect the data such as temperature and pressure. WSN has potential to design many new applications for handling emergency, military, and disaster relief operations that require real-time information for efficient coordination and planning. The self-organization and cooperation of wireless sensor network have opened up a broad application prospect. However, wireless sensor network has the characteristics of resource limitation and design constraint. Resource limitation means that sensor nodes have limited wireless communication range, limited power supply, low bandwidth, limited computing power, and storage capacity. Design constraint means that the design of wireless sensor network depends on its application purpose and the monitored physical environment. This paper first introduces the concept of wireless sensor and expounds the challenges brought by the characteristics of wireless sensor network. Secondly, this paper introduces the representative research work from the aspects of network standard communication protocol network management technology and data compression technology.

1 Introduction

The sensor node is composed of radio transceiver, microcontroller power supply, and multi-type sensor. Sensor nodes with sensing, computing, and communication capabilities are deployed in the designated monitoring area. These nodes are homogeneous and coordinate to collect and process the information of sensing objects in the monitoring area. WSNs construct efficient connections between the physical world and the virtual computing world, making possible the interconnection between physical objects and logical cells. WSNs meet people's real needs to explore and monitor

Z. Huanan (✉) · X. Suping · W. Jiannan
School of Data and Computer Science, Guangdong Peizheng College, Guangzhou, China
e-mail: Topwn2008@163.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_13

the physical world at a low cost and represent a major advance in the traditional way of perception [1].

Sensor nodes can be randomly placed in the monitoring area. In order to ensure the overall performance of the linear sensor network, the strategies and algorithms applied in the network must be self-organized and different from traditional sensing methods. Sensor nodes do not send all the initial sensing data to sink node, but the collected sensing data is processed, calculated, and fused, and only the data of interest to the user is sent to the sink node. Therefore, wireless sensor network has efficient collaboration [2].

The self-organization and collaboration of wireless sensor networks have opened up broad application prospects. For example, in the fields of space exploration, power monitoring, and also medical treatment and nursing, doctors can monitor and control patients' indicators remotely. The physiological data of high frequency and multiple indicators are helpful for doctors to know the health status of patients more accurately and provide reliable basis for them to make effective treatment plans. Sensor networks can also be used in pollution monitoring and control. In fact, the wireless sensor network can intelligently obtain the effective information of the monitoring environment, giving people a clear and accurate picture of the physical world. Wireless sensor network has been widely used in intelligent transportation, green building, intelligent home disaster warning, and other fields [3].

Different from traditional network systems, wireless sensor networks have design constraints and resource constraints. Specifically, resource-constrained means that sensor nodes exhibit restricted wireless communication range, limited power supply low bandwidth, and limited computing capacity and storage capacity. The design constraint means that the design of wireless sensor network depends on its application purpose and the monitored physical environment. Among them, the physical environment determines the factors such as network topology of network scale node deployment. In general, the network size will depend on the physical environment. Obstacles in the physical world also limit wireless communication between sensor nodes, thereby reducing the connectivity of the network topology. Based on this, research on wireless sensor networks is committed to introduce new design concepts, innovating or improving new algorithms for current network protocol design, aiming to meet design constraints and improve network performance [4]. In the following sections, we will introduce wireless sensor key technologies.

2 Key Technologies of Wireless Sensor Networks

The development and key technologies of wireless sensor network can be broadly divided into three groups. As shown in Fig. 1, there is a separate system for each sensor in the first group. In order to meet the application requirements of sensor network, new operating system platform and storage model were designed and developed. The second group is the communication protocol. The communication protocol not only guarantees the wireless communication between sensor nodes, but also establishes

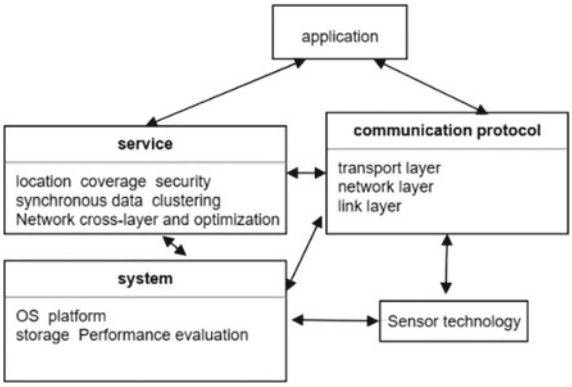


Fig. 1 Key issues of wireless sensor networks

the communication between the upper application and the lower node. The third group is the service research for serving wireless sensor networks, aiming to improve network performance and system efficiency, so as to improve the level of industrial application.

From the perspective of application requirements and network management, the self-organization of sensor nodes is of global importance. Sensor nodes can not only spontaneously organize multi-hop wireless network systems, but also it is possible for nodes to be efficiently distributed and controlled. Due to the limited power supply and storage capacity of sensor nodes, it is of great significance and value to carry out the research on communication protocol and management service to meet the above requirements [5].

The standard communication protocol is generally divided into five layers, specifically, the application layer, the transmission layer, the network layer, the data link layer, and the physical layer. The communication protocols of different layers have important effects on energy consumption, end-to-end delay network congestion, and network throughput. Therefore, the study of efficient communication algorithms has become the focus of the development of the field. The information sharing and interaction between different protocol layers make the cross-layer optimization algorithm possible. Sensor nodes only use batteries to supply energy, so the quota is naturally limited. Once the power supply of the node is exhausted, the node will fail and logically leave the sensor network, which reduces the reliability of the sensor network. The lifetime of the sensor network depends on the effective number of nodes and network connectivity. Therefore, energy efficiency has become the primary index to measure the performance of sensor network algorithm [6].

A great deal of research work is to establish a reliable wireless communication to build an optimized node layout and design a complete data aggregation method to achieve high energy efficiency and good quality of service (QoS) requirements. Moreover, for the communication result pattern, the proposed and implemented data recovery strategy of the congestion control buffer monitoring information confirmation mechanism provides a basic technical guarantee for the reliable transmission

of packets. In addition, the existing research shows that the wireless communication strength depends on the sensor nodes configuration. The sparse configuration scheme will lead to the remote communication between nodes, resulting in high energy consumption. Although the intensive configuration scheme reduces the communication energy consumption, it is easy to cause communication conflicts [7].

3 Network Standard

Different types of sensor hardware is uneven, and how to make a variety of sensor nodes to achieve stable organic combination for sensor network platform development and system design puts forward important research subject. Network standard defines local area network (LAN) for the sensor nodes and the physical media access control protocol in order to meet a variety of network, the IEEE 802.15.4 standard makes the low rate, low overhead, low power, and low complexity of the topological structure of the wireless local area network to become a reality; also, IEEE802.15.4 provides a standard for physical layer and data link layer, and has been widely used in hot fields such as industrial automation of intelligent home environment monitoring.

Based on the IEEE 802.15.4 standard, ZigBee standards have defined the high-level network communication protocol. ZigBee is a kind of simple wireless communication technology and has advantages of low power consumption. ZigBee uses the network topology structure and can connect lots of wireless communication devices at the same time. Safe and energy-efficient wireless communication technology can provide technical support for channel frequency hopping and message passing synchronization, and is applicable to a variety of network topologies, including mesh network, star network, and hybrid network.

4 Communication Protocol

Sensor nodes communicate wirelessly with other nodes through the protocol stack. In order to ensure that, sensor nodes can efficiently and cooperatively conduct intra-network data transmission. A lot of important achievements have been carried out around the design of data link layer and cross-layer network communication protocol of transport layer, such as the sensor transmission control protocol (SCTP), which provides congestion detection for the transport layer as a kind of reliable transport layer protocol, under the condition to meet the reliability requirements of the transport layer, minimizing communication overhead; the pump slowly, fetch quickly (PSFQ) protocol is a scalability and robustness of the ideal transport layer protocol and the agreement to protect the data block transmission, but also provides data recovery strategy in the above results; the congestion detection avoidance (CODA) protocol can alleviate network congestion rapidly and is a kind of energy-efficient congestion control strategy. In conclusion, the network layer protocol is mainly aimed at the

data routing within the network and aims to realize efficient data transmission under the condition of limited energy supply and communication bandwidth [8].

5 Network Management Technology

Network management technology is mainly engaged in node localization topology control time. The mobile-assisted localization (MAL) location algorithm is the root, according to the mobile node and the distance between the static sensor nodes to determine the location of the sensor node immediately. SeRloc algorithm is based on directional antenna. It is equipped with locator node monitoring information and provides location information for sensor nodes. The research is of security positioning algorithm, the algorithm to ensure the safety of the positioning process. In addition, the distributed reputation and trust-based security (DRBTS) protocol can identify false bit information based on the monitoring information of beacon nodes [9].

6 Data Compression Technology

Distributed data compression methods of sensor network are mainly divided into the following four types: distributed source modeling (DSM) technology, distributed transcoding (DTC) technology, distributed source coding (DSC) technology, and compressed sensing (CS) technology. The DSM methods are largely divided into model parameter model method and nonparametric method. Simulation algorithm based on parameter model will perceive the data as a stochastic process, and parameter estimation is optimized to achieve the statistics when sensory data structure of statistics known parameter model. However, when the perception data is sparse or the prior knowledge is fuzzy, the nonparametric model method shows high robustness while the nonparametric model method is mainly devoted to the study of the fitting function of the perception data [10].

A WSN lossless compression algorithm is a kind of lossless compression technology. The technology is based on the same sensor node with no similar cycle and the characteristics of the sensor at the same time. Collecting information is very similar to information preprocessing based on data compression technology. Before compression, the information is compressed more effectively to improve the compression rate and make the energy saving effect more obvious [11].

Based on the original data of temporal correlation and spatial correlation, the data compression method of sensor network based on sequence correlation is carried out considering multidimensional correlation information. An algorithm evaluation model can be designed based on the energy criterion. This algorithm can effectively reduce the amount of data in the network, prolong the network life cycle, and achieve good energy-saving effect [12].

7 Conclusion

With the continuous development of sensor technology, monitoring application based on large-scale sensor network has become possible. However, the research on the availability of large-scale sensor network data and perceived data is still an academic field that has not been deeply studied. Therefore, in the next step, we should take the above characteristics into consideration and propose a more effective and practical algorithm based on the application scenario of sensor network.

References

1. Collotta, M., Pau, G.: An innovative approach for forecasting of energy requirements to improve a smart home management system based on BLE. IEEE Trans. Green Commun. Netw. **1**, 112–120 (2017)
2. Yang, C.-H., Ssu, K.-F.: An energy-efficient routing protocol in underwater sensor networks. In: Proceedings of the 2008 3rd International Conference on Sensing Technology, pp. 114–118. Tainan, Taiwan, 30 Nov.–3 Dec. 2008 (2008)
3. He, T., Vicaire, P.A., Yan, T., Luo, L., Gu, L., Zhou, G., Stoleru, R., Cao, Q., Stankovic, J.A., Abdelzaher, T.: Achieving real-time target tracking using wireless sensor networks. ACM Trans. Embed. Comput. Syst. (2007)
4. Han, B., Zhang, D.Z., Yang, T.: Energy consumption analysis and energy management strategy for sensor node. In: Proceedings of the 2008 International Conference on Information and Automation, vol. 6, pp. 211–214. IEEE (2008)
5. Tsiatsis, V., Zimbeck, S., Srivastava, M.: Architectural strategies for energy efficient packet forwarding in wireless sensor networks. In: Proceedings of the 2001 International Symposium on Low Power Electronics and Design, vol. 3, pp. 25–31 (2001)
6. Xing, G., Tan, R., Liu, B., Wang, J., Jia, X., Yi, C.W.: Data fusion improves the coverage of wireless sensor networks. In: Proceedings of 15th Annual International Conference on Mobile Computing and Networking, pp. 157–168. New York (2009)
7. Kwon, H., Kim, T.H., Choi, S., Lee, B.G.: A cross-layer strategy for energy-efficient reliable delivery in wireless sensor networks. IEEE Trans. Wirel. Commun. **5**(12), 3689–3699 (2006)
8. Hong, Y.W., Scaglione, A.: Energy-efficient broadcasting with cooperative transmissions in wireless sensor networks. IEEE Trans. Wirel. Commun. **5**(10), 2844–2855 (2006)
9. Kulik, J., Heinzelman, W., Balakrishnan, H.: Negotiation-based protocols for disseminating information in wireless sensor networks. Wirel. Netw. **8**(2–3), 169–185 (2002)
10. Zhu, J., Feng, Y., Liu, M., Chen, G., Huang, Y.: Adaptive online mobile charging for node failure avoidance in wireless rechargeable sensor networks. Comput. Commun. **126**, 28–37 (2018)
11. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Comput. Netw. **52**(12), 2292–2330 (2008)
12. Zhou, K., Fu, C., Yang, S.: Big data driven smart energy management: from big data to big insights. Renew. Sustain. Energy Rev. **56**, 215–225 (2016)

Characteristics of Scholte Wave for Target Detection

Bo Ren and Huan Li

Abstract ANSYS/LS-DYNA in the flow-solid coupling method is used to establish a model of the marine environment to simulate the seabed seismic wave which is caused by low-frequency point sound sources. The characteristic and propagation law of Scholte wave are obtained through analyzing and calculating. Experiments show that Scholte wave has good performance in underwater target detection.

1 Introduction

The low-frequency noise of navigating ships causes vibration in the bedrock of the seafloor, producing elastic waves which are transmitted to distant places. This wave is called ship seismic wave. It is suitable for underwater target detection because of the characteristics of the low-frequency, large amplitude than body waves, slow attenuation, propagation distance [1, 2]. The study of the Scholte surface wave characteristics and the law of propagation used in the horizontal layered marine environment low-frequency point source sound wave seismic submarine with the help of the finite element software ANSYS/LS-DYNA in the fluid–solid coupling method, combined with non-reflective boundary conditions.

2 Fluid Solid Coupling Algorithm

In the finite element analysis, the software ANSYS/LS-DYNA provides a fluid–solid coupling method, which can be used to calculate the fluid–solid interaction. It also

B. Ren
College of Equipment Engineering, Shenyang Ligong University, 110159 Shenyang, China

H. Li (✉)
School of Information Science and Engineering, Shenyang Ligong University, 110159 Shenyang, China
e-mail: lihuan9999@yeah.net

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_14

can be used for low-frequency point source seismic wave numerical simulation. The governing equations of the solid and fluid are [3]:

$$M_s\ddot{U} + C_s\dot{U} + K_sU = F_s + RP$$

$$M_fP + C_fP + K_fP = F_f - \rho_0 R^T\ddot{U} \tag{1}$$

where M_s, C_s, and K_s are the mass matrix, the damping matrix, and the stiffness matrix of the solid structure, respectively. M_f, C_f, and K_f are the mass matrix, the damping matrix, and the stiffness matrix of the fluid medium. U and P are the finite element nodes displacement and sound pressure. F_s and F_f are the external forces applied to the solid structure and the pressures acting on the fluid medium. And R is the coupling matrix, representing the effective surface area associated with the node on the fluid–solid structure interface. Let us put F_s and F_f on the right side of the equation and the unique harmony pressure of other unknown nodes on the left. Combining the above two equations into one equation, the finite element discrete equation of the fluid–solid coupling problem can be obtained as follows [3].

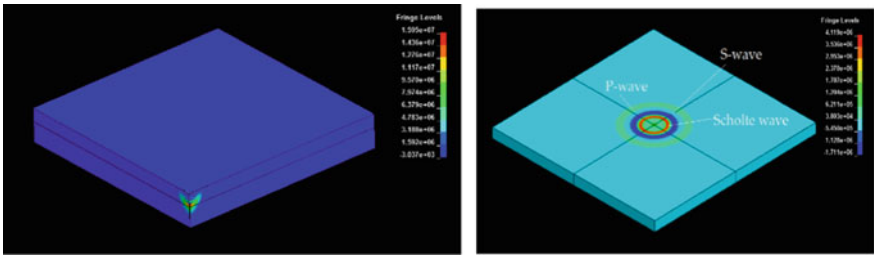
$$\begin{bmatrix} M_s & 0 \\ M^{fs} & M_f \end{bmatrix}\begin{bmatrix} \ddot{U} \\ \ddot{P} \end{bmatrix} + \begin{bmatrix} C_s & 0 \\ 0 & C_f \end{bmatrix}\begin{bmatrix} \dot{U} \\ \dot{P} \end{bmatrix} + \begin{bmatrix} K_s & K^{fs} \\ 0 & K_f \end{bmatrix}\begin{bmatrix} U \\ P \end{bmatrix} = \begin{bmatrix} F_s \\ F_f \end{bmatrix}, \tag{2}$$

where $M^{fs} = \rho_0 R^T$, $K^{fs} = -R$ are the mass matrix and the stiffness matrix of the fluid–solid interface.

3 Modeling of Finite Element

In the model, the upper part is seawater, the lower part is sea bedrock, and the interface between seawater (fluid) and sea bedrock (solid) is the flow–solid interface FSI. The top surface is free surface, the left and front of the calculation model are symmetric surfaces (xoy and yoz) under the real ocean environment, the horizontal and vertical directions are infinite, but the numerical calculation model cannot be infinite. The original infinite field can be truncated to a finite fluctuation by applying non-reflective boundary conditions on the boundary conditions on the boundary surface calculated domain [4]. In the computational model presented here, non-reflective boundary conditions are applied to the right, back, and bottom of the model.

Due to the symmetric low-frequency underwater seismic wave model of noise source, simplicity, a quarter model is set up. The model's length, width, and height are 500 m, 500 m, and 100 m, respectively. The parameters of the model are as follows. The density of seawater is 1000 kg/m^3, the velocity of wave is 1500 m/s, the density of sea bed rock is 2650 kg/m^3, the longitudinal wave velocity is 3098 m/s, the shear wave velocity is 1789 m/s, the depth of water is 40 m, and the depth of bedrock is 60 m.



(a) To bedrock (b) At the fluid-solid interface

Fig. 1 Wave propagation

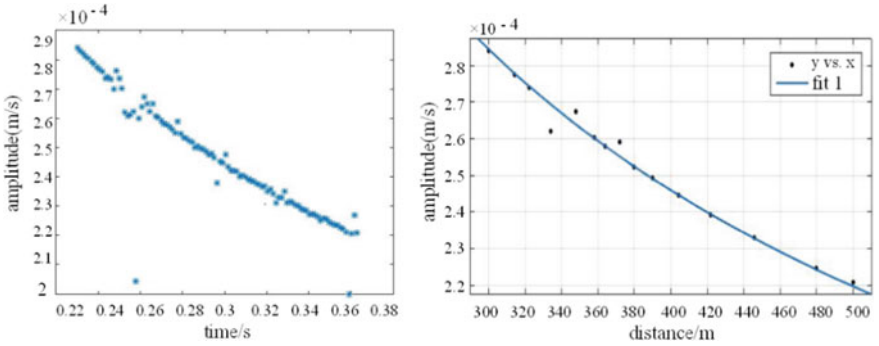
4 Simulation Results and Analysis

4.1 The Propagation of Submarine Seismic Waves

Figure 1a is a waveform diagram of wave propagation to bedrock under the action of point sound. At the time of 29 ms, the pressure waves generated by the point sound begin to contact the bedrock. In Fig. 1b, the pressure wave is produced under the action of point source, which was propagated through the seawater medium to the subsoil medium, and propagates at the fluid–solid interface. Due to the different propagation velocity of the wave, it can be obviously seen an array of three waves (*P*-wave, *S*-wave, and Scholte wave). The *P*-wave travels fastest, and the *S*-wave takes the second place. The Scholte wave propagates the slowest at the flow–solid interface and travels far away.

4.2 The Characteristic of Scholte Wave in Horizontal Direction

In horizontal direction, the data of the time–amplitude curve of Scholte wave between 300 and 500 m is extracted and drawn in Fig. 2a. As the distance increases, the amplitude decays according to a certain rule. From the trend of the curve, the decay relation is set as $s = ar^{-x}$ where s and r are amplitude and distance, respectively [5]; x is unknown and can be obtained from the MATLAB fitting toolbox. The results are shown in Fig. 2b. We select 15 sets of data for fitting, and get $x = -0.5047, -0.488$. As the distance increases, the amplitude of the Scholte wave in horizontal and vertical direction decreases approximately to $1/\sqrt{r}$. So, Scholte waves can propagate further in the shallow sea than the transverse and longitudinal waves.



(a) Time-amplitude (b) Fitting graph of amplitude-distance

Fig. 2 Curves of Scholte wave attenuation in horizontal direction

4.3 The Characteristic of Scholte Wave in Vertical Direction

In vertical direction, amplitude–time curves of the seabed depths of 600 m from the source at 0 m, 1 m, 2 m, and 3 m, respectively, were extracted.

Figure 3 shows the motion trajectory of particles extracted in the depth direction when the sea bed medium is bedrock.

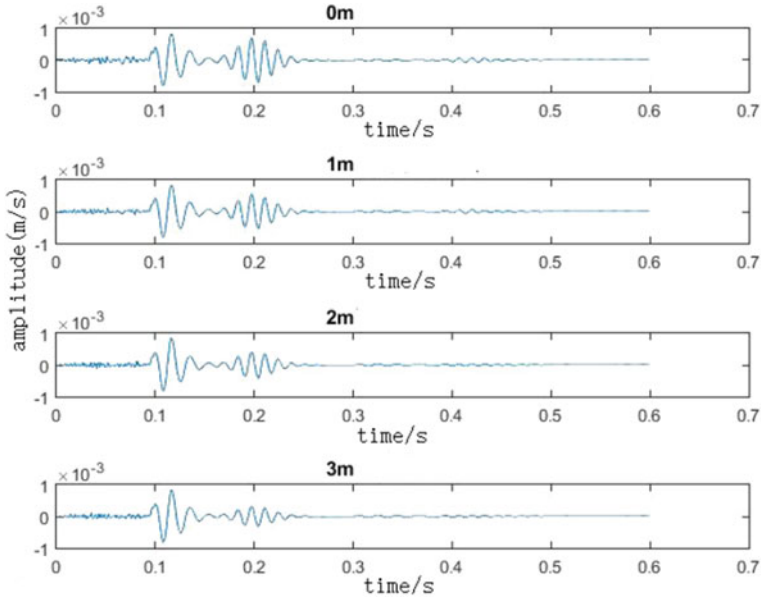


Fig. 3 Horizontal components of the particles at different depths

The Scholte wave in Fig. 3 is polarized in order to analyze the attenuation of the Scholte wave in the depth direction, as is shown in Fig. 4. We can see that the Scholte wave's motion trajectory is an inverse ellipse. As the depth increases, the amplitude of the Scholte wave gradually decays, and the maximum value of the horizontal component in the graph is extracted and given in Table 1.

The maximum amplitude at different depths is given in Table 1. According to the data in Table 1, Scholte wave fitting curve can be obtained as shown in Fig. 5.

From the curve in Fig. 5, we can see that as the distance increases, the amplitude of the curve decays according to a certain rule. From the trend of the curve, the

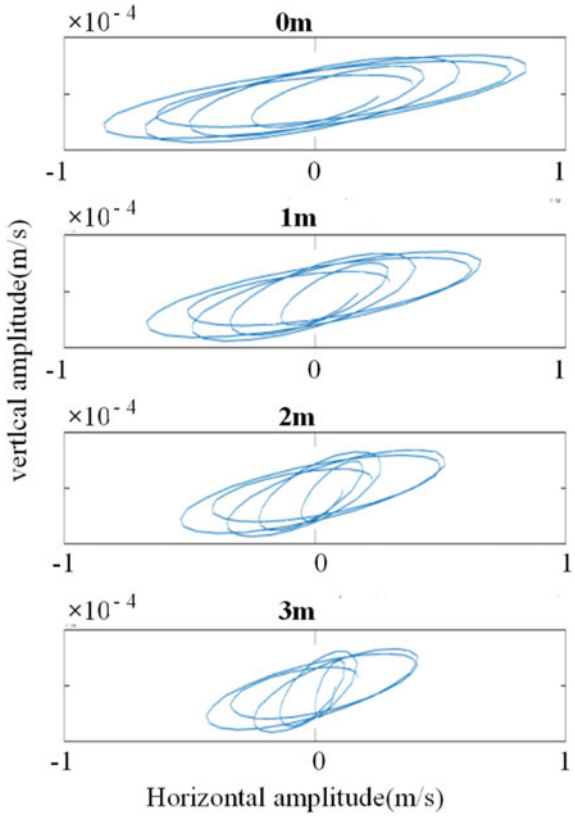


Fig. 4 Polarization curves of Scholte wave at different depths

Table 1 Maximum amplitude at different depths

Depths (m)	Amplitude (m/s)
0	0.00006871
1	0.00005214
2	0.00003859
3	0.00003031

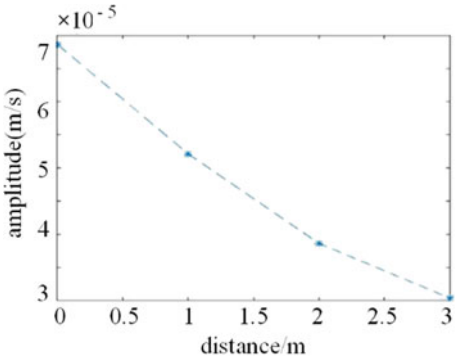


Fig. 5 Fitting curve of Scholte wave

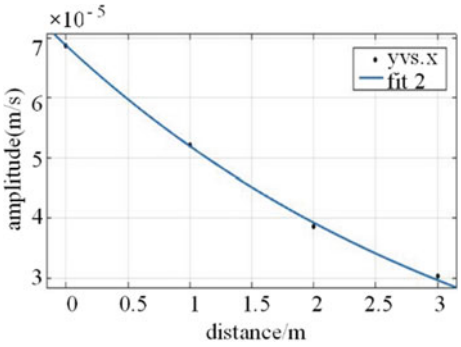


Fig. 6 Decay curve of Scholte wave

attenuation relationship between the amplitude and the distance is set as $s = ae^{br}$ where s and r are amplitude and distance, respectively, a is a known data, b is unknown which can be obtained by the fitting toolbox in MATLAB. The four sets of data in Table 1 are selected for fitting as shown in Fig. 6. From Fig. 6, we get $b = -0.2407$.

5 Conclusions

With the increase of depth, the Scholte wave attenuates exponentially in the depth direction. The Scholte wave travels the slowest in each wave, and its trajectory of the mass point is an inverse ellipse in the depth direction. At the interface, as the horizontal distance increases, the amplitude of Scholte wave decays. Scholte wave decays exponentially in the vertical direction away from the interface. And Scholte wave has the slowest attenuation among the three waves and has good detection performance in submarine target detection.

Scholte wave can work at a frequency band. For different earthquake sources, the center frequency is different. Target detection can be realized by identifying Scholte waves of different frequencies. The Scholte wave is a band consisting of a set of

nonlinear frequencies whose center frequencies are different for different sources. Buried mines can be probed by identifying the frequencies of different Scholte wave. The experimental verification of the target detection of buried mine by Scholte wave proves that the Scholte wave has good detection performance in the detection of the underwater target.

Compared with the difference calculation method, the finite element method adopted in this paper does not eliminate the high-order terms in the calculation process, so the calculation accuracy is higher.

Acknowledgements The authors acknowledge the financial support from natural science foundation of Liaoning Province under Grant 2015010028-301.

References

1. Simonetti, F., Cawley, P.: On the nature of shear horizontal wave propagation in elastic plates coated with viscoelastic materials. Proc. R. Soc. A Math. Phys. Eng. Sci. **1284**(2048), 2197–2221 (2004)
2. Lu, Z., Zhang, Z., Gu, J.: Numerical simulation of low frequency seismic wave in layered seabed of porous media. J. Vet. Sci. **35**(12), 2065–2071 (2014)
3. Yan, B., Zhou, W., Gong, S.: A simple normal wave model for the propagation of shallow sea seismic waves. J. Wuhan Univ. Sci. Technol. **30**(5), 804–807 (2006)
4. Chen, Y., Lu, J.: Shipwreck seismic wave and its application in underwater target detection. J. Ship Sci. Technol. **27**(3), 62–66 (2005)
5. Takiy, A.E., Granja, S.C., et al.: Theoretical analysis and experimental validation of the Scholte wave propagation in immersed plates for the characterization of viscous fluids. In: 2013 IEEE International Ultrasonics Symposium (IUS), 21–25 (2013)

Visible and Infrared Image Fusion Based on Online Convolutional Dictionary Learning with Sparse Matrix Computation

Chengfang Zhang and Xingchun Yang

Abstract To overcome the pseudo-Gibbs effect caused by multi-scale transformation fusion method and the block effect caused by sparse representation fusion method, this paper first learns redundant dictionary filter using online convolutional dictionary learning method with sparse matrix computation. The learned dictionary filters are then applied to convolution sparse representation image fusion. Five infrared and visible images are used to prove the performance of the proposed algorithm. The experimental results show that our algorithm not only objectively obtains a higher evaluation index, but also meets the subjective evaluation of human eyes. Compared with the ASR-based method, the normalized mutual information (NMI), the Tsallis entropy (QTE), and the nonlinear correlation information entropy (QNCIE) are increased by 8.34%, 9.16%, and 0.04% averagely.

1 Introduction

With the continuous improvement of image quality acquired by infrared and visible light imaging sensors, multi-modal images obtained have been widely used in aerospace, safety monitoring, and other fields. However, different imaging sensors are limited by the imaging principle. When the same working environment and the same area are observed at the same time, they can get their own unique information, which contains redundancy and a lot of complementary information. For example, infrared thermal imager is used to detect the thermal radiation energy distribution of the target and is not affected by the illumination conditions; visible light sensor is used to detect the spectral reflection characteristics of the scene, which can reflect the texture details of the scene, but is limited by the illumination conditions. Therefore, for the disadvantage of single sensor imaging is not ideal, the use of multi-sensor cooperation can make full use of the complementary information between sensors and synthesize the target features of infrared image and the texture details of the

C. Zhang · X. Yang (✉)
Sichuan Police College, Luzhou, China
e-mail: zcf1838725417@163.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_15

scene in visible image, so that the image contains more information, and the visual effect is better.

The main fusion methods of infrared and visible images are in the transform domain, represented by multi-scale transform, such as discrete wavelet transform [1], dual-tree complex wavelet transform, and contour wave transform, but these transform methods do not have translation invariance, which results in blurred edge details of the fused image. At the same time, if the gray level of the visible image changes a little, it will not concentrate on the low-pass part. This fusion method will lead to fused image background (not clear), and the overall contrast is low. Sparse representation (SR) can express the deeper structural features between low-frequency subband coefficients, making the linear combination of a small number of atoms in the dictionary approximately perfect. In recent years, many image fusion algorithms based on sparse domain [2–11] have been proposed.

At present, the most popular convolutional dictionary algorithm is batch learning method, which alternately minimizes coefficient maps and dictionary filters, and processes the entire training set at each iteration. When the number of training signals is large, the dictionary filters update sub-problem is computationally expensive. To solve the above shortcomings, Liu proposed a new online convolutional dictionary learning method based on stochastic gradient descent with sparse matrices [12]. In this paper, a stochastic gradient descent with sparse matrices is used in convolution dictionary learning phase, and the convolution dictionary is used for the convolutional sparse coding. The proposed image fusion algorithm not only objectively evaluates the fusion, but also preserves the edges and the details of the fused image better.

This paper is organized as follows. Section 2 briefly introduces the online convolutional dictionary learning method based on the stochastic gradient descent with sparse matrices. Section 3 presents the fusion framework. The experimental results are given in Sect. 4. Section 5 summarizes the conclusions of this paper.

2 Online Convolutional Dictionary Learning

In the literature [12], Liu presented a new online convolutional dictionary learning method based on the stochastic gradient descent with sparse matrix computation technique. Specifically, for a given set of K training images $\{s_k\}_{k=1}^{K}$, the online convolutional dictionary learning is executed via minimizing the following problem

$$\arg\min_{\{x_{k,m}\},\{d_m\}} \frac{1}{2}\sum_{k=1}^{K}\left\|\sum_{m=1}^{M} d_m * x_{k,m} - s_k\right\|_2^2 + \lambda\sum_{k=1}^{K}\sum_{m=1}^{M}\left\|x_{k,m}\right\|_1$$
$$\text{subject to } \|d_m\|_2 \le 1, \quad \forall m \in \{1,\ldots,M\} \tag{1}$$

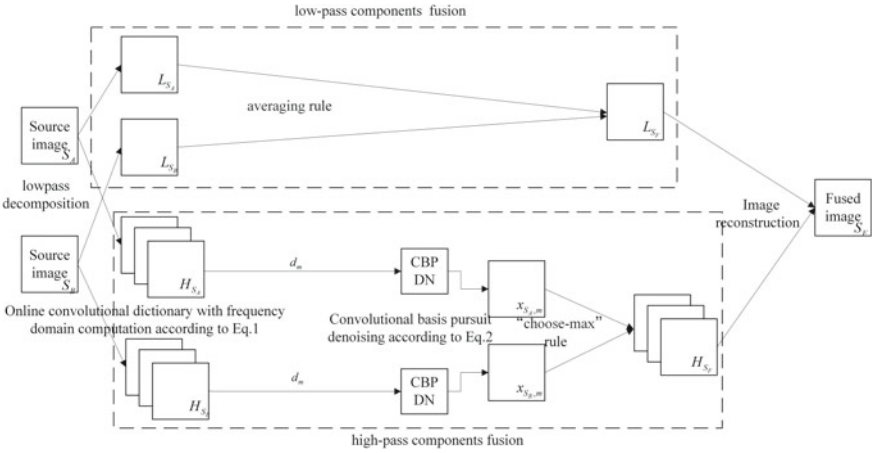


Fig. 1 Fused diagram of proposed method

where * is the convolution operator,$x_{k,m}, k \in \{1, \ldots, K\}, m \in \{1, \ldots, M\}$ is the coefficient maps. The size of the dictionary filter d_m is $8 \times 8 \times 32$, and the training images were from the USC-SIPI image database [5].

3 The Proposed Method

Figure 1 shows the framework of the proposed fusion algorithm. Here, S are the source images, and the dictionary filter is $\{S_A, S_B\} \in R^N$, $d_m, m \in \{1, \ldots, M\}$, respectively. For given d_m and S, x_m can be solved by the convolutional basis pursuit denoising (CBPDN) ℓ_1-minimization problem

$$\arg\min_{\{x_m\}} \frac{1}{2} \left\| \sum_m^M d_m * x_m - S \right\|_2^2 + \lambda \sum_m^M \|x_m\| \tag{2}$$

where * is the convolution operator and λ denotes the regularization parameter.

4 Experiments

In our experiment, the scheme is compared with discrete wavelet transform (DWT), adaptive sparse representation (ASR), and joint sparse representation (JSR) to verify the effectiveness and excellence. Five infrared–visible images were tested under MATLAB R2016a with a 3.20 GHz CPU and a 4.00 GB RAM. In our paper, three common objective evaluation indicators (such as the normalized mutual information

(NMI) [13, 14], the fusion metric-based on Tsallis entropy (QTE) [13, 15], and the nonlinear correlation information entropy (QNCIE) [13, 16]) are used.

The fusion result of an infrared–visible image is shown in Fig. 2. The first and second columns of Fig. 2 are the source visible image and the infrared image, respectively. The third column of Fig. 2 is the fusion result of DWT, the fourth is the fusion result of ASR, and the fifth is the fusion result of JSR. The fusion result of our algorithm is the sixth column of Fig. 2. According to the experimental results of Fig. 2, we can see that the DWT information is blurred, and the contrast is not high; the ASR fusion image is distorted, there are artifacts in some local areas, and the spatial continuity of the image is not strong; the JSR partial fusion image information is seriously damaged (such as row 4, column 4, and column 5). The results obtained by our scheme retain the advantages of infrared and visible light.

In order to quantitatively evaluate the performance of the algorithm, the fusion image is measured by the objective evaluation indicators mentioned above. The measurement results are given in Table 1. As can be seen from Table 1, the measurement results of this algorithm are NMI = 0.3481, QTE = 0.4920, QNCIE = 0.8048. NMI, QTE, and QNCIE are all evaluation criteria based on the information theory, and the above metrics value are the highest, indicating that the fusion image is the

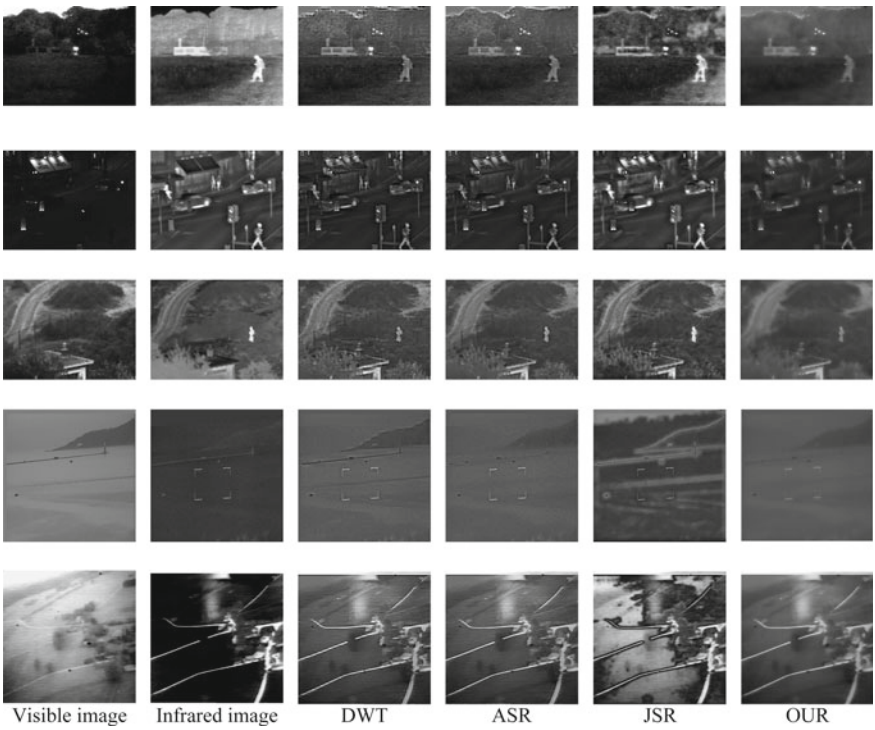


Fig. 2 Fused result of the visible–infrared image

Table 1 Objective assessment of various methods for the visible–infrared image

	NMI	QTE	QNCIE
DWT	0.3031	0.4276	0.8042
ASR [12]	0.3213	0.4507	0.8045
JSR [3]	0.2627	0.4017	0.8039
OUR	**0.3481**	**0.4920**	**0.8048**

most similar to the source image and the more relevant information between the fusion result and the source image. Compared with the other three methods, the three objective evaluation indexes of our algorithm are the highest. Compared with ASR-based method, NMI, QTE, and QNCIE increased by 8.34, 9.16, and 0.04% averagely.

5 Conclusion

To overcome the low contrast of the traditional multi-scale transform fusion algorithm and the block effect of the traditional sparse domain fusion, a new infrared and visible image fusion method based on online convolution dictionary learning is proposed according to characteristics of the infrared-visible images. Through subjective and objective evaluation of the four methods, it is shown that our method is better than the other three methods, and the clarity of the fusion image is improved, which is more in line with the visual characteristics of human eyes.

Acknowledgements This work was supported by Luzhou Science and Technology Program (2019-SYF-34), Scientific Research Project of Sichuan Public Security Department (Grant 201917), and the Sichuan Science and Technology Program (Grants 2019YFS0068 and 2019YFS0069).

References

1. Zhou, Z.H., Tan, M.: Infrared image and visible image fusion based on wavelet transform. Adv. Mater. Res. **756–759**, 2850–2856 (2013)
2. Liu, Y., Wang, Z.: Simultaneous image fusion and denoising with adaptive sparse representation. IET Image Proc. **9**(5), 347–357 (2015)
3. Yin, H.T., Li, S.T.: Multimodal image fusion with joint sparsity model. Opt. Eng. **50**(6), 067007 (2011)
4. Zhang, C.F., Yi, L.Z., Feng, Z.L., Gao, Z.S., Jin, X., Yan, D.: Multimodal image fusion with adaptive joint sparsity model. J. Electron. Imag. **28**(1), 013043 (2019)
5. Gao, Z.S., Zhang, C.F.: Texture clear multi-modal image fusion with joint sparsity model. Optik Int. J. Light Electron Opt. **130**, 255–265 (2016)
6. Zhang, C.F., Yi, L.Z., Yan, D., Pei, Z.: Visible and infrared image fusion based on convolutional sparse coding with gradient regularization. In: The 14th International Conference on Intelligent Systems and Knowledge Engineering (2019)

7. Zhang, C.F., Yue, Z., Yan, D., Yang, X.C.: Infrared and visible image fusion using joint convolution sparse coding. In: The 2nd International conference on Image, Video Processing and Artificial Intelligence (2019)
8. Zhang, C.F.: Multi-focus image fusion based on convolutional sparse representation with mask simulation. In: The 3rd International Conference on Mechatronics and Intelligent Robotics (2019)
9. Zhang, C.F., Chen, Y.L., Yi, L.Z., Yang, X.C., Jin, X., Yan, D.: Image fusion based on convolutional sparse representation with mask decoupling. In: The 3rd International Conference on Mechatronics and Intelligent Robotics (2019)
10. Zhang, C.F., Yue, Z., Yi, L.Z., Yan, D., Jin, X., Yang, X.C.: Infrared and visible image fusion using NSCT and convolutional sparse representation. In: The 10th International Conference on Image and Graphics (2019)
11. Wang, K.: Rock particle image fusion based on sparse representation and non-subsampled contourlet transform. Optik **178**, 513–523 (2019)
12. Liu, J.L., Cristina, G.C., Brendt, W.: First- and second-order methods for online convolutional dictionary learning. SIAM J. Imaging Sci. **11**(2), 1589–1628 (2018)
13. Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., Wu, W.: Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. IEEE Trans. Pattern Anal. Mach. Intell. **34**(1), 94–109 (2011)
14. Qu, G., Zhang, D., Yan, P.: Information measure for performance of image fusion. Electron. Lett. **38**(7), 313 (2002)
15. Cvejic, N., Canagarajah, C.N., Bull, D.R.: Image fusion metric based on mutual information and Tsallis entropy. Electron. Lett. **42**(11), 626 (2006)
16. Madhuri, G., Bindu, C.H.: Performance evaluation of multi-focus image fusion techniques. In: 2015 International Conference on Computing and Network Communications. IEEE (2015)

Optimization of Maintenance Force Dispatch of Artillery Command Information System Based on Mission Success

Hongtu Cai, Yuwen Liu, Pengfei Ma, Yaoze Han, and Yong Wang

Abstract Aiming at the problem that maintenance force is limited and maintenance time is short, this paper constructs an optimization model of maintenance support force dispatch for artillery command information system based on the success of the task. And the validity of the method is verified by an example.

1 Introduction

Maintenance force dispatch is an important part of maintenance support of artillery command information system. The result directly determines whether the failure can be handled in time and whether the system can efficiently complete the scheduled tasks. In the case of limited maintenance support force, how to achieve the overall recovery and capability rebirth of the combat effectiveness of the system in the shortest time and ensure the successful completion of the task is an important issue to be solved urgently.

2 Problem Analysis

2.1 Maintenance Task Relation Analysis

According to the composition of artillery command information system, maintenance task relationship can be divided into series relationship, parallel relationship and series-parallel hybrid relationship.

Series relation means that the maintenance of system equipment must be carried out in chronological order.

H. Cai · Y. Liu (✉) · P. Ma · Y. Han · Y. Wang
Army Artillery and Air Defense College, 230031 Hefei, China
e-mail: sypycht@126.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_16

Parallel relation means that the maintenance of system equipment can be carried out at the same time without time.

Series-parallel hybrid relation refers to the simultaneous existence of series relation and parallel relation in maintenance task of system equipment.

The above three relationships are not only a full reflection of the internal relationship of the system equipment, but also an objective reflection of the maintenance sequence of the system equipment.

2.2 Maintenance Force Dispatch Classification

According to the relationship between maintenance tasks and the number of maintenance personnel, maintenance tasks can be divided into the following categories [1].

The first is to share a task with one person. It mainly diagnoses the maintenance tasks with series or parallel relations in the system.

The second is one person shares many tasks. It mainly aims at the maintenance tasks with series-parallel hybrid relationship or without any correlation in the system.

The third is more people share a task. It mainly aims at the maintenance task of artillery command information system network communication.

The fourth is assigned according to the temporal relationship. It is mainly aimed at tasks with series relationship or urgent need of tasks in the system.

2.3 Maintenance Force Dispatching Principle

The following principles must be followed in dispatching the maintenance force of artillery command information system equipment [2].

First, consider the importance of equipment. Priority should be given to repairing equipment of great importance and to repairing equipment of minor importance.

Second, consider the timeliness of the task. Only when the maintenance task is completed in an effective time can play its due role in combat.

Third, consider the differences of personnel. Reasonable allocation should be made according to the ability and expertise of maintenance personnel.

3 Maintenance Force Dispatch Model

3.1 Force Dispatch Model of One Man Executing a Task

Set n maintenance personnel to complete n maintenance tasks. Each person undertakes one task. The time that maintenance personnel i to complete task j is t_{ij} $(i = 1, 2, \ldots n; \quad j = 1, 2, \ldots, n)$. The objective function is to minimize the time required to repair the whole artillery command information system, then the model is:

$$\min\ y = \sum_{j=1}^{n}\sum_{i=1}^{n} x_{ij}t_{ij} \quad \text{s.t.} \begin{cases} \sum\limits_{i=1}^{n} x_{ij} = 1, & j = 1, 2, \ldots, n \\ \sum\limits_{j=1}^{n} x_{ij} = 1, & i = 1, 2, \ldots, n \end{cases} \tag{1}$$

This problem is a typical balanced assignment problem [3]. The Hungarian algorithm can be used to solve the model, and the flowchart is shown in Fig. 1.

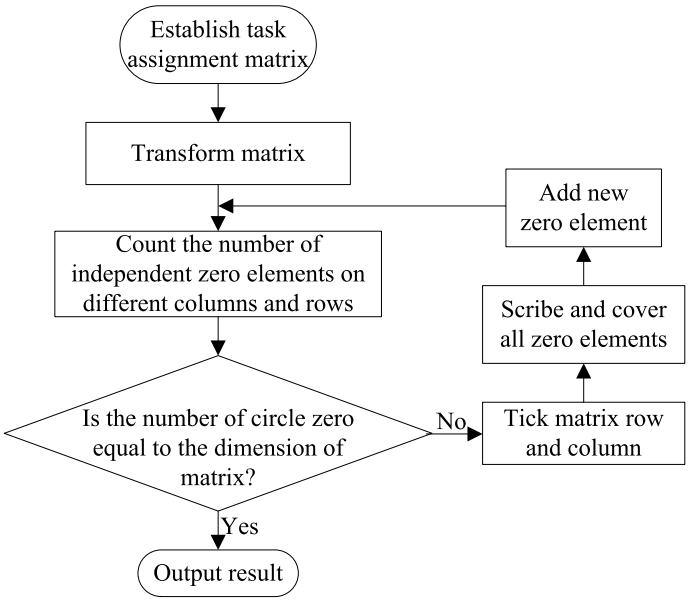


Fig. 1 Flow of Hungarian algorithm

3.2 Multi-task Force Scheduling Model for One Person

Assuming that m personnel complete n tasks and $n > m$ are established, it belongs to non-equilibrium assignment problem. Using Hungarian algorithm to solve the model may not be able to obtain the optimal solution [4]. In this paper, the Hungarian algorithm is improved by "one person into many people" and "adding edge to make up small" method. Set $n = km + i$, where: $k \geq 1, 0 < i < m$, then the force dispatching matrix is

$$Y' = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \\ \min t_{i1} & \min t_{i2} & \cdots & \min t_{in} \end{bmatrix}_{n \times n} \tag{2}$$

The force dispatch matrix is

$$Y = \left[t_{ij}^{'} \right]_{n \times n} \tag{3}$$

Then, the model of force dispatch is

$$\min y = \sum_{j=1}^{n} \sum_{i=1}^{n} x_{ij} t_{ij}^{'} \tag{4}$$

3.3 Force Dispatch Model of More People Executing a Task

Suppose that there are y persons in total to undertake x tasks that need to be completed by multiple persons, and the number of people needed for each task cannot be greater than y. Queuing theory method [5] can be used to solve the problem, the process is as follows:

Step 1 Determine maintenance sequence. The maintenance order is determined according to the maintenance task relationship, equipment status role and operational task requirements determine the order.

Step 2 Maintenance capability ranking. According to the skill level of maintenance support personnel, they are arranged in order from high to low, construct the ability level set of support personnel.

Step 3 Dispatch maintenance force. The maintenance force is determined according to the ability of maintenance personnel, the time sequence of maintenance tasks and the number of maintenance tasks required.

3.4 Power Dispatching Model Based on Time Series

For this situation, the number queue should be made according to the role of equipment in the system, the contribution to combat and the logical relationship of equipment construction, giving priority to the execution of personnel with strong ability and high efficiency.

4 Examples Analysis

In a training of artillery command information system, through condition monitoring and life prediction, it was found that many equipments and communication networking system had faults. The items to be completed are server (A_1), reinforcement computer (A_2), portable command terminal (A_3), communication network control machine (A_4), ultra-short wave radio (A_5), emergency communication radio (A_6), short wave radio (A_7), multi-function network security protection device (A_8), network switch (A_9), router (A_{10}) and high-speed wired data transmission equipment (A_{11}). The information link between the first company command vehicle and the battalion information machine t (A_{12}), the communication channel between the battalion command vehicle and the second company command vehicle t (A_{13}), the target information cannot collected by the third company laser instrument t (A_{14}), and the voice communication between the company command vehicle and the artillery terminal are intermittent t(A_{15}). Where A_1–A_{11} belong to one-person multitasking, A_{12}–A_{15} belong to multi-person one-thing. There are two tasks in these projects which contain time series association classes. One is that the server should be repaired before the reinforced computer, that is, the priority A_1 is higher than A_2. The other is that the emergency communication ratio should be repaired before the ultra-short wave ratio and the short wave ratio, that is, the priority A_6 is higher than A_5 and A_7. A total of five personnel participated in maintenance. As shown in Tables 1 and 2.

According to the previous analysis, the task A_{12} should rank first in the whole maintenance sequence, and the task A_{13}, A_{14} and A_{15} should rank behind the one-man multi-task maintenance task (A_1–A_{11}). In the dispatch of one-personnel multi-task,

Table 1 Time required to complete a multitasking task

Person	Time (h)										
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
B_1	4.6	3.8	4.3	3.4	2.3	3.5	1.6	1.6	1.0	2.5	3.0
B_2	5.2	1.7	2.1	4.4	5.7	1.3	1.1	3.6	4.3	3.1	4.5
B_3	3.3	1.4	2.6	2.8	4.1	2.6	2.5	4.6	3.6	4.7	5.8
B_4	2.5	3.9	3.0	1.5	1.7	3.3	2.2	2.6	3.1	5.8	2.9
B_5	5.5	4.0	5.0	4.9	3.6	4.1	3.1	3.2	2.9	1.7	3.7

Table 2 Number and time required for multi-person one-thing tasks

Task	Number	Time (h)
A_{12}	5	2.7
A_{13}	4	2.0
A_{14}	3	1.4
A_{15}	2	0.9

there is a fixed sequence of A_1, A_2, A_5, A_6 and A_7, and they are numbered A_{1a_1}, A_{2a_2}, A_{6b_1}, A_{5b_2}, A_{7b_2}.

1. One person performs multi-task force dispatch
   Get $x_{1,8} = x_{1,9} = 1$, $x_{2,6b_1} = x_{2,7b_2} = 1$, $x_{3,2a_2} = x_{3,3} = 1$, $x_{1,4a_1} = x_{4,4} = x_{4,5b_2} = 1$, $x_{5,10} = x_{5,11} = 1$ according to Formulas (2)–(4). That is arrange person B_1 to complete task A_8 and A_9, B_2 complete A_{6b_1} and A_{7b_2}, B_3 complete A_{2a_2} and A_3, B_4 complete A_{1a_1}, A_4 and A_{5b_2}, B_5 complete A_{10} and A_{11}.
2. More person perform one task force dispatch
   According to the conditions, we should arrange these five people to complete A_{12} together at first. B_1 and B_2 perform A_{15} after 5.3 h, used 0.9 h, B_1, B_2 and B_3 perform A_{14} after 6.7 h, used 1.4 h, B_1, B_2, B_3 and B_5 perform A_{13} after 8.1 h, used 2 h.
3. Dispatch forces according to time sequence
   At first, B_1, B_2, B_3, B_4 and B_5 complete A_{12} together. Secondly, B_1 complete A_8 and A_9. B_2 complete A_{6b_1} and A_{7b_2}. B_3 complete A_{2a_2} and A_3. B_4 complete A_{1a_1}, A_4 and A_{5b_2}. At last, B_1 and B_2 complete A_{15}. B_1, B_2 and B_3 complete A_{14}, B_1, B_2, B_3 and B_5 complete A_{13}.
4. Maintenance force dispatch optimization scheme
   To sum up, the maintenance force dispatch optimization scheme can be obtained, as shown in Fig. 2. The artillery command information system can restore the overall combat effectiveness after 10.1 h.

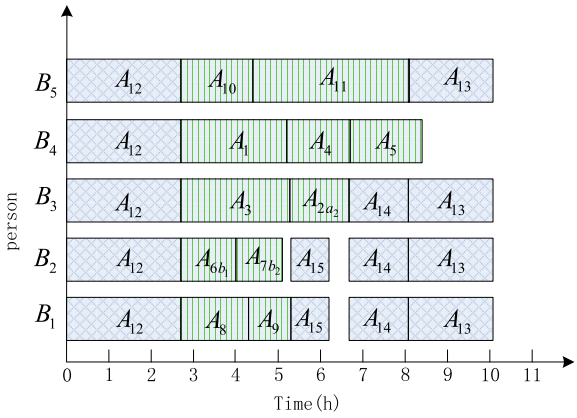


Fig. 2 Gantt chart of maintenance force dispatch

5 Conclusion

Based on the assignment problem, this paper takes the shortest time needed to repair the whole system as the optimization goal. The maintenance force dispatch optimization model is simple and easy to implement, which can provide reference for the maintenance force scheduling of the army.

References

1. Zhu, A.G., Ma, P.F.: Technical Service of Command Information System. PLA Press, Beijing (2017)
2. Chen, Z.: Research on the Optimal Allocation of Maintenance Personnel for Maintenance Tasks. China Civil Aviation University (2018)
3. Lu, L., Cheng, S.P.: Cloud computing energy consumption optimization based on Hungary method in a package-gluster framework. Comput. Appl. Softw. **35**(9), 85–91 (2018)
4. Yang, F., Li, H., Hu, Y.N.: Hungarian algorithm for teaching task assignment problem. China Edu. Technol. Equip. **14**(7), 12–15 (2017)
5. Zhong, L.M.T., et al.: Application of Hungarian algorithm in assignment problem. W. Leather **17**, 96 (2018)

Application of an Improved Genetic Algorithm in Tourism Route Planning

Xiaoyan Chen, Kun Zhang, Haifeng Wang, Jing Chen, and Bhatti Uzair Aslam

Abstract In this paper, after determining the needs of the users, the mathematical model of travel planning is established, and the time and cost constraints are studied. The traveling salesman problem (TSP), from the dimension of time and cost, is defined as the objective function. The basic genetic algorithm is studied and is put forward an improved genetic algorithm. The simulation result shows that the convergence speed of the improved genetic algorithm is better than genetic algorithm (GA). Finally, the application of tourism planning path in the field of wireless network and big data emerging technology and its good meaning are shown.

1 Introduction

At present, the supply of traditional tourism products exceeds that of new tourism products; the supply of tourism group products are surplus, but the supply of leisure, outdoor exploration, human tourism and other tourism products are insufficient. In view of the current situation, the tourism industry urgently needs to promote the supply side reform [1]. Most of the existing tourism services are still based on the tourism products provided by traditional tourism enterprises and institutions such as travel agencies [2–4].

In this paper, with the help of the travel agent model, the improved genetic algorithm is used to solve the travel planning problem, and the artificial intelligence algorithm is used to realize more effective route planning [5].

X. Chen · H. Wang · J. Chen
College of Computer Science and Technology, Hainan Tropical Ocean University, 572022 Sanya, China
e-mail: Xiaoyan0205@126.com

K. Zhang (✉)
Education Center of MTA Hainan Tropical Ocean University, Sanya, China
e-mail: kunzhang@hntou.edu.cn

B. U. Aslam
College of Computer Science and Information Technology, Nanjing Normal University, 210023 Nanjing, Jiangsu, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_17

2 TSP Problem Description and Solution Model

The traveling agent problem is that the traveler should travel through a given number of cities, provided that these cities only allow the traveler to visit once, and ask to find the shortest path after the traveler has traveled through each city. Generally, there are two kinds of algorithms to solve the TSP problem: precise algorithm and heuristic algorithm. There are mainly bound methods, planning method and so on. Heuristic algorithms include genetic algorithm, simulated annealing, particle swarm optimization, ant colony optimization, etc. Among them, the genetic algorithm has the idea of evolution, which provides a new way to solve the traveling salesman problem [6].

The TSP problem requires businessmen to visit all n cities in the map and finally return to the starting point. The goal is to select the shortest path to complete the task. In graph theory, TSP is to find a Hamilton loop, and the path length of the Harman loop is the shortest.

TSP is described in mathematical language as looking for a circuit path and satisfying the following objective function:

$$f(c) = min \sum_{i=1}^{m-1} d(c_i, c_{i+1}) + d(c_m, c_1) \tag{1}$$

In this function, c_i is the city number, $c_i \in M$, $1 \leq c_i \leq m$, $d(c_i, c_j)$ representing the weight between city i and city j; if the TSP is symmetric, there is $d(c_i, c_j) = d(c_j, c_i)$.

Let $G = (C, S)$ is the weighted complete graph, $C \in M$ represents the vertex set, S represents the side set.

3 TSP Calculation Complexity

If $f{:}S \rightarrow R$, S represents a finite set, f represents mapping functions, each $x \in S$ gets the only one real number $f(x)$. The corresponding maximum (minimum) problem can also find the corresponding unique $x' \in S$, so that for any other $x \in S$, $f(x') \geq (\leq) f(x)$.

From the theoretical point of view, any finite combinatorial optimization problem can find the best solution. But suppose the city size is m TSP, then the possible number of circuits are $\frac{(m-1)!}{2}$; each circuit needs the sum of n distances. The amount of calculations will be proportional to $\frac{m!}{2}$. So when m grows to a certain scale, this leads to a "combination explosion."

Supercomputer is used to carry out the simulation experiments, and the obtained results show the high complexity of the TSP time calculation. Therefore, a good algorithm needs to be used in route planning; the genetic algorithm is a good choice.

4 Genetic Algorithm

Genetic algorithm is a good choice. It is an efficient optimization method to find the global optimal solution without any initialization information [7, 8]. The solution set of the problem is regarded as a population. Through the iterative operation of continuous selection, crossover and mutation, the quality of the solution becomes better and better until it converges to the local or global optimal solution. Compared with other algorithms, this algorithm has the characteristics of multi-point search, parallel computing, strong adaptability, no specific restrictions on the problem, and easy to combine with other algorithms. Genetic algorithm is widely used in function optimization, combination optimization, automatic control, system identification, image processing, machine learning and data mining [9].

4.1 Basic Steps of Genetic Algorithm

Genetic algorithm (GA) mainly uses the data of the gene string structure in genetic space to replace the data of the original solution space before iteration, and then combines the data of the gene string structure to form different points [10].

4.2 Formation of Initial Population

First, randomly generate NIND (random generation proper noun) string data of the initial structure, and then let each string form different individuals, so that NIND individuals can form an initial population. Finally, the initial value of the genetic algebra gene and the number of times of maximum iteration max generation are set; the initial value is set to 0.

4.3 Fitness Evaluation

The value of fitness is the interpretation of the advantages and disadvantages of chromosomes or solution space and is the only basis for natural selection of chromosomes. It can be transformed by the objective function. Different solutions take different methods of fitness function. In general, the smaller the value of the function that conforms to the individual, the larger the fitness value corresponding to it, which indicates better the individual of the required solution. So we can take the fitness function:

$$F[f(x)] = \frac{1}{f(x)} \tag{2}$$

4.4 Selection Operation

The main purpose of the selection is to select an excellent individual from the group, so that this excellent individual can be used as a parent to reproduce the next generation of children. The possibility of individual selection is mainly related to the size of the corresponding fitness of this individual. The greater the value of fitness, the greater the probability of being selected. The selection operation of genetic algorithm mainly adopts roulette method, that is, proportion selection strategy based on fitness value.

4.5 Crossover Operation

The crossover operator is applied to the group according to the probability, so that the excellent part of the two parent individuals can be exchanged and recombined to generate a new and better individual operation.

4.6 Variation Operation

Mainly maintaining population diversity.

4.7 Judgment of Termination Conditions

If the number of iterations reaches the maximum, that is, Gen (generation) = MAXGEN (max generation), and the global or local optimal solution can be output, the calculation will be terminated; if the above requirements cannot be met, go to step 4.3.

5 An Improved Genetic Algorithm

The basic genetic algorithm has the disadvantages of slow convergence and easy to fall into the local optimal solution. This paper optimizes the initialization, crossover and mutation operations of the genetic algorithm and proposes an improved genetic algorithm (IGA).

5.1 Initialize the Population

Most of the chromosomes involved in the genetic algorithm are derived from the initial population. The basic genetic algorithm randomly generates the initial population, the individual's adaptive ability is biased, and the algorithm has poor convergence. In this paper, a two-way greedy selection strategy is used to initialize the population. Compared with the general one-way greedy selection, this method can include more optimal sub-paths and improve the fitness of the population. The basic idea is as follows:

Step 1: Randomly generate a city number C_n, and then select the city C_m closest to the city C_n to form a path $C_m - C_n$.
Step 2: Find the two cities closest to the two ends of the path (set points as C_a, C_b, $a, b \in C$), assuming C_s, C_t (C_s, C_t are not selected), and if $C_a\ C_s < C_s\ C_b$, select C_s, constitute the path $C_s - C_a - C_b$, otherwise select C_t, and constitute the path $C_t - C_b - C_a$;
Step 3: Return to step 2 until all cities are in the path;
Step 4: If all cities are already in the path and have not reached the population size, then randomly select two cities to form the path, and then return to step 2 until the population size is reached.

5.2 Fitness Function

In genetic algorithm, fitness function plays a very important role, mainly to judge the quality of individuals. The trip planning model is to minimize the objective function, and the larger the individual fitness value is, the better the individual performance is. Here, the reciprocal of the objective function is taken as the fitness function.

$$f(i) = \frac{1}{\phi(\omega)} \tag{3}$$

$\phi(\omega)$ is objective function, which can be obtained from the following formula:

$$\phi(\omega) = min\left(T + t\frac{1}{c_{max} - c}\right) \tag{4}$$

T travel time of the line;
c travel expenses of the line;
c_{max} the budget of the highest expenses at the user's disposal;
t it is used to measure time (unit: hour) the weight coefficient of the relationship with money (unit: yuan), which is set by tourists.

5.3 Cross Operation

Cross operation can generate new individuals, set low cross probability for individuals with high fitness value, and retain their excellent performance; on the contrary, appropriately increasing the cross probability for individuals with low fitness value is conducive to the continuous optimization of individuals.

$$P_{ci} = P_{cmax} - \frac{(P_{cmax} - P_{cmin})(f_i - f_{\text{ave}})}{f_{max} - f_{\text{ave}}} \tag{5}$$

P_{ci} is the probability of mutation of individual I, f_{ave} is the average fitness value of the group, f_{max} is the maximum fitness value of the current individual, $P_{cmin} = 0.5$.

5.4 Mutation Operation

Mutation operation is conducive to maintaining population diversity. The smaller the individual fitness value is, the lower the possibility of variation is. When the number of iterations increases, the similarity of genes tends to be the same. Increasing the probability of individual variation is conducive to the generation of new individuals and avoiding the local optimal solution

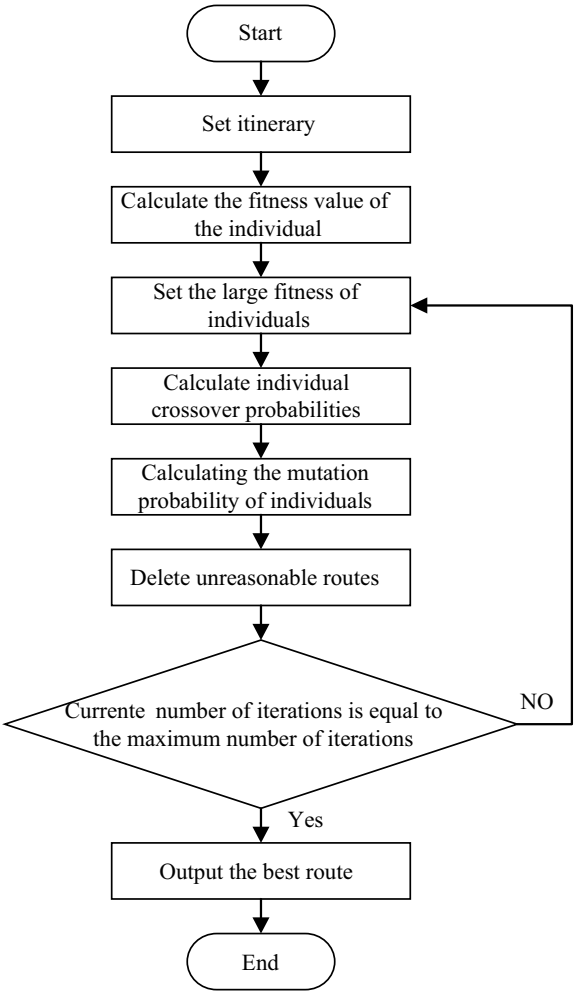
$$P_{mi} = P_{mmin} + \frac{(P_{mmax} - P_{mmin})(f_i - f_{\text{ave}})}{f_{max} - f_{\text{ave}}} \tag{6}$$

P_{mi} is the probability of mutation in individual I, f_{max} is the maximum fitness value of the current individual, f_{ave} is the average fitness value of the group, $P_{mmax} = 0.05$.

5.5 Algorithm Flow

Step 1: Set itinerary according to the sequence of attractions in the route (Fig. 1).
Step 2: Take the inverse of the objective function as the fitness function and calculate the fitness value of the individual $f_{i.}$
Step 3: Use the proportional selection factor P to select individuals with large fitness from the old population to the new population, $P = f_i / \sum f_i$, the individual with the highest fitness will be selected;
Step 4: Calculate individual crossover probabilities P_{ci}, Perform crossover operations, select several parents to pair, and generate new individuals.
Step 5: Calculating the mutation probability of an individual P_{mi} to perform mutation operations to identify mutant individuals.
Step 6: Delete unreasonable routes.

Fig. 1 Algorithm flow



Step 7: Determine if the current number of iterations is equal to the maximum number of iterations. If they are not equal, go to step 3. If they are equal, continue to execute.
Step 8: Output the best route.

6 Simulation Results and Analysis

Set $P_c = 0.6$, $P_m = 0.1$, use GA and IGA to conduct simulation experiments to find the optimal route of the trip planning problem (Fig. 2).

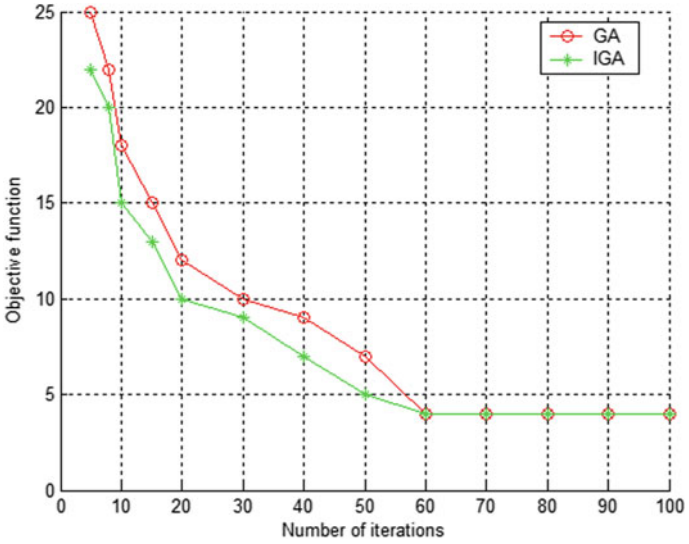


Fig. 2 Results of experimental simulation

Although the final solutions of the two algorithms are the same, the IGA algorithm finds the better solution before GA algorithm in the initial iteration, and the convergence speed is obviously better than GA algorithm, which has more advantages in practical application. Experimental simulation shows that the IGA algorithm used in the travel planning system can be better than the basic genetic algorithm.

7 Conclusion

The main characteristics of genetic algorithm are simple and general. Based on the analysis of the basic principles of genetic algorithm, this paper proposes an improved genetic algorithm and uses it to solve the problem of tourism routes. Simulation results show that this algorithm is relatively basic and it has more advantages in convergence.

Acknowledgements This work was financially supported by the 2019 Research Project on Education and Teaching Reform in Universities of Hainan Province (No. Hnjg2019-78), and the 2019 Teaching Reform Project of Hainan Tropical Ocean University (No.RHYJG2019-28, No. RHYJG2019-17); the 2018 Philosophy and Social Science Planning Project of Sanya (No. SYSK2018-06).

References

1. Hunter, W.C., Chung, N., Gretzel, U., et al.: Constructivist research in smart tourism, Asia Pacific. J. Inf. Syst. **25**(1), 105–120 (2015)
2. Zhang, K. Shen, C. Zhou, Q., etc.: A combined GPS UWB and MARG locationing algorithm for indoor and outdoor mixed scenario. Cluster Comput. J. Netw. Softw. Tools Appl. **22**(3), 5965–5974 (2019)
3. Zhang, K., Shen, C., Gao, Q., etc.: Ultra wideband positioning technology for accident ships under adverse sea condition. J. Coastal Res. **83**(1), 902–907 (2018)
4. Zhang, K., Shen, C., Gao, Q., etc.: Precise positioning system of ship interior based on UBW ultra wideband technology. J. Coastal Res. **83**(1), 908–912 (2018)
5. Shen, C., Wang, C.X., Zhang, K., etc.: A time difference of arrival/angle of arrival fusion algorithm with steepest descent algorithm for indoor non-line-of-sight locationing. Int. J. Distrib. Sens. Netw. **15**(6), 1–8 (2019)
6. Zhang, K., Shen, C., Gao, Q., etc.: Design of intelligent medical service robot drug delivery system based on UWB precise indoor positioning technology. Basic Clin. Pharmacol. Toxicol. **123**(1), 12–13 (2018)
7. Zhang, K., Shen, C., Wang, H.F., etc.: Cluster computing data mining based on massive intrusion interference constraints in hybrid networks. Cluster Comput. J. Netw. Softw. Tools Appl. **22**(3), 7481–7489 (2019)
8. Zhang, J.W., Xu, M.J., Zhao, M.X., et al.: Design of tourism route based on genetic algorithm and analytic hierarchy process. China Bus. **7**, 79–81 (2015)
9. Zhang, K., Shen, C., Huang, M.X., etc.: Interrupt protection control of anti-interference nodes in network based on band sampling decision filter modulation. Cluster Comput. J. Netw. Softw. Tools Appl. **22**(3), 7569–7576 (2019)
10. Zhang, K., Shen, C., Li, H.W., etc.: Direction of arrival estimation and robust adaptive beamforming with unfolded augmented coprime array. IEEE Access **8**(1), 22314–22323 (2020)

Classification of Medical Symptoms Based on Social Sensors

Yong Li, Si-xuan Chen, Xia Wang, Hai Jia, and Yan-bin Shi

Abstract Due to the high degree of classification of the current hospital departments, it is difficult for patients to accurately select hospital departments for registration based on their limited medical knowledge and physical illness status. With the development of technologies such as the Internet, big data, and artificial intelligence, we have ideas and methods to solve this problem. This article treats users on the Internet as social sensors. Users spontaneously generate massive medical data on medical Web sites, medical symptom libraries, and medical professional e-books. These data provide a way to solve the lack of medical knowledge when patients seek medical treatment. Due to the sparsity and lack of information in the short text data of the symptoms, the LDA-Naïve Bayes (L-NB) model is used to expand the features and classify the text of "symptoms–departments". Through a lot of experiments, the algorithm proposed in this paper has high accuracy and can meet the needs of practical applications.

1 Introduction

With the advent of the era of big data, online booking registration systems have been widely used. We know that registration is the first step for patients to seek medical treatment, but it is difficult for patients to accurately judge the complex symptoms of their own bodies without medical background knowledge. In order to solve the above problems, save patients' time and improve the efficiency of medical treatment, we first extract massive medical symptom data from medical Web sites, medical

Y. Li · S. Chen (✉)
College of Computer Science and Engineering, Northwest Normal University, Lanzhou, China
e-mail: 153115981@qq.com

X. Wang · H. Jia
Department of Pharmacy, Gansu Provincial Hospital, Lanzhou, China

Y. Shi
School of Pharmacy, Lanzhou University, Lanzhou, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_18

symptom libraries, and medical professional e-books. Then, we train the symptom data through L-NB algorithm, classify and predict according to the probability model. When the patient is registered, they can enter some keywords for the symptoms of physical illness. According to a certain probability, we can provide the patient with the recommended ranking for the corresponding department of the hospital.

The data of medical symptoms are typical short text data, which are characterized by short length, sparsity, noise, and ambiguity. Therefore, the general text classification algorithm does not work well on short text classification problems [1]. In order to overcome the shortcomings such as feature sparseness, we use the L-NB model to classify the symptom texts [2].

2 Related Work

As far as this article is concerned, there is currently no research work in the academic community to predict specific diseases based on limited symptom keywords. Most of the existing studies have targeted specific diseases and classified their symptoms to help doctor's diagnosis and treatment. Medeiros et al. [3] continuously monitored the motor symptoms caused by Parkinson's disease, and used the support vector machine (SVM) to identify the symptoms of Parkinson's disease retardation. Mohamed et al. [4] combined convolutional neural networks (CNN) and question-and-answer models to classify skin symptoms for better decision making.

The focus of this paper is on the "symptoms–departments" classification. The symptom text is vectorized using a bag of words [5], and then classified and predicted based on a machine learning algorithm. Since the symptom text in the data used in this article is short and the amount of data per text is small, the symptom text is extended by the short text understanding model before classification. The short text understanding model include latent semantic analysis (LSA), neural network language model (NNLM), probabilistic latent semantic analysis (PLSA), latent Dirichlet analysis (LDA), and explicit semantic analysis (ESA).

In most text classification tasks, text is treated as a word package for feature extraction and word frequency statistics. The statistical method ignores the semantics of the text. Machine learning algorithms commonly used for text classification tasks include *K*-nearest neighbor (KNN), decision trees, Naïve Bayes [6], SVM, etc.

Based on the comparative analysis and the data characteristics of this paper, L-NB model was used to classify the data.

3 Theory and Method

3.1 LDA Topic Model

The LDA model assumes that a document is a collection of words, and there is no order relationship between words and words in the document. The probability generation process of the relationship between text and words is represented in Fig. 1.

Suppose $W = (w_1, w_2, \ldots w_n)$ represents a set of words, $D = (d_1, d_2, \ldots d_m)$ represents a set of texts, and $Z = (z_1, z_2, \ldots z_k)$ represents a set of topics. Then, Fig. 1 can be represented by Formula (1):

$$P(w_j|d_i) = \sum_{\substack{0 < i < m \\ 0 < j < n}} \sum_{k=1}^{K} P\left(w_j|z_k\right) P(z_k|d_i) \tag{1}$$

The LDA model generates θ by a Dirichlet distribution with a parameter of η and obtains ψ by a Dirichlet distribution with a parameter of ε. "Text-theme" obeys the multi-distribution of θ and "theme-word" obeys the multi-distribution of ψ. The parameter $Z_{m,n}$ represents the subject corresponding to the nth word in the mth document, and $W_{m,n}$ represents the word corresponding to the subject k. The formula for calculating the probability is shown in Eq. (2). The LDA model is shown in Fig. 2.

$$P(\theta, Z, W|\eta, \varepsilon) = P(\theta|\eta) \prod_{n=1}^{N} P(Z_n|\theta) \cdot P(W_n|Z_n, \varepsilon) \tag{2}$$

The basic problem in the LDA model is to define the parameters η and ε.

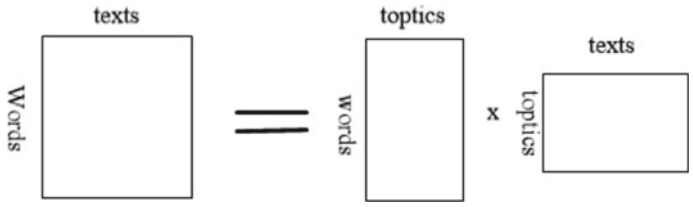


Fig. 1 Simulation process of implicit topics

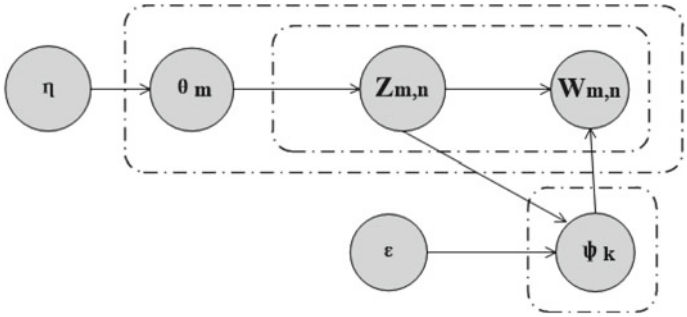


Fig. 2 LDA model

3.2 Naïve Bayes

Naïve Bayes is a highly practical machine learning algorithm whose performance is comparable to algorithms such as decision trees and neural networks [7]. Assuming the training set is n dimensions of m samples, the data set is expressed as follows:

$$\left(x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}, y_1\right), \left(x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}, y_2\right), \ldots, \left(x_1^{(m)}, x_2^{(m)}, \ldots, x_n^{(m)}, y_m\right)$$

Suppose there are a total of k types, such as $C_1, C_2, \ldots, C_k$, and each sample is $m_1, m_2, \ldots, m_k$. The calculation process is as follows:

Step 1. Get the prior probabilities of k types from the sample showing in Eq. (3):

$$P(Y = C_k) = (m_k + \lambda)/(m + k\lambda) \tag{3}$$

Step 2. Calculate the conditional probability distribution $P(X = x|Y = C_k)$ of the symptom sample x under k categories of conditions. Since the symptom words are discrete, therefore, the probability distribution is shown in Eq. (4):

$$P(X = x|Y = C_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n|Y = C_k) \tag{4}$$

Step 3. According to symptom sample $X^{(\text{text})}$, find the posterior probability $P(Y = C_k|X = X^{(\text{text})})$ is shown in Eq. (5):

$$P\left(Y = C_k|X = x_i^{(\text{test})}\right) = \frac{P(Y = C_k) \cdot \prod_i^n P(X_i = x_i^{(\text{test})}|Y = C_k)}{P\left(X^{(\text{text})}\right)} \tag{5}$$

Step 4. The classification result of test sample $X^{(\text{text})}$ is shown in Eq. (6):

$$C_{\text{result}} = \arg\max P\left(Y = C_k|X = x_i^{(\text{test})}\right) \tag{6}$$

3.3 Symptom Classification Model

3.3.1 Framework

The symptom classification model is constructed based on the network text information mining technology, as shown in Fig. 3.

3.3.2 Obtaining Data

In this paper, we use reptile technology to collect data on medical Web sites, obtaining data from medical symptom database and medical classic textbook. At the same time, we use ICD-10 to contribute symptom dictionary, which is imported into the Jieba, to get more accurately of word segmentation.

3.3.3 Feature Extension

Since the content of each symptom text is short and the feature matrix is sparse, before classification the data is trained using the L-NB to obtain the distribution matrix of "theme-word". Then, the theme obtained from the training data is used. Finally, the feature words under the corresponding theme are added to the original data of the category as a new feature vector [8].

3.3.4 Training Classifier

The training process is detailed in Algorithm 1.

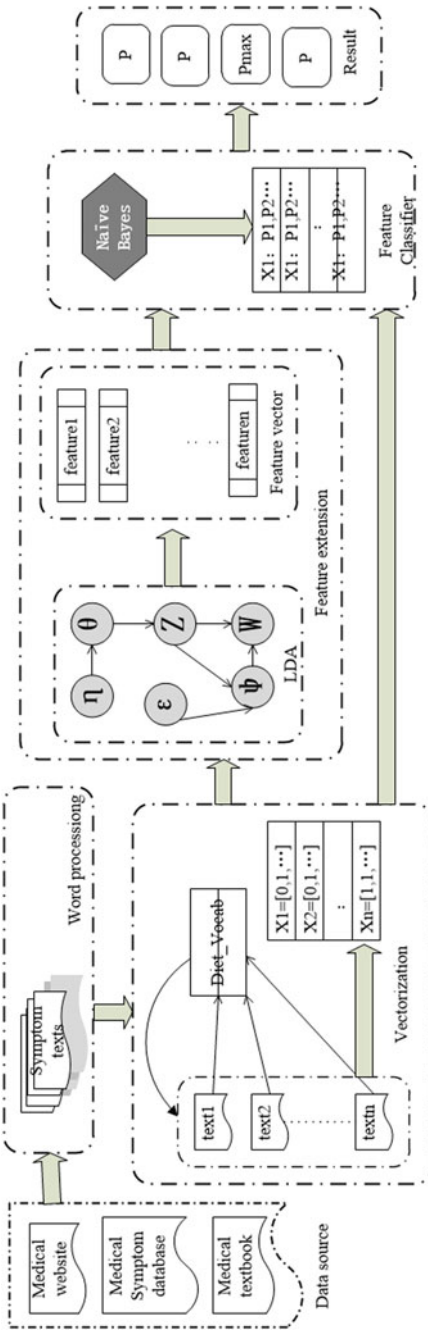


Fig. 3 Symptoms–department classification model

***Algorithm 1**: classification model*

Input: text matrix $M_{m\times n}$, set $V_a = \{(vec_i, Y_j) | 1 \le i \le n, 1 \le j \le 4\}$, set $V_b = vec_p, p \in [1, n]$

Output: classifier model, test set results Y_j.

1. Set parameter k ∈{1,2,⋯,10}, z for topic, d for text, w for words, q for number of words.

2. Derived the subject of each word by Gibbs sampling.

3. for i∈{1,2,⋯,n} do

4. Gibbs sampling=> $P(z_k|d_i)$

5. end for

6. Derive hyperparameters.

7. for i∈{1,2,⋯,n} do

8. by $P(z_k|d_i)$ and formula (5) =>ε,η

9. end for

10. Find the distribution of symptom word characteristics under each topic.

11. for i∈{1,2,⋯,n} do

12. After obtaining ε~θ, η~ψ, training data

13. end for

14. Train the model with feature data and training sets.

15. for i ∈ {1,2, ... , V_a} do

16. Calculate the probability distribution

17. end for

18. Add test set data for classification

19. for i ∈ {1,2, ... , V_b} do

20. Put V_i into the Model in turn, calculate arg max P($Y_j|V_i$)

21. end for

22. return Y_j

4 Empirical Analysis

4.1 Dataset

The data in this article comes from the medical search network, medical symptom library, and medical books. The experiment was mainly divided into four departments of obstetrics and gynecology, internal medicine, surgery, and ear nose throat branch (ENT). The experiment selected 70% of the data in the symptom text to train the L-NB as the training set data, and 30% as the test set data.

4.2 Experimental Evaluation

For the final classification, the classification accuracy rate is calculated as:

$$\text{accuracy} = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{\text{true}_i}{\text{Count}_i} \tag{7}$$

where true_i indicates the correct number of predictions and Count_i indicates the number of all test texts in the department.

4.3 Experimental Results and Analysis

The number of L-NB parameter is important to this paper, we use Gibbs sampling [9] to determine. Finally, the number of topics is set to 10. The result is shown in Fig. 4.

The L-NB classifier is constructed after the feature expansion is completed. In the algorithm, the word frequency of each word is first calculated, the word frequency is arranged in descending order, and then the first N invalid words are eliminated. The accuracy rate is shown in Fig. 5.

According to the data in Table 1, the accuracy of surgical symptom testing is almost unaffected by the model. The distinction between surgery and other departmental symptoms is higher and the features are more obvious. From the experimental comparison, the L-NB model proposed in this paper has good classification accuracy.

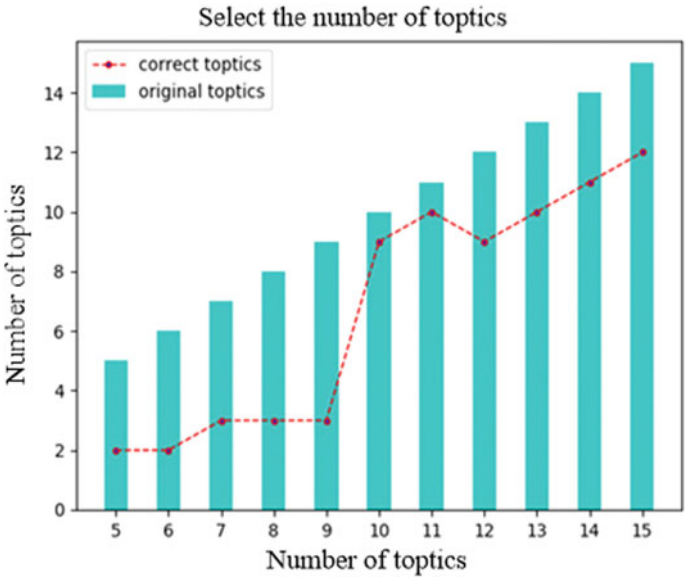


Fig. 4 Effect of theme selection

Fig. 5 Parameter *N* and classification accuracy

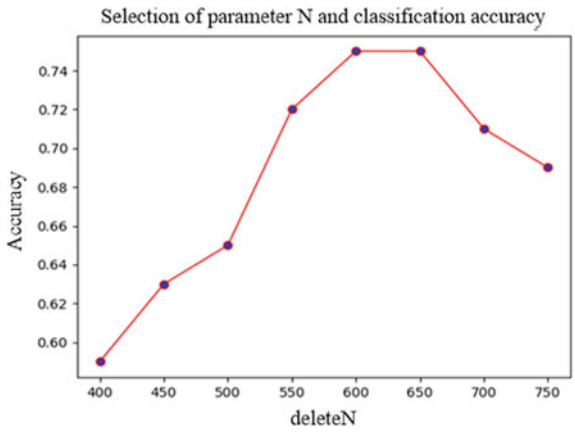


Table 1 Comparison of accuracy

	Naïve Bayes (%)	KNN (%)	SVM (%)	L-NB (%)
ENT	60.00	70.00	70.00	80.00
Internal	80.00	70.00	70.00	70.00
Surgical	100.00	90.00	95.00	100.00
Maternity	70.00	70.00	75.00	80.00
Total	77.75	75.00	77.50	82.50

5 Conclusion

The patient's medical background is lacking, and it is difficult to accurately judge the disease according to his or her symptoms and select the corresponding department to see a doctor. This paper studies the classification relationship between symptoms and departments based on massive medical symptom data extracted from medical Web sites, medical symptom libraries, and medical professional e-books.

In the experiments, not only the accuracy of the L-NB model and the comparison model in the whole test set are given, but also the accuracy comparison of the departments. The "symptom–department" relationship which is finally established in this paper establishing a patient online registration guidance system.

In the future research, we need to build a Probase similar to Microsoft Research Asia, which is a knowledge base based on word probabilities [10]. The relationship between the disease entity and the symptom entity is reflected by the entity relationship [11] in the knowledge base, and then the probability model is used for decision making to achieve the symptom-to-disease correspondence.

Acknowledgements This research was partially supported by the grants from the Natural Science Foundation of China (No.71764025, 61863032, 61662070); the Scientific Research Foundation of the Gansu Provincial Administration of Traditional Chinese Medicine (Grant No. GZK-2019-40); the Research Project on Educational Science Planning of Gansu Province, China (Grant No.GS [2018] GHBBKZ021 and GS [2018] GHBBKW007); and the Scientific Research Foundation of the Higher Education Department of Gansu Province, China (Grant No.2018A-001). Author contributions: Li Yong and Chen Si-Xuan are co-first authors who jointly designed the research. Wang Xia and Jia Hai are co-corresponding authors.

References

1. Wang, Z.Y., Cheng, J.P., Wang, H.X., Wen, J.R.: Short text understanding: a survey. J. Comput. Res. Dev. **53**(02), 262–269 (2016)
2. Song, Y., Wang, H., Wang, Z., Li, H., Chen, W.: Short text conceptualization using a probabilistic knowledgebase. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, pp. 2330–2336.Barcelona, Catalonia, Spain (2011)
3. Medeiros, L., Almeida, H., Dias, L., Perkusich, M., Fischer, R.: a game-based approach to monitor Parkinson's disease: the bradykinesia symptom classification. In: 2016 IEEE 29th International Symposium on Computer-Based Medical Systems, pp. 337–342. Belfast and Dublin, Ireland (2016)
4. Akrout, M., Farahmand, A.M., Jarmain, T., Abid, L.: Improving skin condition classification with a visual symptom checker trained using reinforcement learning. CoRR abs. arXiv:1903.03495v2 (2019)
5. Xing, S.: Research on Image Classification Algorithms Based on Word Bag Model and Transfer Learning. Zhejiang University, Zhejiang (2016)
6. Li, X., Yang, B.: A pseudo label based dataless naive bayes algorithm for text classification with seed words. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 1908–1917. Santa Fe, New-Mexico, USA (2018)
7. Shi, H.B., Wang, Z.H., Huang, H.K., Li, X.J.: A restricted double-level bayesian classification model. J. Softw. **15**(02), 193–199 (2015)

8. Efron, M., Organisciak, P., Fenlon, K.: Improving retrieval of short texts through document expansion. In: Special Interest Group on Information Retrieval 2012, pp. 911–920. Portland, Oregon, USA (2012)
9. Su, J.Y., Xu, J.C., Qiu, X.P., Huang, X.J.: Incorporating discriminator in sentence generation: a gibbs sampling method. In: 2018 Association for the Advancement of Artificial Intelligence, pp. 5496–5503. New Orleans, USA (2018)
10. Wu, W., Li, H., Wang, H., Zhu, K.Q.: Probase: a probabilistic taxonomy for text understanding. In: Proceedings of the 2012 ACM International Conference on Management of Data, pp. 481–492. Scottsdale, Arizona, USA (2012)
11. Zhang, C.Z., Li, Y.L., Du, N., Fan, W., Yu, P.S.: On the generative discovery of structured medical knowledge. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2720–2728. London, United Kingdom (2018)

Research on Method and Technology of Emergency Logistics Intelligent System Engineering

Liang Wu and Dong Xu

Abstract Focusing on the future emergency logistics support needs, this paper studies the evolution trend and application focus of system engineering methods and intelligent technologies in the field of emergency logistics, analyzes the application of various system engineering methods and key technologies in logistics modeling, simulation, analysis, prediction, evaluation, decision-making, management, control and planning design. Using intelligent system engineering methods to achieve accurate prediction, optimal planning, rapid decision-making, independent management and real-time control of emergency logistics, the utilization efficiency of emergency logistics resources is improved, and material security effectiveness is fully utilized. It provides systematic engineering basis and technical solutions for the realization of the leapfrog development of emergency logistics.

1 Introduction

At present, a new round of technological and industrial revolution is at the historical juncture of achieving a major breakthrough. With the rapid development of artificial intelligence and its accelerated transfer to the emergency field, the trend of systematization, engineering and intelligence of emergency logistics is increasingly prominent.

The essence of emergency logistics is resource integration. The core is to establish a rapid, accurate, reliable, safe and low-cost emergency material support system through the overall integration and multi-body combination of all links of emergency logistics. System science and system engineering are exactly the theory and method to study and solve the system problems such as "comprehensive integration." With the continuous development, progress and integration of system science and intelligent technology, the physical space of emergency logistics maps to the

L. Wu · D. Xu (✉)
Joint Logistics College, Beijing, China
e-mail: 165647015@qq.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_19

digital space through ubiquitous sensor, Internet of things and communication technology and accelerates the transition to the information space and cognitive space. Based on dynamic perception, data mining, simulation and operational optimization, the emergency logistics system simulates the operation situation in different environments. The output of the executable intelligent auxiliary results is fed back to the digital process, and the intelligent decision is mirrored to the physical world. Through the coordination of the physical domain, information domain and cognitive domain, the flow of material, information and command are highly coupled, and the value-added chain of data, information, knowledge and wisdom is accompanied, so as to realize the accurate prediction, optimal planning, rapid decision-making, independent management and real-time control, promote the whole process of resource reorganization, power reconstruction and process reengineering, such as planning, procurement, transportation, warehousing, distribution, settlement and recycling.

2 Intelligent Modeling and Simulation of Emergency Logistics

Simulation technology is a model-based experimental science, which is a powerful tool for analyzing, designing and optimizing complex systems. In the aspect of simulation and analysis of complex network model, in 1998, scholars found that regular network and random network could not reflect the characteristics of the real world, while small world and scale-free network had better applicability, and then complex network research began to rise. Many scholars use some abstract methods to build the real-world network model, and then analyze the parameters to study the characteristics of these complex networks. The purpose is to build a more realistic network model by mastering the characteristics and commonness of the network [1]. The cyber-physical systems (CPS) is a direction of intelligent logistics system modeling [2]. Based on the automatic mapping technology, real-time computing and massive information processing technology, dynamic network technology, self-organization and adaptive technology of virtual and real systems, it realizes the deep integration of logistics elements and information elements. The difficulty of building such models is to adapt to the characteristics of both discreteness and continuity. It cannot only explicitly represent the time-domain information of the physical system, but also explicitly represent the execution order of the information system.

Because the traditional simulation technology can not or is difficult to describe the characteristics of the randomness and complexity of logistics activities, an intelligent simulation model based on expert system is proposed in the field of emergency logistics, that is, in the stage of modeling, simulation model design, simulation result analysis and processing, using the knowledge base and reasoning machine of expert system, using expert instructions and experience for reasoning, judgment and decision-making, establishing a foundation [3]. The simulation closed loop is based on the intelligent feedback of expert system. This kind of intelligent simulation system starts from the scheme, carries on the modeling and the simulation experiment, obtains the conclusion through the computer operation, carries on the analysis

and compares with the anticipated goal. If the simulation results are consistent with the predetermined objectives, the original plan is feasible. If not, the existing problems will be analyzed through the expert system. According to the simulation results, the decision-making information will be automatically output with the help of the expert system, such as whether the system model needs to be modified and adjusted, the performance and quality of the system, the proposed countermeasures and explanations, and then the modified plan will be simulated again, until a feasible solution is found [4] (Fig. 1).

The research and application of intelligent modeling and simulation solve the problems in the process of modern emergency logistics modeling and simulation to a certain extent, improve the accuracy and analysis ability of emergency logistics simulation, help to shorten decision-making time, provide intelligent decision-making, and have profound theoretical and practical value for the research of emergency logistics intelligent.

3 Intelligent Analysis and Prediction of Emergency Logistics

The method of system analysis is to find the best scheme through the overall planning of the system. Mathematical programming is a method of operational research. The specific theories and methods include linear programming, dynamic programming, integer programming, queuing programming and inventory theory, which are used to solve the problems of location selection of logistics facilities, resource allocation of logistics operations, cargo loading, time and quantity of material storage, etc. Information intelligence analysis is often applied to fuzzy mathematics, neural network, analytic hierarchy process (AHP), principal component analysis (PCA), rough set, genetic algorithm and other intelligent algorithms, as well as statistical analysis system (SAS), statistical product and service solutions (SPSS), Excel, MATLAB and other data analysis and processing tools, while artificial intelligence and machine learning are the basis of big data analysis and its decision-making system. The analysis and processing technologies, such as massively parallel processor (MPP) database, Hadoop ecosystem based on MapReduce and so on, realize the massive logistics data search and screening, and mining the association relationship through the analysis of the overall sample.

System prediction methods mainly include qualitative prediction, time series analysis prediction, regression analysis prediction, state-space analysis prediction, Markov prediction and so on. Regression analysis prediction model, moving average prediction model, trend extrapolation prediction model, exponential smoothing prediction model and so on are often constructed. With the support of the new generation of information infrastructure, the physical space is rapidly mapping to the network space, and the accelerated integration of the human machine object three-dimensional world of emergency logistics system has led to the explosive growth

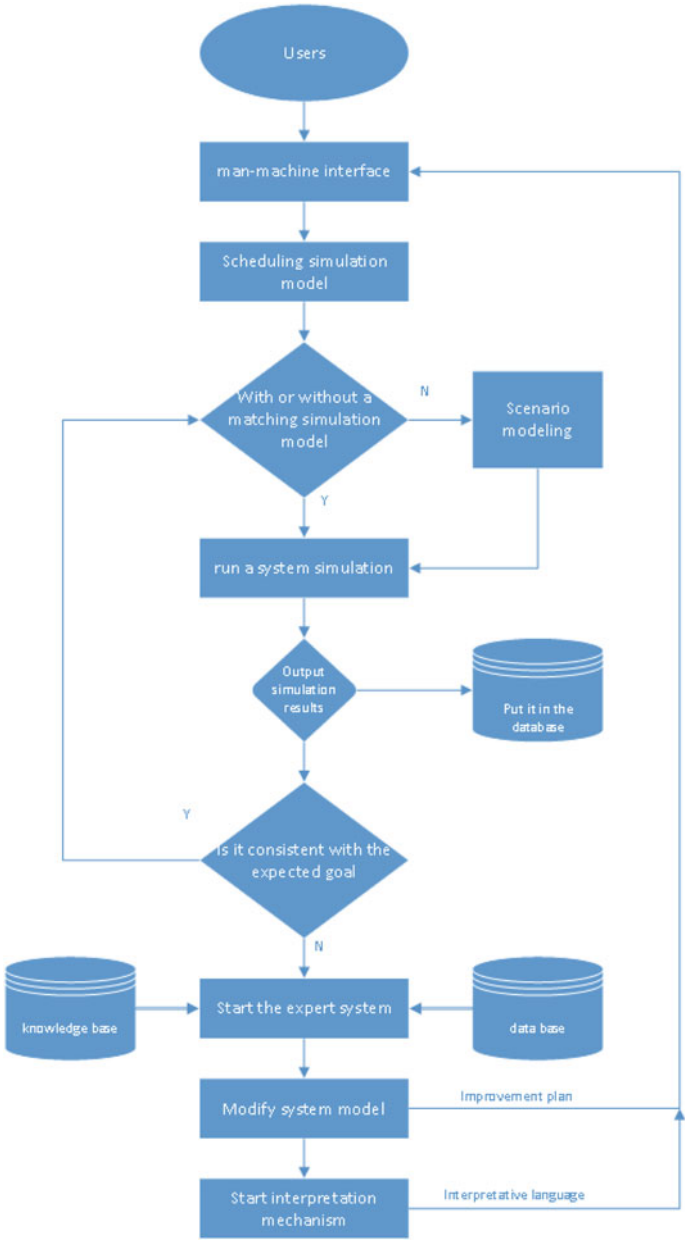


Fig. 1 Flowchart of intelligent simulation based on expert system

of data scale and the highly complex evolution of data model. According to the law of emergency material consumption, the intelligent system of emergency logistics predicts the demand of emergency materials, and ensures the simultaneous system of object demand acquisition and logistics link. The main development direction is the application of big data prediction analysis and real-time decision system based on artificial intelligence, deep learning and convolution neural network in the intelligent emergency logistics Internet of things and cloud platform.

An important development direction in this field is intelligent combination forecasting. All kinds of single forecasting methods contain some independent information of the forecasting system. By properly combining different methods and effectively using the useful information provided by different methods, we can more comprehensively describe the characteristics and laws of logistics forecasting problems, reduce one sidedness and randomness, and improve the forecasting accuracy.

4 Intelligent Evaluation and Decision of Emergency Logistics

Emergency logistics and supply chain system evaluation and decision-making are inseparable. Evaluation is the basis of decision-making, and some decision-making methods are also an evaluation method. The research in this field focuses on the classification, process, steps and methods of all kinds of evaluation methods and decision-making systems, including the evaluation and decision-making process of finding the optimal scheme in all stages and levels of logistics system planning, such as logistics network technical scheme, site selection scheme, facility master planning scheme and material handling system layout scheme. The commonly used evaluation methods of emergency logistics system include cost-benefit method, value analysis method, fuzzy comprehensive evaluation method, AHP, interval decision-making evaluation, etc., the commonly used models include index value standardization model, weighted average evaluation model, approaching ideal scheme evaluation model, fuzzy comprehensive evaluation model, fuzzy clustering evaluation model, gray correlation evaluation model and so on. Common emergency logistics system decision-making includes decision-making, risk decision-making, uncertainty decision-making, probability interval decision-making, complete information value and Bayesian decision-making, utility decision-making, competitive decision-making, multi-objective decision-making, etc. The existing evaluation and decision-making methods of emergency logistics system mainly adopt the multi-level and multi-index evaluation index system constructed by human, build the evaluation and decision-making model and adopt the qualitative and quantitative combination method to evaluate the supply chain reliability, logistics guarantee performance, emergency logistics system benefit, emergency logistics base layout, emergency logistics capacity building and other aspects.

The decision-making of emergency logistics support has the characteristics of large-scale, multi-objective and multi-constraint. These methods have their own limitations and shortcomings, so it is difficult to make accurate and timely evaluation and decision-making. The traditional decision-making based on the experience, knowledge and preference of decision makers cannot meet the needs of "timely, correct location, appropriate quantity and qualified quality" of logistics support. Therefore, it is necessary to study how to make decisions on the large-scale, complex structure and multiple objectives of emergency logistics support. With the continuous application of artificial intelligence technology in various fields, the intelligent level of the decision-making system is improved by the aid of the decision-making system, and the application of intelligent optimization algorithm to solve the complex emergency logistics support decision-making problem has become an important means and development trend. Ant colony optimization, particle swarm optimization, genetic algorithm, neural network, simulated annealing method, differential evolution algorithm, biogeography optimization can be introduced to optimize, improve the decision-making accuracy and operability, and verify and improve in emergency logistics and emergency logistics practice, so as to improve the intelligent level and comprehensive benefit of logistics support decision-making evaluation. With the development of intelligent technology such as genetic algorithm, evolutionary computation, fuzzy logic and immune algorithm, the hybrid intelligent system with multiple intelligent technologies will be an important means of evaluation and decision-making in the field of logistics.

The next key research direction is to build the scientific evaluation system and intelligent decision-making platform of emergency logistics system. The basic idea of emergency logistics intelligent system evaluation is to first define the evaluation purpose, then establish a reasonable evaluation index system, and finally use intelligent means to select evaluation methods and build evaluation model. Intelligent system decision-making is based on the scientific analysis of the system, using intelligent decision-making technology based on emergency logistics big data to better assist the command personnel to make efficient reasoning and judgment, and make scientific decisions. Emergency logistics intelligent system evaluation and decision-making, the goal is to simulate the thinking method of emergency logistics experts to solve emergency logistics problems at a higher level, better mine effective information and carry out integrated application, and provide alternative plans, measures or suggestions for commanders in the process of logistics decision-making. "Intelligence" comes from data. The core ability of intelligent logistics is the intelligent decision-making system based on big data. Through machine learning, it can improve the ability to solve the emergency logistics decision-making problems, and assist the emergency logistics commanders in decision-making at a higher level. With the development of decision-making theory, emergency operation research theory, system engineering theory and artificial intelligence technology, the scheme evaluation method is intelligent, which provides a new tool for logistics support plan evaluation and provides the advantages that the traditional evaluation method does not have. The degree of intelligence will promote the logistics index efficiency and guarantee efficiency greatly.

5 Intelligent Management and Control of Emergency Logistics

With the continuous improvement of the cognitive efficiency and management decision-making level of intelligent logistics, the value of data is continuously transferred and added along the logistics chain, and data has gradually become the main body of the intelligent emergency logistics, and the scientific management has ushered in a new change [5].

In warehouse management and inventory control, common models include deterministic inventory model and stochastic inventory model. The intelligent system of emergency logistics based on artificial intelligence technology will assist the management personnel to analyze the historical consumption data and in and out information of all kinds of emergency materials in real time, predict the demand change of all kinds of materials and equipment in different emergency stages, and effectively control and dynamically adjust the inventory. Through this method, we can ensure the orderly balance of emergency materials reserve, avoid overstocking, waste or short supply, reduce the cost of logistics storage, and improve the efficiency of emergency support. The intelligent simulation of inventory decision model will be the development direction of this field.

Intelligent control technology (fuzzy control technology, neural network control technology, learning control technology, expert control technology, etc.) is a new control technology which develops continuously by applying intelligent theory to control technology. The application of this technology in the field of intelligent logistics is still relatively small, and it is still in the initial stage in emergency support, especially in the field of logistics operation, the logistics equipment monitoring, handling robot, sorting robot, intelligent warehouse control and so on are becoming more and more complex, which has a broad application prospect. We concluded these methods improve logistics operation, optimizing decision-making, and avoiding losses during the logistics flow [6] (Fig. 2).

6 Conclusion

The combination of system engineering method and intelligent technology in the field of emergency logistics is a new field. It is the interdisciplinary science to study large-scale complex artificial composite system from the perspective of artificial intelligence. It is not only a technical process, but also a management process. The construction of intelligent system of emergency logistics includes not only the construction of hardware infrastructure, but also the improvement of decision-making management level. The improvement of the overall level of emergency logistics not only depends on the improvement of infrastructure invested by a large number of human, material and financial resources, but also pays attention to the following aspects from the perspective of system engineering: The management and operation

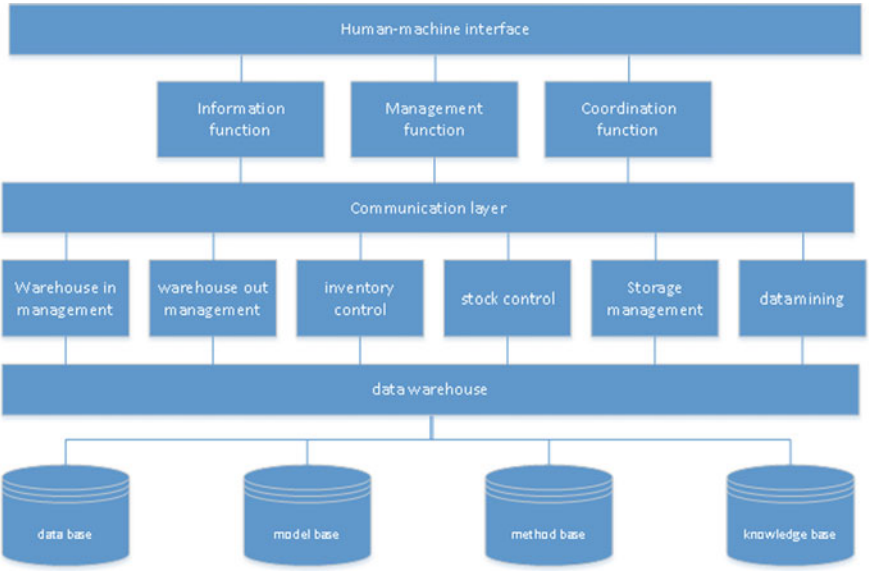


Fig. 2 Structure chart of intelligent warehouse management and inventory control

level of emergency logistics is improved, such as the renewal of emergency logistics concept, the adjustment of logistics organization system, the improvement of emergency logistics business process, the utilization of various resources and the construction of information system. Through the wide application of artificial intelligence, machine learning, big data, cloud computing, Internet of things and other intelligent technology groups, we will improve the intensive, standardized and intelligent level of modern emergency logistics system, ensure the mutual promotion of emergency benefits and economic benefits, and reduce or even avoid the blindness, randomness and irrationality of emergency logistics system, and improve the overall efficiency and comprehensive support capacity of the system.

References

1. Barceló, J., Grzybowska, H., Pardo, S.: Vehicle routing and scheduling models, simulation and city logistics. Dynamic Fleet Management, pp. 163–195. Springer, Boston, MA (2007)
2. Cyber-physical system (CPS) program solicitation. National Science Foundation of the United States. http://www.nsf.gov/pubs/2010/nsf10515/nsf10515.html (2010)
3. Gattuso, D., Cassone, G.C., Pellicano, D.S.: A micro-simulation model for an intelligent logistics platform: specification and calibration results. Supply Chain Forum Int. J. **15**(4), 52–69 (2014)
4. Kovalský, M., Mičieta, B.: Support planning and optimization of intelligent logistics systems. Proc. Eng. **192**, 451–456 (2017)

5. Min, Hokey: Artificial intelligence in supply chain management: theory and applications. Int. J. Log. Res. Appl. **13**(1), 13–39 (2010). https://doi.org/10.1080/13675560902736537
6. Oliveira, Rodrigo R., et al.: An intelligent model for logistics management based on geofencing algorithms and RFID technology. Exp. Syst. Appl. **42**(15–16), 6082–6097 (2015)

A Technical Survey on Approaches for Detecting Rogue Access Points

Jianwei Hu, Yetao Li, Yanpeng Cui, and Le Bu

Abstract With the popularity of wireless local area networks (WLANs) and access points (APs) which play an integral part in the wireless infrastructure coordinating wireless users and connecting them to the wired side of networks to the Internet become increasingly vulnerable and are exposed to malicious attacks. This paper focuses on rogue access points (RAPs) common attacks. The attack principle of different types of RAPs and corresponding RAPs detection methods are presented. Besides, the disadvantage and strength of these RAPs detection methods are also compared in this survey. Finally, some possible issues and future research trends are introduced.

1 Introduction

Wireless networks are rapidly spreading due to their advantages such as convenience, flexibility, mobility, scalability, and easy installation. However, the universalization of this technology also increases the security risk. RAPs are one of the most dangerous security risks.

RAP usually has the same service set identifier (SSID) with the legitimating one to cheat users. As a wireless network attack reported in 2015, attackers successfully tricked users into accessing their APs and obtain sensitive information such as bank account numbers, passwords, and photos. As the RAPs set by the attacker usually provides the network services normally, users will not feel a significant difference when accessing the RAP. In enterprise networks, attackers can also use RAPs to invade internal networks. Therefore, both wireless networks in public environments and the internal wireless networks of enterprises need security detection.

J. Hu · Y. Li (✉) · Y. Cui
Xidian University, No. 2 South Taibai Road, 710071 Xi'an, China
e-mail: liyetao2012@outlook.com

L. Bu
Imperial College, London, UK

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_20

2 Taxonomy of RAPs

In the literature, RAPs are classified into three categories: twin APs, incorrectly configured APs, and unauthenticated APs.

2.1 Twin APs

Twin APs are RAPs that highly mimic legitimate APs [1], which is usually carried out by portable devices, small in size, and difficult to detect [2]. They exist in two types: substitution and coexistence. Substitution means that RAPs take place to legitimate APs by disconnecting users or someway else [3], and trick users into accessing them. Coexistence represents that a RAP and a legitimate AP coexist in a certain space and compete providing services at the same time.

2.2 Incorrectly Configured APs

APs in enterprises configured incorrectly by network employees could cause security problems [4, 5]. For example, attackers could steal network certificates by attacking mobile devices used by users to gain access to the enterprise internal network [6, 7]. Moreover, some personal APs in small shops or restaurants usually post the passwords in public or do not have a password [8].

2.3 Unauthenticated AP

An unauthenticated AP refers to an AP built privately on the network. Such APs are a part of the network but neither uniformly deployed nor controlled by the administrator. For example, if employees share the network privately, unauthorized users will access the network through these unauthenticated APs or sniff the network traffic [9]. Attackers can also set up unauthenticated APs deliberately, harming the network and stealing sensitive information.

3 Existing RAP Detection Approaches

Several novel approaches have been proposed by researchers. The perspective of the administrator and client detection are summarized below.

3.1 Detection from Client-Side

Use special length frame arrival time (SLFAT) to detect RAPs. It provides that the same gateway is used, as the legitimated APs [10]. SLFAT monitors the traffic sent by the target APs at the detection node and extracts the arrival time of special frames of the same length to determine whether there is a RAP forwarding data packet.

Detection of routing options based on IP headers [11]. The IP header has a record routing option function. When this function is turned on, the router address passed between the source address and the destination address will be recorded in the IP packet header of the data packet. After receiving the data packet, you can view all the passed IP addresses. If an abnormal path is found, a RAP exists. However, due to security and other considerations, many firewalls will disable or ignore packets that record routes [12], and the average number of routes on the Internet is 19–21, which is much higher than the number of 9 addresses that can be recorded [13].

Detection based on TCP connection [14]. The client connects to a nearby AP and establishes a TCP connection with the public server. When switching to a different AP with the same wireless SSID, the client sends a heartbeat request to the former public server. Only the gateway keeping, switches AP will not affect the previous TCP connection. Then, the public server can respond to heartbeat requests from the client. Conversely, a RAP exists. Meanwhile, if an attacker impersonates a public server to perform a man-in-the-middle attack, this detection will not work.

Detect duplicate association [15]. During the authentication stage, the retry bits, sequence number, and association ID (AID) of the two association responses are analyzed to verify whether a RAP exists. However, for some open-mode RAPs, this detection approach fails.

Authentication using a dedicated public server and watermark packets [16, 17]. The server continuously replies to the watermark data packet which the client sends to it before to the client. At the same time, the client detects whether other channels are transmitting the watermark. If the data packet exists, the initial AP is a RAP, and the others are legitimate. This applies when an attacker uses a legitimate access point to provide network services. However, the attacker can avoid detection by storing the watermarked data packet and disconnecting with the legitimate AP at their replay arrival time and the round-trip time.

Detect base on RAP and rogue wireless client (RWC) [14]. When the wireless client (WC) sends/receives data via RAP, the attacker's RWC and legitimate AP will also send/receive the same data, which means WC data will appear twice on the wireless channel. However, this detection method is not valid in the mode of the man-in-the-middle attack.

Detection based on the confirmation number and serial number [18]. The confirmation number and serial number in the IP packet are used as the basis for judging whether there is forwarding in the wireless network. However, this detection method requires a lot of details of the IP packet header. Encrypted wireless network services cannot be supported.

3.2 Detection from Administrator Side

Using authorization lists to detect RAPs [19]. Administrators usually form authorization lists (white lists) that contain the identity information of legitimate APs by scanning APs neighboring of their identity information and then comparing them with lists to detect RAPs' existence. Therefore, focusing on significant information and ignoring unnecessary data can do much help for reducing detection time, while the information in the data packet is not hard to alter.

Clock deviation detection approach [20]. It is a passive detection method that uses information extracted from the beacon frame to combine a clock offset with the device's inherent temperature to detect a RAP. However, this method is limited to detecting RAPs released by software.

Honeypot-based wireless intrusion detection scheme [16]. First is distinguished whether the packet comes from an authorized host, to detect the man-in-the-middle attack by the packet flow rate decreases with the increase of the packet spacing, Snort detects DoS, DNS spoofing, and then redirects malicious traffic from Snort to the KFSensor honeypot for deepening analysis. The scope of this system is a small network, and it remains to be determined whether it can be extended to a larger network.

Use dedicated hardware to interfere with channels for detection [21]. The AP2 throughput of non-adjacent channels is detected by interfering AP1 channels. If a drop occurs, AP2 may be a RAP because the wireless connection is provided by replaying the signal from AP1. However, this approach interferes with the normal user's online experience. And deploying wireless sensors across different wireless networks or gathering traffic at a centralized site is expensive and complex.

Detection based on protocol modification [22]. It mainly uses received signal strength (RSS) which changes on the wireless channel between the client and the AP. The client and the AP need to exchange challenge and response packets to complete the detection. This detection method is efficient, but modifying the protocol involves driver and firmware upgrades, which make it difficult to be popularized.

Group-aware approach [13]. It uses the spatial correlation of RSS for detection. RSS measurements collected from the population helps to provide a robust profile and minimize the impact for the inaccuracies on individual RSS values. The measurement can also dynamically match the configuration file to filter out the abnormal samples detected in real-time. The efficiency of this method depends on the number of users in the detection area.

Physical layer channel state information (CSI) [23]. A position model which is based on the edge of the landmark area, combining a large amount of crowdsensing data, is used to determine whether the detected AP is a RAP. This method uses a crowdsourcing strategy which is also depends on the number of users.

Table 1 Strengths and weaknesses of current techniques

Method	Technical requirements	Strengths	Weaknesses
Time-based detection	None	Passive detection based on the frame information	Only applicable to software release attacks
Authorization list	None	Fast detection speed	AP fingerprint spoofing can bypass the detection
Honeypot	Honeypot system	Semi-defense detection	Suitable for small network environment
Active interference	Dedicated sensor hardware	Low false positives	Affects normal network communication, high hardware deployment costs
Protocol-based	Modify agreement	High efficiency	Difficult to popularity
Automated detection	Sensor device	Passive, little infrastructure	Takes time and energy a lot

4 Conclusion

Because the simplicity of RAP creation takes a security threat to the wireless network, several detection approaches proposed by researchers. The current techniques have several weaknesses, as listed in Table 1.

Early RAP detection mainly used the authorization list to identify APs and RAPs by media access control address (MAC) and SSID. Later, multi-parameter detection was developed, and more fingerprint information of AP devices was involved in the detection work. The detection range was limited based on the time method. Later interference technology affects network user. Meanwhile, other testing methods are introduced with their strength and limitations. With the development of intelligent technology, the research of detection technology is gradually developing toward automation. Wireless intrusion detection system (WIDS) technology is the trend of future wireless security research.

References

1. Breński, K., Chołuj, M., Luckner, M.: Evil-AP-mobile man-in-the-middle threat. In: IFIP International Conference on Computer Information Systems and Industrial Management, Springer, Cham. pp. 617–627. (2017)
2. Lanze, F., Panchenko, A., Ponce-Alcaide, I., Engel, T.: Undesired relatives: protection mechanisms against the Evil Twin attack in IEEE 802.11. In: Proceedings of the 10th ACM Symposium on QoS and Security for Wireless and Mobile Networks, pp. 87–94 (2014)
3. Khan, W.Z., Aalsalem, M.Y., Saad, M.N.B.M., Xiang, Y.: Detection and mitigation of node replication attacks in wireless sensor networks: a survey. Int. J. Distrib. Sens. Netw. **9**(5), 149023 (2013)

4. Bartoli, A., Medvet, E., Onesti, F.: Evil Twins and WPA2 enterprise: a coming security disaster? Comput. Secur. **74**, 1–11 (2018)
5. Vanhoef, M., Piessens, F.: Key reinstallation attacks: forcing nonce reuse in WPA2. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1313–1328 (2017)
6. Bartoli, A., Medvet, E., De Lorenzo, A., Tarlao, F.: In: Secure Configuration Practices of WPA2 Enterprise Supplicants. arXiv: Cryptography and Security (2018)
7. Brenza, S., Pawlowski, A., Pöpper, C.: A practical investigation of identity theft vulnerabilities in eduroam. In: Proceedings of the 8th ACM Conference on Security and Privacy in Wireless and Mobile Networks, pp. 1–11. (2015)
8. Ma, L., Teymorian, A. Y., Cheng, X., Song, M.: RAP: Protecting commodity Wi-Fi networks from rogue access points. In: The 4th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness and Workshops, pp. 1–7 (2007)
9. Alotaibi, B., Elleithy, K.: Rogue access point detection: taxonomy, challenges, and future directions. Wire. Pers. Commun. **90**(3), 1261–1290 (2016)
10. Lu, Q., Qu, H., Ouyang, Y., Zhang, J.: SLFAT: client-side Evil Twin detection approach based on arrival time of special length frames. Secur. Commun. Netw. **2019** (2019)
11. Nikbakhsh, S., Manaf, A. B. A., Zamani, M., Janbeglou, M.: A novel approach for rogue access point detection on the client-side. In: 2012 26th International Conference on Advanced Information Networking and Applications Workshops, pp. 684–687. IEEE. (2012)
12. Sherwood, R.: Discovering and securing shared resources on the internet (Doctoral dissertation). (2008)
13. Zhou, T., Cai, Z., Xiao, B., Chen, Y., Xu, M.: Detecting rogue AP with the crowd wisdom. In: International Conference on Distributed Computing Systems. (2017)
14. Nakhila, O., Amjad, M. F., Dondyk, E., Zou, C.C.: Gateway independent user-side Wi-Fi Evil Twin attack detection using virtual wireless clients. Comput. Secur. 41–54. (2017)
15. Agarwal, M., Biswas, S., Nandi, S.: An efficient scheme to detect Evil Twin rogue access point attack in 802.11 Wi-Fi networks. Int. J. Wire. Inf. Net. **25**(2), 130–145. (2018)
16. Agrawal, N., Tapaswi, S.: The performance analysis of honeypot based intrusion detection system for wireless network. Int. J. Wire. Inf. Netw. **24**(1), 14–26 (2017)
17. Mónica D., Ribeiro, C.: Wi-Fi Hop—mitigating the Evil Twin attack through multi-hop detection. In: Atluri, V., Diaz, C. (eds.) Computer Security—ESORICS 2011. Lecture Notes in Computer Science, vol 6879. Springer, Berlin, Heidelberg. (2011)
18. Hsu, F.H., Wang, C.S., Hsu, Y.L., Cheng, Y.P., Hsneh, Y.H.: A client-side detection mechanism for Evil Twins. Comput. Elect. Eng. **59**, 76–85 (2017)
19. Vanjale, S.B., Mane, P.B.: Multi parameter based robust and efficient rogue AP detection approach. Wire. Pers. Commun. **98**(1), 139–156 (2018)
20. Lanze, F., Panchenko, A., Braatz, B., & Engel, T.: Letting the puss in boots sweat: Detecting fake access points using dependency of clock skews on temperature. In: Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security, pp. 3–14 (2014)
21. Jang, R., Kang, J., Mohaisen, A., Nyang, D.: Highly-accurate rogue access point detection using intentional channel interference: poster. In: Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks, pp. 277–278. (2017)
22. Altaweel, A., Stoleru, R., Gu, G.: EvilDirect: a new Wi-Fi direct hijacking attack and countermeasures. In: 2017 26th International Conference on Computer Communication and Networks (ICCCN) (pp. 1–11). IEEE. (2017)
23. Wang, C., Zhu, L., Gong, L., Liu, Z., Mo, X., Yang, W., … Li, Z.: Detecting Evil-Twin attack with the crowd sensing of landmark in physical layer. In: International Conference on Algorithms and Architectures for Parallel Processing, pp. 234–248. Springer, Cham. (2018)

A Matching Algorithm Based on the Fusion of SAD and Census

Han Liu, Yan-tong Zhao, and Wei Wang

Abstract This paper analyzes the characteristics of SAD algorithm and Census algorithm, improves SAD algorithm and Census algorithm, respectively, merges them, and proposes a SAD and Census algorithm based on adaptive window and adaptive weight. The experimental results show that the algorithm has stronger robustness and higher matching accuracy than SAD algorithm and Census algorithm.

1 Introduction

Binocular stereo vision is designed to extract 3D information from 2D images to recover 3D scenes. It is widely used in 3D measurement, virtual reality, robot navigation, and other fields [1]. Stereo matching technology is the key technology of binocular stereo vision [2]. Its task is to obtain the corresponding points in the two images and calculate the parallax, so as to obtain the depth information in the scene.

At present, stereo matching can be divided into four steps: matching cost calculation, matching cost aggregation, parallax calculation and optimization, and parallax refinement [3]. According to the different methods of parallax calculation, stereo matching algorithm is divided into global algorithm and local optimization algorithm. The global algorithm obtains the matching result by constructing the energy function and minimizing it, with low accuracy and high efficiency, but it is not suitable for the realization of a low-cost independent stereo vision system. However, although the local optimization algorithm has high efficiency, it is easy to cause mismatching and is difficult to be applied in practical situations.

After studying SAD and Census algorithm, this paper fuses the SAD algorithm with Census algorithm, and proposes a stereo matching algorithm based on SAD and

H. Liu (✉) · Y. Zhao · W. Wang
Key Laboratory of Space Manufacturing Technology, Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China
e-mail: liuhan171@mails.ucas.edu.cn

H. Liu · Y. Zhao
University of Chinese Academy of Sciences, Beijing 100094, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_21

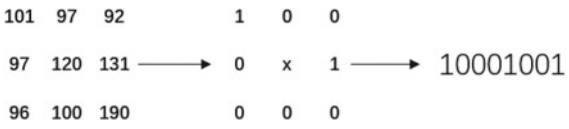


Fig. 1 Schematic diagram of census transformation

Census algorithm with adaptive weights and windows. This algorithm can improve the matching accuracy and has good robustness to the difference of image brightness caused by the change of illumination and different exposures.

2 Stereo Matching Fundamentals

2.1 Sum of Absolute Differences

The sum of absolute differences (SAD) algorithm [4] has only plus and minus operation, so it is simple and efficient in respect of algorithm complexity and hardware implementation. As shown in Fig. 1, such an approach first establishes a window to assist in matching. As the point to be matched in the base graph, a certain point to be matched is taken as the center point in the registration graph to create an identical sub-window, and then the gray value of multiple pixel points in the window is used to represent the center point. In the search range of the registration map, that is, the parallax range, the comparison continues, and then is calculated the sum of the absolute value of the gray difference of all the pixels. The algorithm expression of SAD [5] is shown in Eq. (1).

$$\mathrm{SAD}(i, j, \mathrm{disp}) = \sum_{h=-\frac{w-1}{2}}^{h=\frac{w-1}{2}} \sum_{k=-\frac{w-1}{2}}^{k=\frac{w-1}{2}} |I_L(i+h, j+k) - I_R(i+h, j+k+\mathrm{disp})| \tag{1}$$

where $I_L(i, j)$, $I_R(i, j)$ are the pixel gray value of the position in the left and right images, and w is the window size. When the SAD obtained by the two sub-windows is the smallest, the corresponding point is the matching point, and then the parallax value disp is obtained.

2.2 The Census Algorithm

Census algorithm [6] also needs a window. The algorithm starts with the Census transformation. For this, it is selected any point in the view and is created a window centered on that point. Every point in the window except the center point is then

compared to the center point. If the gray value is larger than the center point, it is denoted as 1; otherwise, it is denoted as 0. Take the resulting sequence of only 0 and 1 s is obtained as the sequence of the center point. Then, the Hamming distance is used to calculate the similarity, and the point with the minimum Hamming distance is found in the matching graph, that is, the matching point. Finally, the parallax map is obatained.

3 Improved Algorithm

3.1 Improvement of SAD Algorithm

In the traditional SAD algorithm, the absolute value of the gray difference of each position in the window has equal weight. In this article, the smaller the correlation between the pixels in the window and the center, the lower the weight. Therefore, this paper attempts to calculate the SAD value by taking the set distance between the pixels in the window and the center as a weight factor. In this paper, the weight is normally distributed.

3.2 Census Algorithm Improvement

The Census transformation depends too much on the center pixel. If the selection of the center pixel or the center pixel itself is interfered by noise, the result of the transformation will be greatly affected. In view of this, this paper proposes to replace the gray value of the center pixel with the median value of the pixel in the window.

For this, first are sorted the pixels in the window from small to large, then is obtained the median value, and the center pixel value is replaced by the median value.

In the traditional method, the center pixel is affected by noise, and the comparison results are different, but the median value is replaced by the center value as the new reference value. As shown in the figure below, the Census code obtained is 10001001 by taking the transformation window of 3*3 as the column. When the center pixel is polluted by noise and changes from 120 to 90, the encoding in accordance with the traditional method changes to 11111101, while the result of the improved method transformation will still be 10001001. Therefore, this improved method can effectively reduce the influence of noise on the image and enhances the robustness of the algorithm.

3.3 Adaptive Window Selection

Whether it is Census algorithm or SAD algorithm, their accuracy is related to the selection of windows. In the weak texture region, a larger window is needed to determine the exact location of the current pixel points to improve the matching accuracy; in the discontinuous depth region, a smaller window can be used to protect the edge details of the object.

This paper presents an adaptive window selection method. Let $P(i, j)$ is the point to be matched and is taken as the reference point. The adjacent pixel points around I are $P(i + 1, j)$ and $P(i - 1, j)$. Then, the discriminant criterion of pixel gray difference is shown in Eq. (2)

$$\frac{\sqrt{\left|(I_{i+1})^2 - (I_i)^2\right|}}{\max\{I_{i+1}, I_i\}} \leq \sigma \tag{2}$$

where I_i is the gray value of $P(i, j)$ to be matched, I_{i+1} is the gray value of $P(i + 1, j)$, σ is the preset gray threshold selected according to the actual demand, $0 \leq \sigma < 1$.

However, only one restriction condition will cause infinite selection of the window in the smooth transition region of grayscale. In this paper, the restriction condition of window truncation is added in the process of expansion. The calculation method is shown in Eq. (3).

$$\frac{\sqrt{\left|(I_{i+1})^2 - (I_i)^2\right|}}{\max\{I_{i+1}, I_i\}} > \partial \tag{3}$$

where: I_n is the gray value of the Nth pixel extended to the right by the point to be matched and ∂ is the window truncated gray threshold: $0 \leq \sigma < \partial < 1$.

4 Algorithm Implementation and Experimental Results

4.1 The Experimental Results

In this paper, images of Tsukuba, Teddy, Venus, and Cones provided by Middlebury platform are used for experiments.

The original image and the ideal parallax diagram are shown in Fig. 2.

In the picture, columns 1, 2, 3, and 4 are images of Tsukuba, Teddy, Venus, and Cones, respectively. The first, second, and third lines are the left, right, and ideal parallax images in the original image group, respectively.

These sample graphs are matched by this algorithm, and the parallax results are shown in Fig. 3.

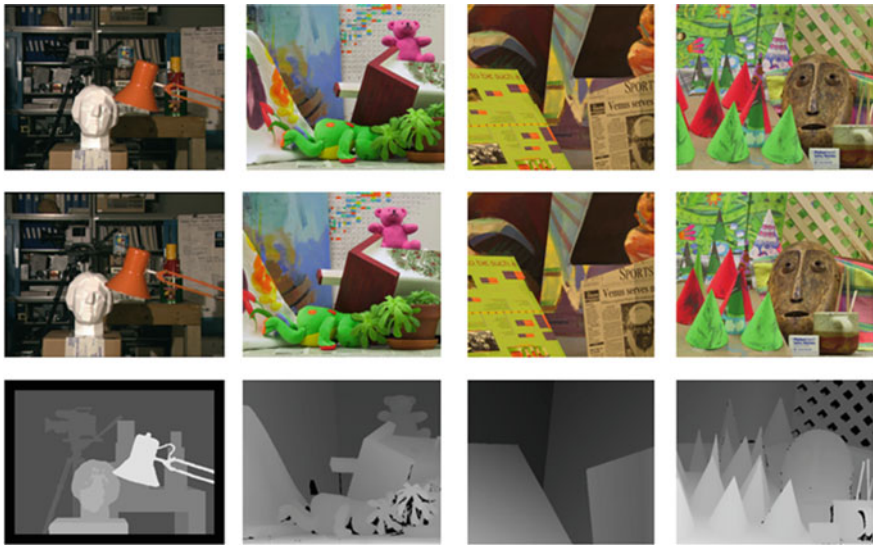


Fig. 2 Middlebury datasets and Middlebury ground truth

Fig. 3 Experimental results of the proposed algorithm

When the parallax algorithm is used to obtain the best parallax figure and ideal figure to make difference, the difference exceeds the threshold value of markers for matching, and adopts the quantitative error matching rate matching quality; the matching error rate is smaller, and the match is better. The algorithm of matching error rate as a result, the result compared with other algorithms, which is reference list of currently published stereo matching algorithms at Middlebury vision. The results are shown in Tables 1 and 2. The "nocc" column shows the mismatching rate of the non-blocking area, and "All" represents the overall mismatching rate of the image.

Obviously, the algorithm presented in this paper is much better than the traditional Census and SAD–IGMCT [7] algorithm. This algorithm is better than a lot of good algorithms, compared to classic AdaptWeight [8] algorithm; in addition to Tsukuba image matching results, the matching results are better than the AdaptWeight algorithm and compared with those of SegmentTree [9], Teddy and Cones matching error rate is low, on the whole, and its precision high. Therefore, the algorithm in this paper has high matching accuracy.

Table 1 Middlebury evaluation results for the Tsukuba and Teddy images

Algorithm	Tsukuba			Teddy		
	Nocc	All	Disk	Nocc	All	Disk
This algorithm	2.19	3.52	7.01	5.11	10.5	12.9
Traditional census	9.06	10.6	21.5	12.7	21.1	28.1
SAD-IGMCT	5.81	7.11	22.6	9.79	15.5	25.7
Adapt weight	1.38	1.85	6.80	7.88	13.3	18.6
Segement tree	1.11	1.37	5.79	4.22	7.06	1.44

Table 2 Middlebury evaluation results for the Venus and Cones images

Algorithm	Venus			Cones			Ave.bad pixels
	Nocc	All	Disk	Nocc	All	Disk	
This algorithm	0.47	0.90	2.89	2.75	8.74	7.42	5.37
Traditional census	2.34	3.72	21.5	7.90	17.0	16.2	14.3
SAD-IGMCT	2.61	3.33	25.1	5.98	11.5	15.0	12.5
Adapt weight	0.71	1.19	6.14	3.97	9.79	8.26	6.67
Segement tree	4.22	7.06	11.8	2.48	7.92	7.32	4.23

5 Conclusion

In this paper, a local stereo matching algorithm based on SAD and Census transform is proposed. Firstly, the appropriate window size required by the algorithm is obtained through the adaptive window, and the pixel weight of each region is assigned to increase the relation between pixels. Meanwhile, the gray value of the center pixel is replaced by the median value of the gray value of the transformation window to increase the robustness of the algorithm. The results show that this approach is better than traditional SAD and Census algorithms, and also better than many excellent classical algorithms.

References

1. Yongjie, Y., Qidan, Z., Zhuang, L., Quanfu, C.: Camera calibration in binocular stereo vision of moving robot. In: World Congress on Intelligent Control and Automation IEEE (2006)
2. Chun-He, Y.U., Jing, Z.: Research on sad-based stereo matching algorithm. J. Shenyang Aerosp. Univ. (2019)
3. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision **47**(1–3), 7–42 (2002)
4. Banks, J., Bennamoun, M., Corke, P.: Non-parametric techniques for fast and robust stereo matching. In: IEEE Tencon 97 IEEE Region 10 Conference Speech and Image Technologies for Computing and Telecommunications. IEEE (1997)

5. Ambrosch, K., Kubinger, W., Humenberger, M., Steininger, A.: Hardware implementation of an SAD based stereo vision algorithm. In: 2007 IEEE Co66mputer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), pp. 18–23 June 2, Minneapolis, Minnesota, USA (2007)
6. Zabih R.: Non-parametric local transforms for computing visual correspondence. Comput. Vision-ECCV'94 (1994)
7. Ambrosch, K., Kubinger, W.: Accurate hardware-based stereo vision. Comput. Vision Image Underst. **114**(11), 1303–1316 (2010)
8. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern. Anal. Mach. Intell. (2006)
9. Mei, X., Sun, X., Dong, W., Wang, H., Zhang, X.: Segment-tree based cost aggregation for stereo matching. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2013)

Evaluation Method of Vehicle Technology Status Based on Big Data Analysis Technology

Hong Jia, Ying-Ji Liu, Hai-Ying Xia, and Guo-Liang Dong

Abstract This paper presents a method, which uses big data analysis for maintaining data of a vehicle type to evaluate the technical status of a vehicle. The failure rate of parts can reflect the technical status of each system of the vehicle. By establishing an evaluation model of vehicle technical status and data mining, the vehicle technical status of the entire vehicle model is displayed as the vehicle mileage changes. At the same time, combined with the current information and maintenance data of the vehicle, its current technical status can be effectively estimated, and it can provide a reference for the owner to target the maintenance of his vehicle.

1 Introduction

With the sustained and rapid development of China's economy and society, the number of motor vehicle ownership in China continues to grow rapidly, and the demand for vehicle maintenance has also greatly increased, which has promoted the rapid development of the automotive maintenance industry. Vehicle transactions are increasingly appearing in people's life. With the improvement of the informatization level of the auto-repair industry and the accumulation of massive auto-repair data, the method of using big data technology to analyze and evaluate the technical status of vehicles has more and more practical value.

Traditional methods for assessing the technical condition of automobiles [1–4] mainly rely on the skills and experience of professional technicians and evaluate only one aspect of the vehicle to make qualitative judgments on the external technical conditions and working conditions of vehicles, supplemented by simple appliances to perform an intuitive inspection of the technical condition of vehicles. These methods are time consuming, low efficient, and have certain limitations.

H. Jia (✉) · Y.-J. Liu · H.-Y. Xia · G.-L. Dong
Key Laboratory of Operation Safety Technology on Transport Vehicles, Ministry of Transport, Beijing, China
e-mail: h.jia@rioh.cn

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_22

Big data analytics are now widely used [6, 7]. This paper presents a method, which uses big data analysis of maintenance data of a vehicle type to evaluate the technical status of a vehicle. The failure rate of parts can reflect the technical status of each system of the vehicle. The more maintenance data of a vehicle type, the more accurate the evaluation of the technical condition of the vehicle is. This method is highly objective and efficient. By establishing an evaluation model of automotive technical conditions and data mining, the technical status of the entire vehicle model as a function of vehicle mileage is demonstrated. At the same time, combined with the current information and maintenance data of the vehicle, the current technical status of the vehicle can be effectively estimated, and it can provide a reference for the owner to repair and maintain the vehicle in a targeted manner.

2 Evaluation Method of Vehicle Technical Condition

2.1 Overview of Evaluation Methods

The evaluation of the technical condition of the vehicle is composed of six aspects: the use of the vehicle, power performance, economic performance, safety performance, stability, and maintenance score. Among them, the use status is reflected by the ratio of the service life of the vehicle to the service life, the ratio of the mileage of the vehicle and the mileage limit of the model; the power performance is the comprehensive status between the failure status of the key components of the vehicle's power system and the failure rate of the power system of the model reflected; economic performance is comprehensively reflected by the number of key component failures of the vehicle fuel system and the fuel system failure rate of the vehicle model; safety performance is comprehensively reflected by the number of key component failures of the vehicle brake system and the vehicle brake system failure rate; The interval between the service mileage and the service time are comprehensively displayed; the maintenance score is based on the auto maintenance of the vehicle to extend the service life as a benchmark to build a maintenance score index.

2.2 Evaluation Model Establishment

2.2.1 Evaluation System

1. Usage

The evaluation of vehicle use status mainly reflects the current status of the vehicle's inherent attributes through the mileage and years of use, and then reflects the vehicle's own technical status. Among them, the mileage can reflect the vehicle condition information, the number of years can reflect the life information of the vehicle, and the vehicle identification number code can match the basic information of the vehicle

model. By comparing the number of years of use with the maximum service life of the model and the ratio of the actual mileage to the reference value of the maximum mileage of the model vehicle, the estimated state of the vehicle over its full life cycle is obtained. Vehicle usage scores are obtained using the following model:

$$U = \left(1 - \frac{U_1}{U_{1\text{lim}}}\right) \times B_{11} + \left(1 - \frac{U_2}{U_{2\text{lim}}}\right) \times B_{12}$$

Among them, U_1, $U_{1\text{lim}}$, U_2, and $U_{2\text{lim}}$ are the numbers of years of use, the mileage, and the corresponding limits. With reference to the "Requirements for the Mandatory Standard for Motor Vehicles," the mileage limit of the vehicle is set to 600,000 km and the service life of the vehicle is set to 12 years. B_{11} and B_{12} are the index weight values, which are obtained by analytic hierarchy process. The score of the vehicle usage status is obtained by combining the mileage and the number of years of use.

2. Power performance

The evaluation of vehicle power performance is mainly based on the number of failures of the vehicle power system and the failure rate of the vehicle power system. Among them, the failure rate of the power system is reflected by the failure rate of the key components of the power system, and the failure rate of the components is calculated from the replacement of the components in the vehicle maintenance data. The dynamic performance evaluation score is obtained by the following model:

$$D = (1 - D_1) \times B_{21} + (1 - D_2) \times B_{22}$$

Among them, D stands for vehicle dynamic performance score, B_{21} and B_{22} are indicator weight values. D_1 is the normalized value of the number of failures of the vehicle power system (mapped to the [0, 1] interval), D_2 is the standardized value of the failure rate of the vehicle's power system (mapped to the [0, 1] interval), and the vehicle's power performance score.

3. Economic Performance

The economic performance is represented by the failure rate of the fuel supply system. The failure rate of the fuel supply system is reflected by the failure rate of key components of the fuel supply system. The failure rate of the components is calculated by the replacement of components in the vehicle maintenance data of the model. The economic performance evaluation score is obtained by the following model:

$$F = (1 - F_1) \times B_{31} + (1 - F_2) \times B_{32}$$

Among them, F stands for vehicle economic performance score, B_{31} and B_{32} are index weight values. F_1 is the standardized value of the number of failures of the

fuel system of the vehicle (mapped to the [0, 1] interval) and F_2 is the standardized value of the failure rate of the vehicle's fuel system (mapped to the [0, 1] interval).

4. Safety Performance

The safety performance evaluation score of a vehicle is comprehensively expressed by the number of times the vehicle's brake system is repaired and the brake system failure rate of the vehicle model. The safety performance evaluation score is obtained by the following model:

$$S = (1 - S_1) \times B_{41} + (1 - S_2) \times B_{42}$$

Among them, S is the vehicle safety performance score, B_{41} and B_{42} are the index weight values. S_1 is the standardized value of the number of failures of the braking system of the vehicle (mapped to the [0, 1] interval), and S_2 is the standardized value of the failure rate of the braking system of the vehicle model.

5. Stability

The vehicle stability score is a combination of the average service interval and the average service mileage interval. Among them, the average maintenance interval refers to the average maintenance frequency of the vehicle in a certain period time. The frequency reflects the stability of the vehicle equipment. Frequently repaired vehicles have poor stability. The average value of the difference in mileage between multiple repairs reflects the stability of the performance of the vehicle through the average interval between repairs. The stability evaluation score is obtained by the following model:

$$T = T_1 \times B_{51} + T_2 \times B_{52}$$

Among them, T is the vehicle stability score, B_{51} and B_{52} are the index weight values. T_1 is the standardized value of the average maintenance time interval of the vehicle (mapped to the [0, 1] interval), T_2 is the standardized value of the average service mileage interval (mapped to the [0, 1] interval), and the standardized value can further intuitively reflect the vehicle in the overall situation.

6. Vehicle Maintenance

Vehicle maintenance can improve vehicle performance and extend vehicle life. Vehicle maintenance score calculation method: If keywords such as "maintenance" are retrieved from the vehicle maintenance records, it is counted as one vehicle maintenance; the maximum number of maintenance times $Q_{\max}$ is 4 in a prescribed unit period, and the minimum number of maintenance times $Q_{\min}$ is 0 within the statistical period. The number of vehicle maintenance Q is normalized (mapped to the [0, 1] interval), and the maintenance score is obtained by the following model:

$$H = y_{\min} + \frac{y_{\max} - y_{\min}}{Q_{\max} - Q_{\min}} \times (Q - Q_{\min}) = \alpha \times Q = \frac{1}{4} \times Q$$

It can be known from the mapping interval that $y_{\min} = 0$, $y_{\max} = 1$, combined with the minimum maintenance times $Q_{\min} = 0$ and the maximum maintenance times $Q_{\max} = 4$, the coefficient is calculated $\alpha = \frac{1}{4}$, and the maintenance score H is calculated by the model.

From this, the vehicle performance evaluation score model is as follows:

$$P = (U \times B_1 + D \times B_2 + F \times B_3 + S \times B_4 + T \times B_5) \times P_1 + H \times P_2$$

Among them, P is the score of vehicle technical status, U, D, F, S, T, and H are, respectively, the use score, power performance score, economic performance score, safety performance score, stability score, and vehicle maintenance score. Among them, the five indicators of usage score, dynamic performance score, economic performance score, safety performance score, and stability score are the key indicators, B_i, $i = 1, 2, 3, 4, 5$ are the key weights, and the key weights are analyzed through hierarchy. The vehicle maintenance score is an auxiliary indicator. The weight of the comprehensive score of the key indicators and the vehicle maintenance score are P_1, P_2, and the default is [0.95, 0.05].

Set the description table of the technical status of the vehicle based on the scoring standard module. As shown in Table 1, the technical status of the vehicle is divided into five levels. There are certain differences between the technical statuses of the different levels. Corresponding the calculated total score P of the comprehensive index with the division values of each level in the description table, the corresponding technical condition level and corresponding technical condition description are obtained.

Table 1 Description table of technical status of vehicles

Technical condition level	Score	Specific description
First level	$P > 90$	During the use of the vehicle, there were no excessive repairs, good maintenance, and key systems remained in good condition
Secondary	$70 < P \leq 90$	During use, the performance of the vehicle starts to deteriorate, and each system is damaged to maintain normal working conditions
Third grade	$50 < P \leq 70$	During use, the number of system assembly failures is high, and the performance of the vehicle is moderately aging
Fourth grade	$30 < P \leq 50$	The vehicle cannot be used normally, and all critical systems are severely damaged
Fifth grade	$P \leq 30$	The vehicle was completely inoperable and was about to be scrapped

2.2.2 Evaluation Method Steps

This method uses the analytic hierarchy process (AHP) method [7, 8] to determine the weight of the key indicators of the vehicle use status and the vehicle performance evaluation system. The comparison matrix of the two places is provided by expert experience.

S_1: extract the vehicle maintenance data of the vehicle type to be evaluated in a certain period of time, and preprocess the data. Among them, the vehicle type of the vehicle to be evaluated is identified through the vehicle identification code (VIN code) of the vehicle basic information. The vehicle maintenance data includes the vehicle's repair date, repair mileage, repair items, and repair accessories. The preprocessing process includes data quality verification and data extraction.

S_2: according to the data preprocessed in step S_1, obtain the evaluation index data of vehicle technical conditions.

S_3: use AHP to determine the weight of each index in the evaluation index system of automobile technical condition, further build the evaluation model and calculate the score of automobile technical condition.

3 Conclusion

In today's society, the pace of life is fast and the cost of living is high. Most car owners do not have a good sense of maintaining their cars on schedule. This phenomenon has exacerbated the deterioration of the technical level of cars. The method for assessing the technical condition of the automobile in this article allows the owner to effectively evaluate the technical condition of the automobile and timely maintenance of the automobile, which mining automobile maintenance data from a new perspective and can prolong the service life of the automobile.

Acknowledgements This work is funded by the special fund project for the basic research work of the central public welfare research institute (2018–9051).

References

1. Yang, X., Yang, D., Jia, H.: Research on reliability big data analysis technology of new energy vehicles. J. Highway Transport. Res. Dev. **34**(S2), 123–127 (2017)
2. Li, X., Lu, Y., Wang, Y., Huo, W.: Reliability evaluation of domestic trucks. Highway Transport. Technol. **05**, 81–84 (2000)
3. Zhong, K.: Research on used car evaluation methods based on performance indicators weighting revising. In: Research Institute of Management Science and Industrial Engineering. Proceedings of 2017 International Conference on Materials, Energy, Civil Engineering and Computer (MATECC 2017). Research Institute of Management Science and Industrial Engineering: Computer Science and Electronic Technology International Society, pp. 135–139 (2017)

4. Liu, J., Wu, X., Xiao, J., Chen, W.: Research on grey theory prediction of high mileage reliability of automobile based on segmented Weibull distribution model. Qual. Reliab. **05**, 30–35 (2019)
5. She, C., Zhang, Z., Liu, P., Sun, F.: Overview of application of big data analysis technology in new energy vehicle industry based on big data of new energy vehicle operation. J. Mech. Eng. **55**(20), 3–16 (2019)
6. Li, M., Wang, H., Chen, S., Liao, Y.: Industrial big data analysis technology and tire sales data prediction. Comput. Eng. Appl. **53**(11), 100–109 (2017)
7. Zhang, J., Xu, X., Ding, M., Li, J., Wang, J., Wu, C.: Transformer state evaluation based on fuzzy analytic hierarchy process. Power Syst. Prot. Control **45**(03), 75–81 (2017)
8. Zhou, Y.: Setting method of telecom industry rating weight based on fuzzy analytic hierarchy process. China Bidding **48**, 35–40 (2019)

Bay Number Recognition Based on Deep Convolutional Recurrent Neural Network

Xingxing Li, Chao Duan, Panpan Yin, Yan Zhi, and Na Li

Abstract The recognition of image text sequences has been the subject of long-term research in computer vision. The recognition problem of the shell number image sequence studied in this paper is based on the deep convolutional neural network architecture, and an image recognition algorithm that integrates feature extraction, sequence modeling, and transcription into a unified framework is adopted. The algorithm used has four characteristics: (1) Compared with most algorithms that require separate training and coordination, the method adopted in this paper has end-to-end characteristics. (2) It can handle indefinite length image sequences without involving character segmentation or horizontal-scale normalization. (3) The algorithm is not limited to any predefined vocabulary, and has achieved remarkable performance in both no-lexicon and dictionary-based scene text recognition tasks. (4) The algorithm produces an efficient and much smaller model, which is more practical for real-world scenarios. Using this algorithm for bay number recognition can assist in scene location and provide smart services and application through AI technology.

1 Introduction

The rise of artificial intelligence is mainly driven by the deep neural network model, especially deep convolutional neural networks have achieved great success in computer vision. However, in these machines vision-related tasks, image classification, and object detection are particularly concentrated [1]. In this article, we will use the image-based sequence recognition to identify the numbers in the bay marker. Unlike general object recognition, identifying such class sequence objects typically requires the system to predict a series of object tags rather than a single tag. Sequence prediction cannot be directly performed using the most popular deep convolutional

X. Li (✉) · C. Duan · P. Yin · Y. Zhi · N. Li
Department of Electronics and Information Engineering, Guangzhou College of Technology and Business, 528138 Foshan, China
e-mail: lixingxing_Edooy@163.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_23

neural network (DCNN) model, because DCNN models require input and output with fixed size and dimensions. Therefore, a variable length image sequence cannot be produced.

The recurrent neural network (RNN) model is another important branch of the deep neural network family, designed primarily to process sequences. One of the advantages of RNN is that no additional training and testing are required for each character in the sequence target image. However, a pre-processing procedure of converting an input target image into an image feature map is usually necessary. For example, Graves et al. [2] extracted series of geometric or image features from handwritten text, while Su and Lu [3] converted character images into sequence histogram of oriented gradient (HOG) features. The image pre-processing step is separated from the subsequent processing components, and is not connected to the entire algorithm flow. Therefore, existing systems based on RNNs cannot be trained and optimized in an end-to-end manner.

The main contribution of this paper is to use the convolutional recurrent neural network (CRNN) proposed by the author for shell number recognition [4] to help autopilot use the position information to locate. It is called CRNN model because it is DCNN, a combination with RNN. For image sequences, two neural networks can each exert their advantages. (1) The method used can learn directly from the sequence label, and does not need detailed annotation; (2) Method can extract features directly from the image data, without adding any manual features, like binarization, contour extraction, etc.; (3) With the same nature of RNN, it can produce a series of labels; (4) The model parameters generated by training are small and have real-time application value.

2 Introduction of Methods

2.1 CRNN Network Architecture

The network architecture of convolutional recurrent neural network (CRNN) is mainly composed of three parts, CNN + RNN + CTC, as shown in Fig. 1.

The network architecture consists of three components: (1) CNN layers, extracting feature sequences from the original image; (2) RNN layers, predicting the labels of each frame separately (3) Connectionist temporal classification (CTC) layers, which transforms the prediction of each frame into the final sequence of labels. At the bottom of the CRNN, the convolutional layer automatically extracts a sequence of features from each original input image. Above the convolutional network, a circular network is constructed to predict each frame of the feature sequence output by the convolutional layer. Each frame prediction of the loop layer is converted to a tag sequence using a transcription layer on top of the CRNN. Although CRNN consists of different types of network architectures (such as CNN and RNN), joint training can be performed through a loss function.

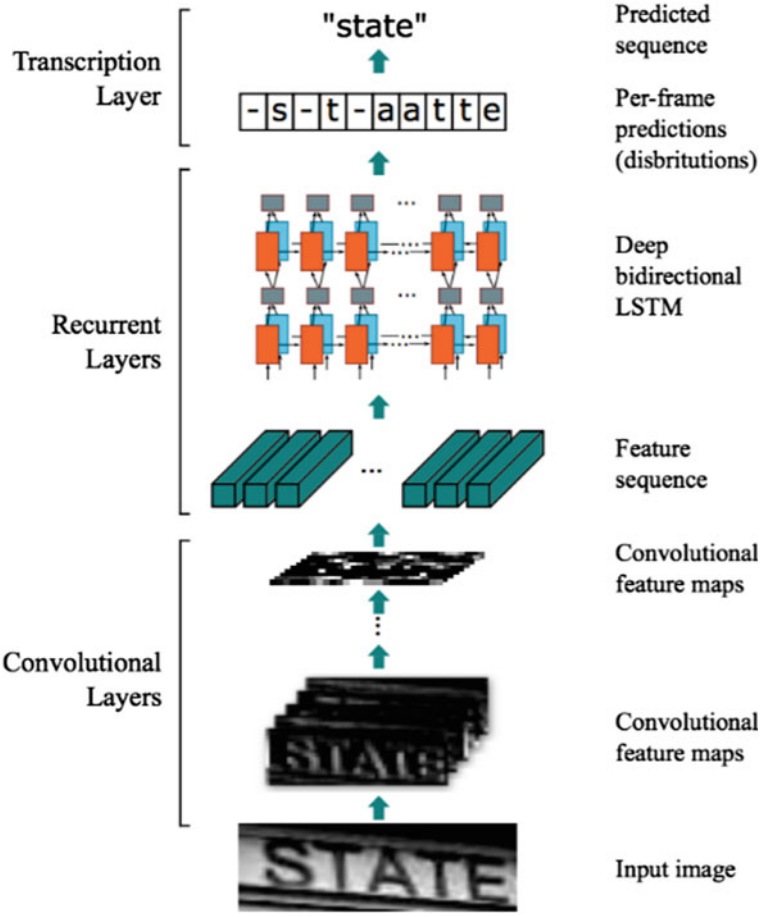


Fig. 1 CRNN network architecture

2.2 CNN Structure

In the CNN network layer, the main body adopts the Visual Geometry Group Network (VGG) structure, and some fine-tuning are made to the VGG network structure, as shown in Fig. 2.

From Fig. 2, the CNN layer makes the following adjustments to the VGG network structure.

The kernel size of the third and fourth pooling layers was changed from the original 2×2 to 1×2 to ensure that the features extracted by CNN can be input into the RNN network.

The method also adds a BN layer after the fifth and sixth convolution layers, thereby speeding up the training of the network.

Fig. 2 CNN layer network structure

Type	Configurations
Transcription	-
Bidirectional-LSTM	#hidden units:256
Bidirectional-LSTM	#hidden units:256
Map-to-Sequence	-
Convolution	#maps:512, k:2 × 2, s:1, p:0
MaxPooling	Window:1 × 2, s:2
BatchNormalization	-
Convolution	#maps:512, k:3 × 3, s:1, p:1
BatchNormalization	-
Convolution	#maps:512, k:3 × 3, s:1, p:1
MaxPooling	Window:1 × 2, s:2
Convolution	#maps:256, k:3 × 3, s:1, p:1
Convolution	#maps:256, k:3 × 3, s:1, p:1
MaxPooling	Window:2 × 2, s:2
Convolution	#maps:128, k:3 × 3, s:1, p:1
MaxPooling	Window:2 × 2, s:2
Convolution	#maps:64, k:3 × 3, s:1, p:1
Input	$W \times 32$ gray-scale image

The kernel size of the third and fourth maxpooling is changed from 2 × 2 to 1 × 2 in the CNN layer in order to facilitate the extraction feature of CNN as the input of the RNN. The requirement of the network for input feature dimension is w (wide) × 32, which means the network only requires that the high dimension of input feature is 32, and there is no limit to the width of feature. In this way, it is convenient to perform the next calculation as the input of the RNN, and each feature input is a one-to-one correspondence relationship. Each vector in the extracted feature sequence is associated with a receptive field of the input image, which can be used as the feature vector of the region.

2.3 RNN Network Structure

The RNN network is a sequence of features input to the CNN $x = x1, x2, \ldots, x$t, and each input xt has an output yt. In order to prevent the gradient from disappearing during the training process, the method uses the long short-term memory (LSTM) neural unit as the unit of the RNN. For the prediction of the sequence in this method, both the forward information propagation and the backward information propagation of the sequence are favorable for the sequence prediction. Therefore, this paper adopts a bidirectional LSTM network structure.

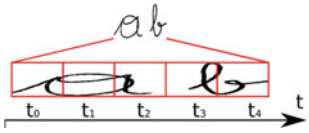


Fig. 3 Handwritten font image

2.4 CTC Translation Layer

A sequence of variable lengths is outputted by the RNN layer in CRNN, such as the original image width is W, and the number of sequences that may be output after passing through CNN and RNN is S. At this time, we need to translate the sequence into the final recognition result. When RNN performs timing classification, inevitably there will be a lot of redundant information. For example, a letter is recognized twice in succession. This requires a set of de-redundancy mechanism, but simply see two consecutive letters to remove redundancy. There are also problems, such as to choose words like trees, so CTC has a blank mechanism to solve this problem. Here is an example shown in Fig. 3.

In Fig. 3, t_0, t_1, and t_2 should be mapped to "a", and t_3 and t_4 should be mapped to "b". If we combine consecutively repeated characters into one output, then "aaabb" will be merged into an "ab" output. However, such a merge mechanism is problematic. For example, when the image is "aab", the sequence output will probably be "aaaaaaabb". There is no way to determine whether the text should be recognized as "aab" or "ab". In order to solve this ambiguity, CTC is proposed to insert the blank mechanism. For example, if we use the "-" symbol to represent blank, then if the label is "aaa-aaaabb", it will be mapped to "aab", and "aaaaaaabb" will be mapped to "ab". With the introduction of the blank mechanism, we can handle the problem of repeated characters very well. Similarly, "aaa-aaaabb" can be mapped to "aab", and similarly, "aa-aaaaabb" can also be mapped to "aab", there is a plurality of different character combinations that can be mapped to "aab".

For a label sequence, the conditional probability is the sum of all path probabilities mapped to it:

$$p(l|x) = \sum_{\pi \in B^{-1}(l)} p(\pi|x) \tag{1}$$

where $\pi \in B^{-1}(l)$ refers to all the sets of paths that can be merged into.

This way of passing the sum of the mapping B and all candidate path probabilities makes it unnecessary for the CTC to accurately segment the original input sequence, which makes it possible to perform task translation of sequence length > label length output by the RNN layer. The CTC can be used with any RNN model, but considering that the annotation probability is related to the entire input string, rather than only the segments of the previous small window range, the bidirectional RNN/LSTM model is more suitable. CTC calculates loss to find the character corresponding to the most likely pixel area. In fact, the computational essence of loss here is the induction

of probability. By calculating the probability, the previous neural network can be back-propagated and updated. Similar to the general classification, the loss function O of CTC is defined as the negative maximum likelihood. For the convenience of calculation, the logarithm of the likelihood is taken.

$$O = -\ln(\Pi_{(x,z)\in s} p(l|x)) = -\sum_{(x,z)\in S} \ln(p(l|x)) \tag{2}$$

The training goal is that the loss function O is optimized to a minimum.

3 Experiment

The experimental environment is ubuntu system, using NVIDIA 1060 GPU graphics card for training test, using Caffe deep learning framework, the experimental data set is the image of the 12,602 shell number, container box number and container box door collected by the camera at the port. And 11,000 images are used as a training set, and 1602 images are used as a test set, and the image is sent to the algorithm by a tool for training, and iteratively 3600 times.

The experiments in this article need to recognize the bay number. First, we need to know the specific position where the bay number appears in the image. This paper uses You Only Look Once (YOLO) [5] target detection algorithm to detect the position of the bet number in the image. YOLO [5] is a network for target detection. Its tasks include determining the location of certain objects in the image and classifying these objects. YOLOv3 [6] is an upgraded version of yolo target detection algorithm, it uses multi-scale prediction, and adopts feature maps of different scales to extract image features to adapt to different sizes of objects and uses a better basic classification network (ResNet-like) and classifiers, using multiple logistic classifiers instead of the softmax classifier solves the problem of multi-label classification. In this paper, a YOLOv3 [6] model is trained by using the captured bay number image, which is used for detection and lays a foundation for subsequent image sequence recognition.

In Fig. 4, the left and right sides of the shell number are all related to each other. For example, the right side of the 03th is the 02 number, and the left side is the 04 number. The shell number can be used for positioning, and the detection result of the YOLOv3 is fed back to the CRNN.

After getting the detection frame given by YOLOv3, we can deduct the image in the detection frame to identify the image sequence. We tested on the 1668 bay image, the test results are shown in Table 1.

In Table 1, 1668 images were tested, the number of positive samples correctly classified by True Positive (TP) is 1648 images, True Negative (TN), the number of samples that misclassified the correct sample is 20, the accuracy was 98.8%, and Frames Per Second (FPS) represented 250 images per second was processed. The specific recognition effect diagram is shown in Fig. 4:

Fig. 4 Bay number recognition results

Table 1 Bay number identification results

Test images	TP	TN	Accuracy	FPS
1668	1648	20	98.8%	250.6

The left and right columns in Fig. 4 are the detection effects from different scales of the bay number, where the red box is the box closest to the center point, that is, the box that needs to participate in the positioning, and it can be seen that the identified image sequence has been displayed on the image, upper right, the image of the corner is framed by the light, even if the human eye is already unrecognizable, and the image recognition error. The image in the lower right corner has a shell number that has disappeared into the field of view. At this time, the image recognition result is given 0, and the result is actually correct.

4 Conclusion

In this paper, the CRNN algorithm is used to realize the real-time recognition of the shell number. Firstly, the detection result given by YOLOv3 is used as the input image of the CRNN algorithm. The CRNN algorithm has end-to-end characteristics, which is not limited to any predefined vocabulary, and significant performance can

be achieved in both no-dictionary and dictionary-based scene text recognition tasks, and only a small parametric model needs to be trained.

Acknowledgements This work was financially supported by fund project, that is, Guangzhou Institute of industry and commerce college level research project in 2019 "Research and design of S band radio frequency front-end" 2019KQNCX231.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **25**(2), 1097–1105 (2012)
2. Graves, A., Liwicki, M., Santiago, F., et al.: A novel connectionist system for unconstrained handwriting recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 855–868 (2009)
3. Su, B., Lu, S.: Accurate scene text recognition based on recurrent neural network. In Asian Conference on Computer Vision, Springer, Cham, pp. 35–48 (2014)
4. Shi, B., Bai, X., Yao, C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Trans. Pattern Anal. Mach. Intell. **39**(11), 2298–2304 (2015)
5. Redmon, J., Divvala, S., Girshick, R. et al.: You only look once: Unified, real-time object detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788 (2016)
6. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement (2018)

Deep Reinforcement Learning-Based Resource Allocation for Smart Grid in RAN Network Slice

Mingyue Liu, Yang Wang, Sachula Meng, Xiongwen Zhao, and Suiyan Geng

Abstract Driven by the construction of the Ubiquitous Electricity Internet of things, various services have increasingly higher requirements for wireless communication indicators. The 5G network with low latency, large connection and large bandwidth is urgently needed to fundamentally meet various business requirements and network security requirements of the smart grid. Because of the type of service arrival of smart grid is unknown and lacks prior knowledge, reinforcement learning (RL) is used to conduct this research. Considering the differentiated service characteristics of the smart grid and the challenges of flexibility and adaptability of the communication platform, this paper aims to solve the resource allocation problem of the radio access network (RAN) slice of smart grid. In this paper, we firstly introduce three typical power services and propose a service priority concept. After reviewing the fundamental concepts and proving the convergence of RL algorithm, we propose a dynamic resource allocation strategy of the RAN slice for smart grid based on RL. Finally, simulation results prove that the proposed RL algorithm can achieve resource utilization and quality of experience (QoE) improvement against the fair allocation scheme.

1 Introduction

Driven by the ubiquitous construction of the electric power Internet of things, there is an urgent need to optimize and upgrade the power grid business. Emerging power businesses are booming, network security requirements are becoming higher and higher and various services are increasingly demanding wireless communication indicators. Combined with 230 MHz power wireless private network test and pilot

M. Liu (✉) · Y. Wang · S. Meng
China Electric Power Research Institute, No. 15, Xiaoying East Road, Haidian District, Beijing, China
e-mail: a18811593699@163.com

M. Liu · X. Zhao · S. Geng
North China Electric Power University, No. 2, Bei Nong Road, Changping District, Beijing, China

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_24

verification, the average transmission delay of the 230 MHz private network system is 50–100 ms, the terminal access rate is less than 2 Mbps and the single base station access capacity is 4000. The average transmission delay of the 4G public network is above 100 ms and the terminal access rate is less than 10 Mbps. Therefore, 230 MHz private network and 4G public network are not enough to support the mass connection, bandwidth access and low-latency service requirements per unit area, and cannot support the development of ubiquitous electric power Internet of things business. Although the optical fiber private network has the advantages of high bandwidth, low delay, and high reliability, it cannot meet the needs of ubiquitous, flexible, and mobile access. In order to respond to the transformation of energy production and consumption to promote the upgrade of the power grid to the energy Internet, it is urgent to build a 5G network with low latency, large connections, and large bandwidth to fundamentally meet the various business needs and network security requirements in the construction and development of smart grids. Then, it can form a new business format for the power grid, promote the improvement of the quality and efficiency of the energy ecosystem, reform the business model and ensure that the power network is safe and autonomous.

The 5G network slice divides the existing physical network into service logic to form an independent business logic network, and provides personalized customized services for differentiated services. According to different business service quality requirements, the industry application of 5G network architecture can be realized by allocating corresponding network resources and network functions. 5G network slicing technology can provide a more efficient and flexible wireless network solution for vertical fields. Smart grid is a representative of a typical vertical industry. 5G network can build a secure and reliable private network for smart grids that can provide differentiated needs.

From the business deployment analysis of smart grids, 5G network slicing can enable grid control, acquisition and mobile services, thus enabling multi-service network slicing hybrid networking and unified operation and maintenance management. Network slicing technology can effectively help power companies to reduce capital expenditures and operation and maintenance costs, and provide a more efficient business management basic communication network. For the smart grid scenario, the application of 5G network slicing technology for resource allocation research is still in the development stage. Smart grid allocates and manages resources through two-way interaction with application devices. Since there are many smart grid users and a large number of demand requests for managing a large number of devices, the demand for resources is not certain, that is, each service flow has its own bandwidth requirements, reliability requirements, and delay tolerance. Therefore, under the rapid development of the new generation of power grids, in the face of the variety of smart grid services, how to efficiently manage resources, minimize network deployment costs, adapt to the diverse needs of multiple types of power services, and obtain satisfactory service quality. It is of great practical significance for the application and popularization of the ubiquitous power Internet of things.

There are very few related tasks in implementing network slice resource management for smart grids. In [1], the trend and feasibility of applying network slicing to smart grids are presented. Saleem et al. [2] analyze typical business scenarios and communication needs of smart grids and application research of 5G network slice in smart grid. The radio access network slice is part of the end-to-end network slice, and the radio resource allocation is an important part of the radio access network slice. There are many resource allocation methods currently. In [3], some scholars have proposed a semi-static resource allocation scheme based on proportional fairness algorithm for the dynamic optimization of network slicing resources. However, the algorithm focuses more on fairness than performance, so there is room for improvement in resource utilization. In [4], to improve the utilization of spectrum resources, an opportunity spectrum sharing with each other algorithm is proposed. However, this literature only considers the utilization of spectrum resources without considering the QoE of users. QoE is a subjective assessment of quality of users and has recently become an essential indicator in 5G wireless communication systems [5]. By taking advantage of non-orthogonal multiple access (NOMA) features, [6] establishes the potential of QoE-based resource allocation in multi-cell NOMA networks. The above article only considers resource utilization or user experience. Therefore, this paper will improve resource utilization while meeting user experience quality.

At present, the reinforcement learning strategy is a popular technology widely used in many wireless resource allocation scenarios, such as multi-beam satellite systems [7], content-centric Internet of things (IoT) [8], and millimeter-wave systems [9]. There are many smart grid users, managing a large number of devices will generate a large number of demand requests, so the demand for resources is not certain. Each service flow has its own bandwidth requirements, reliability requirements, and delay tolerance. Therefore, considering diverse types of smart grid services with the rapid development of the new generation of power grids, we propose a deep reinforcement learning (DRL)-based RAN slicing strategy to realize resource allocation of the smart grid. The main contributions of this paper are as follows:

The power business is divided into three categories with different priority based on the application scenario of the grid business. The state, action, and reward in the power grid scenario are analyzed. The mapping of RAN slice of resource management for smart grid to deep Q-learning algorithm is analyzed.

We briefly introduce the reinforcement learning algorithm and the Q-learning algorithm is introduced in detail.

In dynamic resource management for multiple RAN slices of power grid, we propose a deep reinforcement learning algorithm which adjusts resource allocated to slices based on the corresponding QoE and spectral efficiency (SE).

The remainder of this paper is organized as follows. The system model is introduced in Sect. 2. As the main part of the paper, Sect. 3, we review the fundamental concept and convergence of reinforcement learning algorithm at first and propose slice allocation strategy based on reinforcement learning. The evaluation and simulation of the proposed method is given in Sect. 4. We conclude this work in Sect. 5.

2 System Model

This section is divided into three subsections, system description, power service, and system model. The power business and system model are introduced in detail.

2.1 System Description

Figure 1 shows the basic structure of the RAN network slice in the smart grid scenario including the service plane, the orchestration control plane, and the data plane. Network slicing achieves control plane and data plane decoupling of the network by means of software defined network (SDN) technology. And the open interface is defined between them to implement a flexible definition of network functions in the network slice. To meet the needs of this type of service, network slices only contain

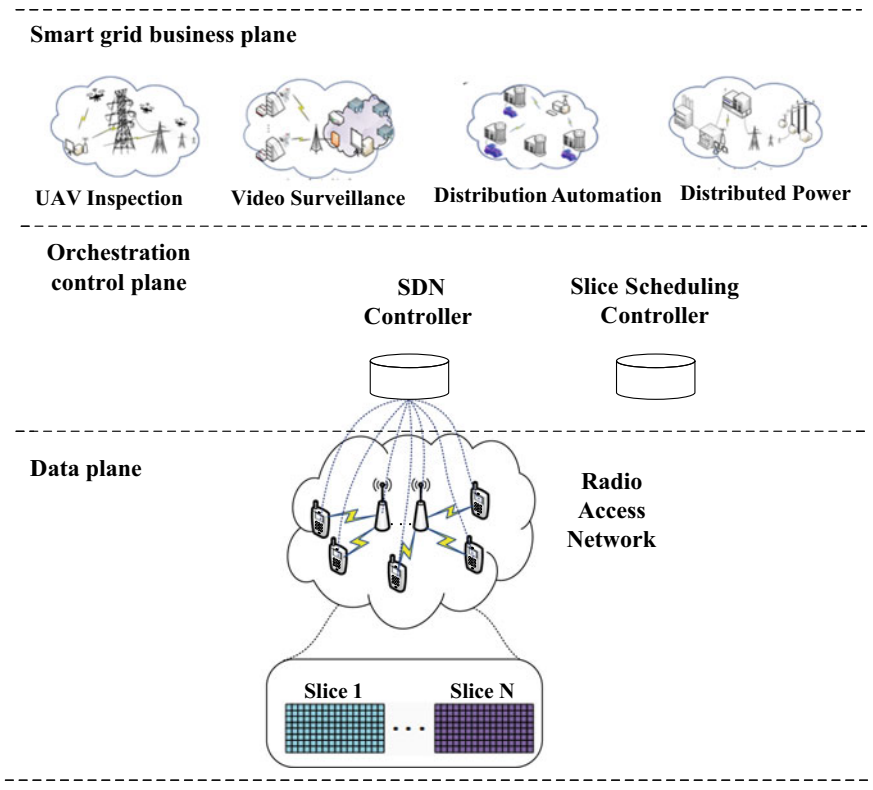


Fig. 1 RAN architecture under smart grid

network functions that support specific services. In this paper, we mainly consider the service plane and control plane.

The service plane mainly includes three major categories of power services: control power services, information collection services, and mobile applications power services. Control power services mainly include distribution automation, accurate load control, distributed power, and so on. The information collection services involve electric information collection, video monitoring and so on. Mobile applications power services mainly include unmanned air vehicle (UAV), mobile officing, etc.

Different services have different quality of service (QoS) requirements like delay, rate, reliability, etc. Therefore, different power services should have different priorities.

The access network orchestration control plane is mainly composed of a slice scheduling controller and an access network SDN controller. The slice orchestration controller completes the division of the RAN slice network according to various types of prior knowledge of the service type, channel conditions, and user requirements of the service plane. The SDN controller is mainly responsible for network function management and coordination of the access network. The SDN controller of access network manages the RAN network slice and execute corresponding algorithm, which is the resource allocation strategy for smart grid based on reinforcement learning proposed in this paper.

2.2 Service Introduction

With the continuous deepening of the market-oriented reform of power companies and the regulation of large-scale new energy integration, the power grid is developing in the direction of smart grid and energy Internet. Demand for control, information collection, and mobile applications will show more diversified characteristics. Different services have different network requirements, and new requirements are put forward for resource allocation schemes in smart grids.

The control business is related to the safe and stable operation of the power grid. The service is mainly used for fault isolation, self-healing control, and precise timing and has typical low-latency, high-reliability service features and 30 ms-1 s delay requirement. Take distribution automation as an example, distribution automation requires the realization of remote control reliability to support the development goal of grid power supply reliability of not less than 99.999%. In the future, it is necessary to support the differential protection. The connection mode will adopt a distributed point-to-point connection, coexist with the substation/master station mode, the main station sinks, the local proximity control, and the delay requirement reaches the millisecond level. The existing 4G public network and 230M private network wireless communication mode cannot meet the needs of millisecond low-latency services such as distribution automation.

Information collection business involving mass communication terminals is widely distributed in all links of the power grid. The frequency of future acquisitions tends to be at the minute level, and the number of connections is at least doubled. Taking electricity information collection as an example, in the future, the development will be mainly in the direction of increasing frequency of acquisition, rich content collection, and two-way interaction. There are about two controllable household appliances per household. There are 1000 electric meters connected per square kilometer, and about 2000 households connected per square kilometer. According to the frequency of acquisition, the frequency of acquisition is doubled, and the maximum rate of concurrent acquisition by the user-side terminal per square kilometer: $1000 * 560 * 2 + 2000 * 560 * 1/3$ kbps $= 1.23$ Gbps. The existing 4G public network and 230M private network wireless communication mode cannot meet the requirements of the large-scale connection and large bandwidth of the power information collection service.

The mobile application business has obvious requirements on network bandwidth and mobility, and has typical large bandwidth and flexible access service features. The bandwidth requirement of a single terminal must not be lower than 2 Mbps. Taking the integrated digital management of the whole process of infrastructure as an example. Video surveillance is measured according to the requirements of 48 channels of video surveillance at the headquarters command center. Minimum bandwidth requirement is about 2 Mbps $* 48 =$ 96 Mbps. Delay requirement is second. The bandwidth required for a single video-aware application is 20Mbps. To support the needs of video-aware applications, the headquarters ICT room is about $48 * 20$ Mbps $=$ 960 Mbps. Other non-video data has a bandwidth of about 10 Mbps, monitoring system bandwidth requirements are about 970 Mbps, and latency requirements are in milliseconds. The existing 4G public network, 230M private network, and fiber communication mode all cannot meet the requirements of the infrastructure digital service in millisecond delay and large bandwidth.

2.3 System Model

The system model of this paper is shown in Fig. 2. We mainly consider the bandwidth allocation of the downlink in the RAN network slices. In this paper, spectral efficiency and QoE are indicators of evaluation. In the network, A set of slices is denoted as $M = \{1, 2, \ldots, M\}$ and a set of total users is denoted as $K = \{1, \ldots, K\}$. A set of users of a specific slice m is denoted by K_m, and k_m denotes a single user of the slice.

Reinforcement learning is a mapping from environmental state to action. Therefore, we firstly define the state space, action space, and reward function of the

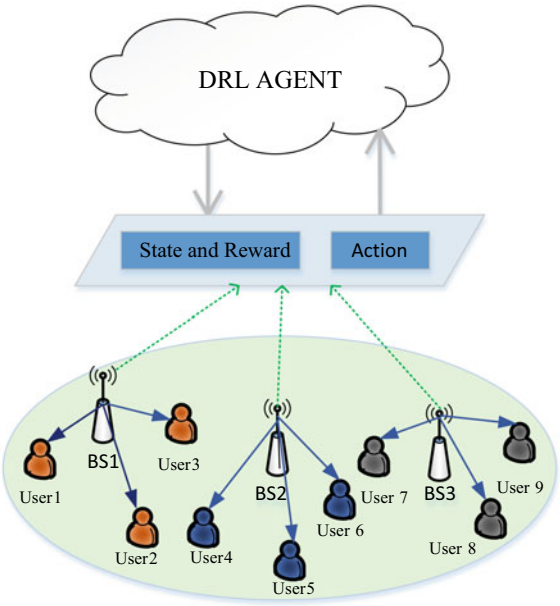


Fig. 2 An illustration of resource allocation for network slicing

DRL-based framework as follows.

State Space: the state of DRL agent consists of two components, the current state $s_m^n \in \boldsymbol{s} = (s_1, s_2, \ldots, s_M)$ of each slice m in the nth state and the demand $d_k \in \boldsymbol{d} = (d_1, \ldots, d_K)$ of each user k. s_m is used to indicate the current state of all slices that can be used to carry related power traffic. d_k represents the service flow of smart grid.

Action Space: the agent of reinforcement learning should provide suitable bandwidth resource to maximize the QoE requirements and SE when faced with unknown power services. The agent can decide what action to perform in the next state according to the current slice state and the reward function. We define the action space as $\{a_n\}$, where the nth element is a_n. Also, we define binary variables of $a_n \in \{0, 1\}$.

Reward: we mainly consider the downlink situation. The spectral efficiency of the system can be defined as:

$$\mathrm{SE} = \frac{R}{B}, \tag{1}$$

$$R = b_k \log_2\left(1 + \frac{P_n h_{k_m,n}}{\sum_{l \in N, l \notin n} P_l h_{k_m,l} + \sigma^2}\right), \tag{2}$$

where B is system bandwidth of BS n in Hz, R is transmission rate, $b_k \in \boldsymbol{b} = (b_1, \ldots, b_K)$ is the bandwidth allocated to user k. σ^2 denotes thermal noise power

at the user, $h_{k_m,n}$ denotes the channel gain, the transmit power of BS n is denoted by P_n.

Driven by the requirements of high quality power services like distribution automation, emergency communications, precision load control, etc., an appropriate level of quality of experience in smart grid is desired. In this paper, the QoE is measured by considering the effective throughput experienced by the users, normalized according to their maximum demanded data rate. With this target, the resources allocated to a service with low priority could be reduced, if needed, down to the minimum amount capable of meeting the basic QoS requirements in order to admit new services with higher priority.

We define the QoE of users:

$$\mathrm{Q} = \left(\frac{R}{R_{\mathrm{km}}^{\max}}\right)^{u}, \tag{3}$$

$$u1 + u2 + u3 = 1 \tag{4}$$

where R is transmission rate. $R_{\mathrm{km}}^{\max}$ is the maximum data rate constraint of the user in packets per second. u represents the priority of the users, which depends on the type of business it serves. As discussed above, we define the priority of the power service according to the respective QoS requirements. Figure 3 shows the priority level of the power services. The higher the priority the higher the u value. In this article, we define $u1$, $u2$, and $u3$ as the priority of control services, mobile application services and information services respectively.

The reward for learning agent is defined as follows

$$\mathrm{Reward} = \lambda \cdot \mathrm{SE} + \gamma \cdot \mathrm{QoE}, \tag{5}$$

where λ and γ are the weight of SE and QoE respectively.

We consider an access network scenario consisting of multiple base stations (BSs), where there exists a list of existing K users sharing the aggregated bandwidth B and

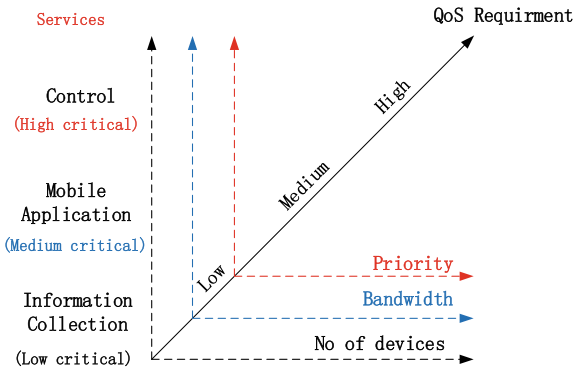


Fig. 3 Services allocation priorities

having fluctuating demands $\boldsymbol{d}$. We aim to maximize the expectation of the utility function E {Reward($\boldsymbol{d}, \boldsymbol{b}$)} by dynamically adjusting the allocated bandwidth.

$$\begin{aligned}
&\arg\max E\{\text{Reward}(\boldsymbol{d}, \boldsymbol{b})\} \\
&\quad = \arg\max E\{\lambda \cdot SE(\boldsymbol{d}, \boldsymbol{b}) + \gamma \cdot QoE(\boldsymbol{d}, \boldsymbol{b}) \\
&\text{s.t.:} \quad C1\text{: } 0 \le b < B \\
&\qquad C2\text{: } b_1 + \cdots + b_K = B \\
&\qquad C3\text{: } 0 < R \le R_{k_m}^{\max} \\
&\qquad C4\text{: } 0 < P_n \le P_{\max} \\
&\qquad C5\text{: } d_k \sim \text{Certain Traffic Model} \\
&\forall k \in [1, \ldots, K]. \qquad (6)
\end{aligned}$$

3 Slice Strategy Based on Reinforcement Learning

In this section, we introduce briefly reinforcement learning algorithm or more specifically Q-learning. Specially, we prove the convergence of the Q-learning. Then, we propose a DRL-based RAN slicing strategy to realize resource management of the smart grid.

3.1 Q-*Learning Algorithm*

The Q-learning algorithm is an enhanced learning method similar to the dynamic programming algorithm proposed by Watkins. Therefore, we firstly introduce the RL algorithm. RL is one of the main methods of machine learning and intelligent control in recent years. RL determines the set of behaviors an agent should take in the environment by maximizing the cumulative utility. The way in which the cumulative utility of the RL is calculated does not depend on past behavior, but on the future states, which means that future states will influence the choice of the current state. By reinforcement learning, an agent can know what action to take in a particular state. Markov decision process (MDP) can be used to model the RL problems. MDP is usually defined as a quad $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}\}$. $\mathcal{S}$ is a collection of all environmental states. $\mathcal{A}$ is a collection of executable actions for the agent.$\mathcal{P}$ is the state transition probability distribution function. $\mathcal{R}$ is the reward function. $r_t \sim \mathcal{R}(s_t, a_t)$ means the immediate bonus value obtained by performing the action of at the state s_t. Strategy π means a mapping from state to behavior, RL through the definition and the optimal value function to get the optimal strategy, the most common form of value function is given by Eq. (7):

$$V^{\pi}(s) = E_{\pi}\left(\sum_{i=0}^{\infty} \gamma^{i} r_i | s_0 = s\right). \tag{7}$$

This is a weighting and expectation for a set of utility functions, where γ is called a factor, which describes how important future utility is to current utility. With the definition of the value function, the optimal strategy becomes the maximized value function, which is Eq. (8)

$$\pi^{*} = \arg\max V^{\pi}(s), \quad \forall s \in S. \tag{8}$$

We can improve existing strategies based on the following principles: If the other behavior of the strategy π is kept the same, only change the action a to a' at the state s, then resulted in a new strategy π'. If the value function $V' > V$, then the strategy π is better than the strategy π'. We can get the optimal strategy through the dynamic programming algorithm π^{*}.

The Q-learning algorithm provides a learning ability for intelligent systems to select optimal actions using the sequence of actions experienced in the Markov environment without building an environment model.

We define the discount accumulation return expectation value $Q^{\pi}(s, a)$, whose expression is

$$Q^{\pi}(s, a) = E\{r_t + \gamma r_{t+1} + \gamma r_{t+2} + \cdots | s_t = s, a_t = a\}, \tag{9}$$

where $0 < \gamma < 1, \pi$ is the strategy, $Q^{\pi}(s, a)$ is the Q function that performs the action a at the state s. Then according to Bellman's optimality criteria, there is at least one optimal strategy in a single environment setting. Therefore, the state value function of the optimal strategy is given by get its optimal Q value: For a policy π, define Q values (or action-values) as:

$$Q^{\pi}(s, a) = \mathcal{R}(a) + +\gamma \sum_{y} P_{xy}[\pi(x)] V^{\pi}(y). \tag{10}$$

The state transition probability distribution function above depends on many factors, such as service type, the demand of delay and rate, traffic load, etc., so it is difficult for us to obtain at the RAN slice. Therefore, due to the absence of prior knowledge, the model-free reinforcement learning algorithm, without the need of the desired reward, is very suitable for the derivation of the optimal strategy. Therefore, in this paper, we chose deep Q-learning algorithm from various RL algorithms.

Mithchell has demonstrated that the Q-learning method can converge to the optimal strategy and optimal values in the deterministic Markov decision process as follows:

The cumulative reward value of the state action can be continuously updated, and its value can be accessed infinitely multiple times.

The step size (learning rate parameter) must be reasonable.

(1) The state transition of the environment must have Markovity.

The Q value function is expressed in the form of a lookup table.

If the above condition is satisfied, the action value function $Q(s, a)$ converges to the optimal action value function $Q^*(s, a)$ with probability 1 when the iteration is infinitely multiple times.

Proof Let the Q value function after the nth iteration be $Q_n(s, a)$, and the optimal action value function is Q^*s, a.

The definition of Δ_n is as follows

$$\Delta_n = \max_{s,a} \left| Q_n(s, a) - Q^*(s, a) \right| \tag{11}$$

The expression of Δ_{n+1} after the $(n + 1)$th iteration:

$$\begin{aligned}
\Delta_{n+1} &= \left| Q_{N+1}(s, a) - Q^*(s, a) \right| \\
&= \left| \left(r + \gamma^* \max_{a' \in A} Q_n\left(s', a'\right)\right) - \left(r + \gamma^* \max_{a' \in A} Q^*\left(s', a'\right)\right) \right| \\
&= \left| \gamma^* \left(\max_{a' \in A} Q_n\left(s', a'\right) - \max_{a' \in A} Q^*\left(s', a'\right)\right) \right| \\
&\leq \gamma^* \max_{a' \in A} \left| Q_n\left(s', a'\right) - Q^*\left(s', a'\right) \right| \\
&\leq \gamma^* \Delta_n
\end{aligned} \tag{12}$$

As for arbitrary (s, a), $Q_n(s, a)$ and $Q^*(s, a)$ have bounds, so Δ_n has bounds. It can be obtained from the Formula (11) that after k stages, since each state action pair is frequently accessed infinitely, the maximum error is $\gamma^k \Delta_n$. When $k \to \infty, 0 \leq \gamma < 1, \gamma^k \Delta_n \to 0$. That is, when the number of iterations n tends to infinity, the action value function $Q(s, a)$ eventually converges to the optimal action value function $\mathrm{Q}^*(s, a)$.

3.2 Autonomous Radio Resource Allocation

In this section, we present the Q-learning-based dynamic resource allocation of network slice strategy, which maximizes the total resource utilization while ensuring that the demands of each user are satisfied. The expressions of the state set, the action set, and the reward function are defined in the above.

In this paper, the RAN slice controller interacts with the wireless environment in a short discrete time period. As stated above, we want to maximize resource utilization while meeting the demand of diverse users. In the Q algorithm, it is to maximize the Q value, getting the $\pi^{\max}(s)$

$$\pi^{\max}(s) = \arg\max Q(s, a), \quad \forall s, \pi. \tag{13}$$

Through continuous iterative learning, based on the known information, the slice controller can get the optimal value of Q. The slice controller can choose any action randomly at every state. After each decision, Q value will be updated. This process can be expressed as:

$$Q(s, a) = Q(s_t, a_t) + \alpha\left[r_{t+1} + \gamma_a^{max} Q(s_{t+1}, a_{t+1}) - Q\left(s_t, a_t\right)\right], \tag{14}$$

where α is the learning rate. By continuously updating the value at multiple times while adjusting the values of the operations, it is guaranteed that $Q(s, a)$ eventually converges to the value of the optimal strategy, i.e., $\pi^{\max}(s)$.

The following algorithm is the resource allocation strategy of the network slice. At first, we set the value of Q to zero. Prior to the application of the Q-learning algorithm, we firstly assign certain bandwidth to different slices according to different power services requirement, in order to initialize the state of the slices. Existing radio resource slicing solutions use bandwidth-based provisioning to allocate radio resources to different slices. Since Q-learning is an online iterative learning algorithm, it performs two different types of operations. In exploration mode, the slice controller randomly selects a possible action to enhance its future decisions. In contrast, in development mode, the slice controller prefers the operations it has tried and found to be effective in the past. We assume that the slice controller in the state s explores with a probability ϵ and uses the previously stored value of Q with a probability $1 - \epsilon$.

Alogorithm 1

1. Initialize the replay memory dataset D with size of N.
2. Initialize the episode index t =0.
3. Initialize the evaluation network Q and the target network Q*
4. **For** episode=1, ..., M **do**
5. The DQL agent observes the current state s_l of the service, the status of the service type s_y, recorded as s_t.
6. Choose epsilon value ϵ.
7. **for** t = 1, ..., T **do**
8. Generate a random number π,$0 < \pi < 1$
9. **if** $\pi < \varepsilon$
10. Choose an action randomly.
11. **Else**
12. Input the state to the ANN and choose the action that has maximum Q-value.
13. **end if**
14. After executing the action a_t, the agent observes the reward $\mathcal{R}(s_t, a_t)$ and a new state $s_{t+1} = s_t$ for the system.
15. The agent stores the episode experience $e_t =< s_t, a_t, s'_t, \mathcal{R}(s_t, a_t) >$into D
16. The agent samples a minibatch of experiences from D and update the Q-value function by
$Q'(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\big(\mathcal{R}(s_t, a_t) + \gamma max Q(s_t', a_t') - Q(s_t, a_t)\big)$.
In cases where episode terminates at t, $Q'(s_t, a_t) = \mathcal{R}(s_t, a_t)$.
17. The episode index is updated by t ← t + 1.
18. Update the gradient descent based on Q′(s, a) -Q (s, a)
19. Update the Q′(s, a) - Q (s, a) each *N* step.
20. **end for**
21. **end for**

4 Simulation Results

To evaluate the performance of our proposed scheme, we perform numerical simulations and analysis using Python software. When the power business arrives, we assume that there are two base stations. In a coverage radius of 1000 m, the central coordinates of base station BS1 is [0, 0] and base station BS2 has a coverage radius of $R = 500$, the central coordinates is [500, 0]. The power services all obeys Poisson distribution, and the generation rates are expressed as follows $\lambda_c = 3.6$, $\lambda_i = 2.4$, and $\lambda_m = 3.0$. The generation coordinates of the service are generated randomly, and the access to which base station is decided according to the distance between the service and the two base stations. CSI between base station and equipment obeys Rayleigh distribution of variance $\sigma_r = 1.5$, channel noise obeys Gauss distribution of the mean value $\mu_g = 0$ and variance value $\sigma_g = 5$, channel number of base station BS1 is 20

and channel number of base station BS2 is 10. For convenience of research, the channel bandwidth is $b = 8$ MHz in Shannon's formula. Based on Q-learning, two DRL networks are constructed, namely deep Q-learning and double deep Q-learning. The main role of the former lies in the target network and experience playback. The main role of the latter is to improve the max action selection operation and solve the overestimation problem. The former has two neural networks, namely the two layer structure of the evaluate net and the two-layer structure of the target net, the latter having only one neural network, consisting of two layers of structure.

Since the application changes the location coordinates over time, the resulting application updates the coordinates within base station coverage at the rate of $upv = 3$. The base station generates a corresponding number of slices according to the access application. The maximum bandwidth $b_{\max} = 5$ MHz can be allocated in the mobile application services, which have higher demand for bandwidth. The unacceptable quality of service caused by the too small bandwidth can be ignored because the algorithm in this paper meets the reward maximization.

With priority of the three power services setting as $u1 > u2 > u3$, Figs. 4 and 5 illustrate the variation of SE and QoE along with each leaning episode when $\lambda < \xi$ and $\lambda > \xi$ respectively. As can be seen from the figures, larger QoS weights result in better QoS performance while at the same time causing some loss in SE performance. What is more, we compared the performance of the three algorithms. The first method is based on a fair allocation scheme, which distributes network resources to each network slice fairly and does not adjust according to changes in network traffic [10]. The latter two algorithms are enhanced double Q-learning and enhanced deep Q-learning. The algorithms realize dynamic allocation of network slice resources taking advantage of reinforcement learning, that is, the schemes described in this paper. As we can see from the figure, the algorithm of this paper can get better QoE and SE.

Figure 6 shows the cost values of the algorithm. The abscissa represents the number of iteration steps, and the ordinate denotes the cost value. The algorithm generates a large cost in the initial stage, but with the algorithm further updating, the cost gradually converges and eventually falls to zero. Comparing the four curves, it can be seen that the enhanced double Q-learning has the fastest cost convergence and smallest cost value in the case of $u1 = u2 = u3$.

Figure 7 shows the reward function values of the algorithm. In the figure, two types of the two algorithms are compared. As can be seen from the figure, in the case of $u1 > u2 > u3$, the reward value of enhanced double Q-learning is greater than the other three cases, but the difference in reward values for the four cases is small.

5 Conclusions

In this paper, we proposed a dynamic resource allocation strategy based on reinforcement learning algorithm of networks slice for smart grid. The algorithm achieves joint optimization of resource utilization and user experience quality. According to

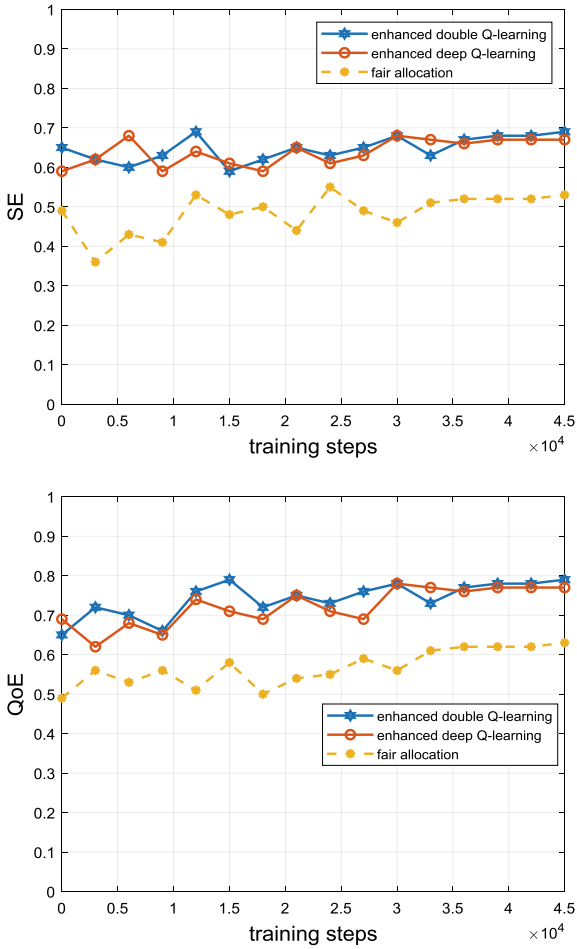


Fig. 4 SE and QoE versus the number of iterations when $\lambda < \xi$

the QoS requirements of different power services, the priorities are determined to achieve a relative maximization of QoE. By continuously updating the parameters and thresholds of the *Q*-learning algorithm, the system benefits can be maximized, and the cost of the system gradually converges to zero. By adjusting the specific gravity coefficients of QoE and SE, the degree of optimization can be selected. Through the simulation results, we can find that double *Q*-learning algorithm has the fastest cost convergence and larger reward value. Therefore, the proposed algorithm can better meet the resource allocation requirements of the smart grid on the RAN side when the neural network layer is double. In the future research, the parameters of the *Q*-learning network are improved, so that the algorithm can converge more quickly, and the quality of service requirements for power services is evaluated in multiple standards.

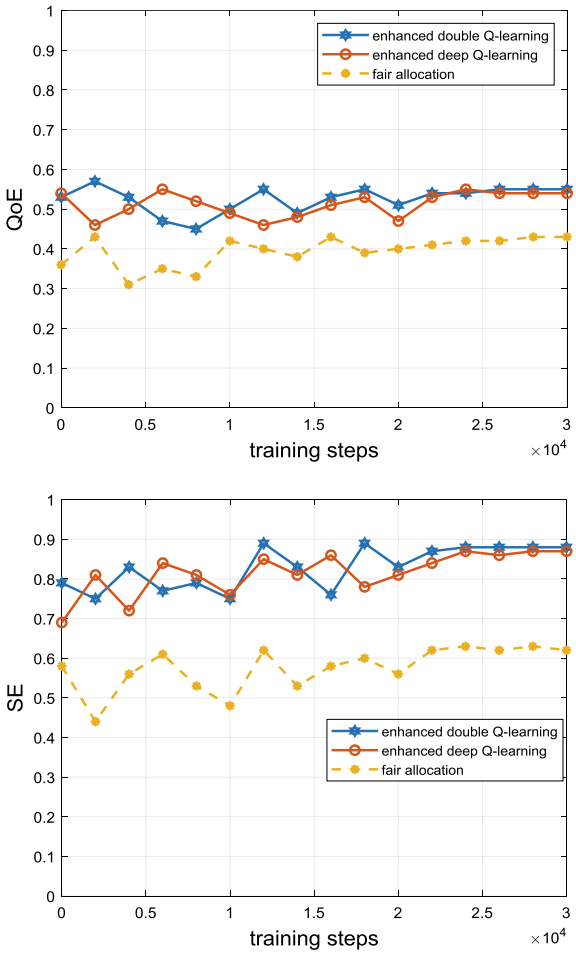


Fig. 5 QoE versus the number of iterations when $\lambda > \xi$

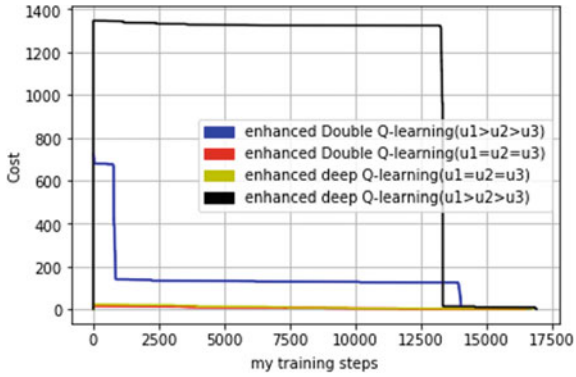


Fig. 6 Training cost versus the number of iterations

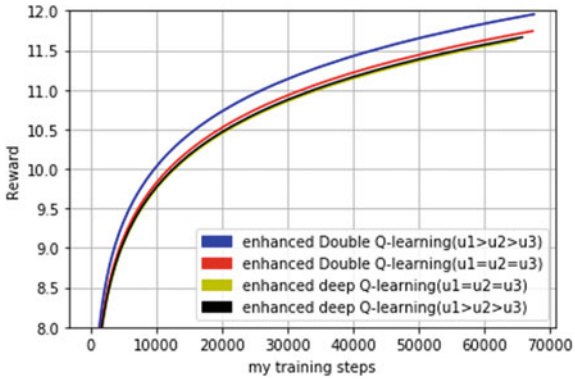


Fig. 7 Reward versus the number of iterations

Acknowledgements This paper is supported by "Project Evolution of Power Wireless Private Network and Application Analysis of 4G and 5G Technologies (5442XX190007)."

References

1. Dorsch, N., Kurtz, F., Wietfeld, C. (2018). On the economic benefits of software-defined networking and network slicing for smart grid communications. NETNOMICS Econ. Res. Electron. Network. 1–30
2. Saleem, Y., Crespi, N., Rehmani, M.H.: Internet of Things-aided smart grid: Technologies, architectures, applications, prototypes, and future research directions. IEEE Access **7**, 62962–63003 (2019)
3. Caballero, P., Banchs, A., Veciana, G.D.: Network slicing for guaranteed rate services: Admission control and resource allocation games. IEEE Trans. Wireless Commun. **17**(10), 6419–6432 (2018)
4. Liu, X., Li, M., Song, M. (2016). Wireless virtual network embedding based on spectrum sharing allocation. In: IEEE 11th International Conference on Computer Science & Education (ICCSE), Nagoya, pp. 670–675
5. Wang, W., Liu, Y., Luo, Z.: Toward cross-layer design for non-orthogonal multiple access: A quality-of-experience perspective. IEEE Wirel. Commun. **25**(2), 118–124 (2018)
6. Cui, J., Liu, Y., Ding, Z.: QoE-based resource allocation for multi-cell NOMA networks. IEEE Trans. Wireless Commun. (2018)
7. Xin, H., Liu, S., Rong, C.: A deep reinforcement learning-based framework for dynamic resource allocation in multibeam satellite systems. IEEE Commun. Lett. **22**(8), 1 (2018)
8. He, X., Wang, K., Huang, H.: Green resource allocation based on deep reinforcement learning in content-centric IoT. IEEE Trans. Emerg. Top. Comput. (2018)
9. Feng, M., Mao, S.: Dealing with limited backhaul capacity in millimeter-wave systems: A deep reinforcement learning approach. IEEE Commun. Mag. **57**(3), 50–55 (2019)
10. Peyman, T., Khosrow, S., Kiseon, K.: A fair and efficient resource allocation scheme for multi-server distributed systems and networks. IEEE Trans. Mobile Comput. **15**, 2037–2050 (2015)

Identification of e-Nose Wine Varieties Based on Feature Extraction and Classification Modeling

Miao Qiao and Lei Zhang

Abstract A new method for rapid identification of wine varieties was established by using transducer electronic nose detection system. System performance: firstly, the electronic nose with ten different sensors is used to collect the information of seven kinds of grape wine, then the sensor information of each kind of grape wine is extracted by polynomial curve fitting method, then the qualitative analysis of different kinds of grape wine is realized by principal component analysis, and finally, the classification algorithm of grape wine is combined with lightGBM algorithm. At the same time, Bayesian optimization is used to optimize the super parameters of the model. Through the classification model in the validation set, the accuracy of the lightGBM algorithm reached 95.24%, achieving the best classification effect, and verifying the effectiveness of the new pattern recognition method, which provides a new idea for wine classification and identification.

1 Materials and Methods

1.1 Materials

The wine samples are provided by COFCO Huaxia Great Wall Wine Co., Ltd. (Qinhuangdao, Hebei Province, China), including seven kinds of wine samples (300 bottles of each type), which are different varieties of wine like Cabernet Sauvignon, matherland, longzhibao, Merlot, Chardonnay, little Munson, pinli pearl, etc., in 2018 in China.

M. Qiao (✉) · L. Zhang
College of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300401, China
e-mail: 314793911@qq.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_25

1.2 Instruments and Equipment

In the experiment, the transducer portable electronic nose made by air-sense company of Germany is used [1]. The electronic nose system consists of metal oxide gas sensor matrix, gas acquisition device, and signal processing unit. The sensor matrix contains ten different metal oxide sensors, each of which has different aroma types and sensitivity. When the sample volatiles enter the collection system from the injection tube, they contact with the heated metal oxidation sensor matrix, and their resistivity G changes, and the ratio G/G_0 (relative conductivity) of the initial resistivity G_0 changes accordingly [2]. When the concentration changes greatly, G/G_0 deviates from 1 (greater than or less than 1). If the gas concentration is lower than the detection limit or there is no inductive gas, it is close to or equal to 1.

1.3 Experimental Method

1.3.1 Test Process

The sensor matrix in the PEN3 electronic nose equipment has high requirements for the detection environment. The pretest of the equipment is required before the test. In the preexperiment, the parameters in the process of electronic nose detection were selected. Through the experiment, the detection conditions were: the headspace volume was 500 ml, the headspace time was 10 min, and the ambient temperature was controlled. The specific detection process is as shown in Fig. 1: put 50 ml wine into a small bottle and seal it with a fresh-keeping film; allow it to stand and balance with the air in the small bottle for 10 min, so that the sample gas can be fully volatilized in a closed beaker, and conduct the formal experiment after the gas reaches a saturated and stable state. Before the gas collection, clean the air chamber and airway of the electronic nose with 500 ml of clean air treated by activated carbon, and the cleaning time is 60 s. During detection, insert the air inlet needle and air replenishing needle into the beaker sealed by the fresh-keeping film at the same time, the internal air pump

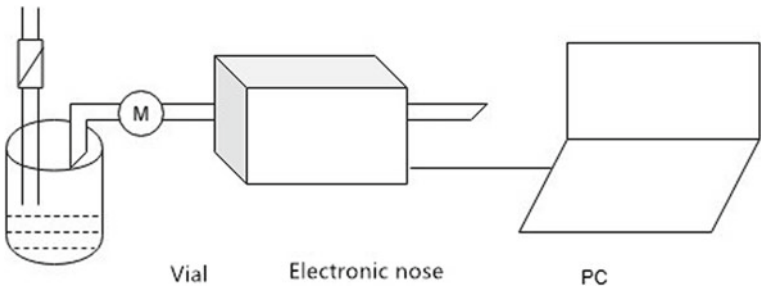


Fig. 1 Test diagram of electronic nose pair sample

of the electronic nose starts to work, and sucks the sample gas with 300 ml/min; the collection time is 90 s. To avoid an accidental error caused by human operation so that to ensure the accuracy and reliability of the sample, the same sample is tested three times. The collected gas information is saved to the computer in text form for subsequent data analysis and processing.

1.4 Data Processing and Analysis

1.4.1 Bayesian Optimization

There are two important steps in Bayesian optimization. First, select a prior function to map the super parameter λ to the loss function L, and select tree Parzen estimator (TPE) for this step [3]. Secondly, we need to select an acquisition function and construct an effective function from the posterior distribution of the model to determine the next sampling set [4]. In this step, the expected improvement (EI) is selected.

In order to improve the efficiency of searching the optimal super parameter set, an optimal agent model is established, and the ten times cross-loss value of the real lightGBM model is used for evaluation. TPE uses the Parzen window density estimator to generate the probability model [5] that divides the initial observation value into two groups, and then adds the multiple Gaussian probability distributions generated at each point to obtain the probability density and loss value performance index of each group. More than 20% of the corresponding points are divided into $l(\lambda)$ groups. The value of 20% of the performance index is represented by c^*, and the rest are classified into $v(\lambda)$ groups [6]. The formula is as follows:

$$p(\lambda|c) = \begin{cases} l(\lambda), c < c^* \\ v(\lambda), c \geq c^* \end{cases}$$
$$c : \text{magnitude of the loss}, \lambda : \text{Super parameter} \tag{1}$$

After defining the proxy function, select the next group of optimal-desired super parameters by maximizing the acquisition function [7]. The acquisition function uses the probability density function to generate the candidate options [8]. The expected improvement EI value of each sampling point is calculated by Eq. (2):

$$\text{EI}(\lambda) = \int_{-\infty}^{c_{\min}} (c_{\min} - c(\lambda)) p(c|\lambda) \mathrm{d}c \tag{2}$$

$c_{\min}$ represents the current minimum loss value, $c(\lambda)$ represents the loss value under the super parameter, λ.

1.4.2 Construction of Bo-LightGBM

As a machine learning algorithm, lightGBM, has many super parameters, the setting of super parameters will have a serious impact on the prediction ability [9]. This paper uses Bayesian super parameter optimization algorithm to adjust the super parameters of lightGBM. The Bayesian optimization uses the probability model to represent the complex optimization function. The priori of the objective to be optimized is introduced into the probability model [10]. The model can effectively reduce unnecessary sampling and is a search method considering historical parameters. The steps are as follows:

Generate observation range through random search: $H_t = \left\{\lambda^{(d)}, C(\lambda^{(d)})\right\}_{d=1}^{t}$;

When $t < T$ circulates;

According to formula (1), it is divided into H_t two groups;

$l(\lambda)$, $v(\lambda)$ are defined as the probability density estimation of Gaussian kernel function of all points in each group;

Maximize formula (2) to generate the candidate sets: $\lambda^* = \text{argmaxEI}(\lambda)$;

Bring the candidate set into the lightGBM model for ten times cross-validation search loss value $c(\lambda^*)$;

The observation range is updated by super parameter setting and corresponding loss value: $H_{t+1} = H_t \cup (\lambda^*, c(\lambda^*))$;

Cycle termination;

Return the optimal parameter H of the minimum loss value in λ_{best}.

2 Results and Analysis

2.1 Response Signal Curve of the Electronic Nose

The sample data is obtained by the electronic nose instrument, and the curve of sensor data of a typical sample is shown in Fig. 2. The horizontal axis represents the sensor response time, the vertical axis represents the sensor conductivity, and the curve represents the change of conductivity of each sensor in time when the volatile of wine reaches the measuring room. In the process of 90 s detection, the response value of the sensor rises suddenly and deviates from the original baseline. With the extension of detection time, the response value of the sensor basically reaches a stable state.

2.2 Qualitative Identification of Wine Based on PCA

PCA analysis was carried out on the electronic nose test data for different varieties of wine by using the pycham software platform [11]. The principal component analysis

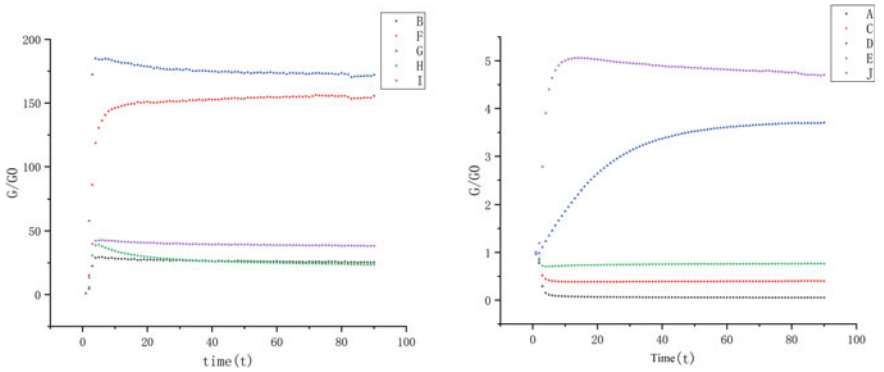


Fig. 2 Sensor response diagram figure

diagram is shown in Fig. 3, in which the contribution rate of the first principal component, the second principal component, and the third principal component are 58.27, 25.60, 5.13%, and the total contribution rate is 89%, which proves that PCA can well explain the information after the electronic nose curve fitting. It can be seen from the figure that different varieties of wine are distributed in different regions, and the PCA method can distinguish different varieties of wine intuitively and qualitatively.

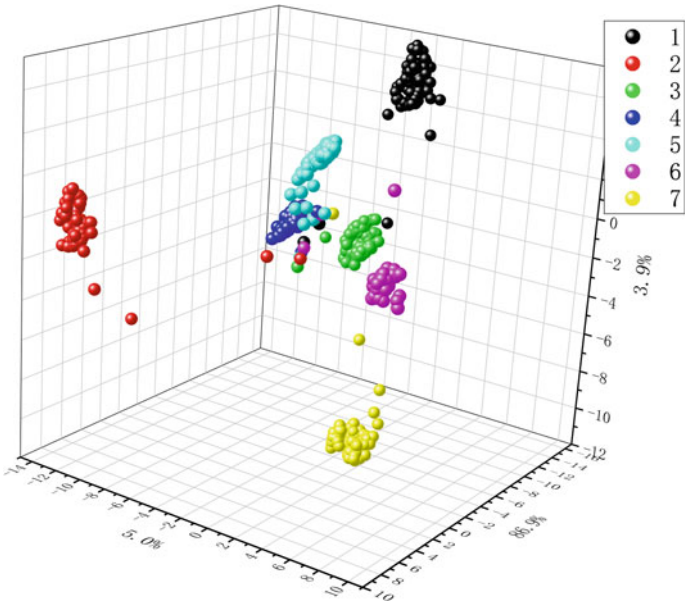


Fig. 3 PCA results of different wines

Table 1 Comparison table of classification model results

Label	SVM	RF	AdaBoost	BO-light GBM
1	157/180	162/180	167/180	170/180
2	150/180	169/180	160/180	171/180
3	163/180	160/180	164/180	173/180
4	160/180	158/180	170/180	172/180
5	158/180	167/180	159/180	174/180
6	161/180	166/180	161/180	170/180
7	151/180	158/180	169/180	170/180
Accuracy (%)	87.30	90.47	91.27	95.24

2.3 Bo-LightGBM Wine Quantitative Model

In this experiment, seven kinds of wine samples, 30 bottles of each kind, 210 samples in total, were taken three times for each sample, forming 6300 groups of samples. The 6300 $\times$ 3 sample data is randomly divided into two parts, 80% of which is used as training data set, 20% of which is used for test data to judge the model classification performance. The parameter expression and the value range: Learning rate from log(0.01) to log(0.2), maximum depth of tree from 1 to 10, leaf number from 5 to 15, and the number of iterations is set at 1000.

In the first 100 iterations, the error rate of Bayes shows a rapid decline trend. Due to many combinations of lightGBM super parameters, random search and grid search cannot quickly select the optimal parameters, so the error rate of Bayes continues to decline after 100 iterations. Finally, compared with the other two methods, the Bayes tone is more ideal.

Train support vector machine, random forest, AdaBoost model on the data set with default parameters, and Table 1 shows the comparison results of the three algorithms with BO-LightGBM. The results show that the improved lightGBM achieves the best classification effect in e-nose wine data.

3 Conclusion

To sum up, this study selects seven kinds of wine odor information, uses lightGBM algorithm to process the electronic nose data in the way of modeling and classification, and introduces Bayesian optimization to automatically adjust the super parameters of the model. Through the comparative experiment with random search and grid search methods, it shows that Bayesian optimization can quickly find the optimal super parameters, and has high accuracy than other search methods. Finally, through the comparison with the other four common modeling methods of electronic nose, through the final experimental results, it can be clearly seen that Bo-LightGBM

model method has higher accuracy compared with other classification algorithms, so it can be seen that the model proposed in this paper can provide some ideas for the application of electronic nose to identify wine types.

References

1. Guo, D., Yang, J., Li, B.: Analysis and design of visual olfactory system. J. Sens. Technol. **19**(3), 562–565 (2006)
2. Wang, L., Qu, J., Yang, J.: Developing electronic nose technology. Measure. Control Technol. **18**(5) (1999)
3. Jiaxu, Cui, Bo, Yang: Review of Bayesian optimization methods and applications. J. Softw. **29**(10), 3068–3090 (2008)
4. Bin, Li, Weixing, Wang: Improvement of classification model for NCA reduction and bayesian optimization. Computer application and software **08**, 281–287 (2019)
5. Shahriari, B., Saversky, K., Wang, Z.: Taking the Human Out of the loop: a review of Bayesian optimization. Proc. IEEE **104**(1), 148–175 (2015)
6. Snoek, J., Larochelle, H., Adams, R.P.: Practical Bayesian optimization of machine learning algorithms. In: International Conference on Neural Information Processing Systems, pp. 2951–2959. Curran Associates Inc. (2012)
7. Snoek, J., Rippel, O., Swersky, K., et al.: Scalable Bayesian optimization using deep neural Nehvorks. Statistics 1861–1869 (2015)
8. Yan, J., Tian, F., He, Q.: Feature extraction from sensor data for detection of wound pathogen based on electronic nose. Sens. Mater. **24**(2), 57–73 (2012)
9. Ke, G., Meng, Q., Finley, T.: Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Proc. Sys. 3146–3154 (2017)
10. Fridman, J.H: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), 1189–1232 (2001)
11. Pereira, A.C., Reis, M.S., Saraiva, P.M., Marques, J.C.: Aroma ageing trends in GC/MS profiles of liqueur wines. Analytica chimica acta **659**, 93–101 (2010)

Comparison of Contrast Enhancement Methods for Underwater Target Sonar Images

Linna Sun and Haitao Guo

Abstract The contrast enhancement of underwater target sonar images plays a key role in the smart perception of underwater targets and underwater smart services. In general, the contrast of sonar images is low, which is not conducive to the subsequent sonar image segmentation and target recognition. In the paper, the four selected methods for sonar image contrast enhancement are gray scale transformation, histogram equalization, unsharp masking and discrete wavelet transform. The four methods are contrasted in terms of contrast and signal to noise ratio of image.

1 Introduction

At present, people are constantly exploring the application field of smart perception, among which the application field of underwater robot is expanding, including ocean exploration and ocean development. The sonars installed on underwater robots can obtain sonar images for the smart perception of underwater objects. In general, the contrast of the sonar images is low. It is not conducive to the subsequent sonar image segmentation and target recognition. The sonar image contrast enhancement refers to improving the contrast of the image, making the dark region of the image darker and the bright region brighter, so as to highlight the target region, facilitating machine recognition and human eye observation. At present, there are many methods applied to image enhancement, among which gray scale transformation, histogram equalization and unsharp masking method [1] are traditional classical algorithms, but have their disadvantages. In the literature [2], an image enhancement method based on fuzzy gray scale transformation is proposed. According to the characteristics of poor contrast of sonar images, the traditional fuzzy enhancement algorithm is improved to achieve the effect of stretching gray scale range and improving the contrast of image.

L. Sun
Inner Mongolia University, Hohhot 010021, China

H. Guo (✉)
Hainan Tropical Ocean University, Sanya 572022, China
e-mail: ghtpaper@126.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_26

Literature [3] proposed a new gray scale enhancement method of sonar image. Firstly, the sonar image is divided into low gray scale region, medium gray scale region and high gray scale region by using the statistical feature of gray scale histogram. Then, the method of piecewise nonlinear enhancement is used to enhance the contrast of the sonar image. The multi-resolution analysis method based on the wavelet transform has become an effective way for image contrast enhancement [4]. Literature [5] used the discrete wavelet transform to enhance the contrast of sonar image. The sonar image contrast enhancement based on wavelet transform uses the wavelet technology to divide a sonar image into low-frequency and high-frequency sub-bands. Most of the information of the image corresponds to the low-frequency sub-band, and the detail information of the image such as edges and noise features corresponds to the high-frequency sub-band. Hence, different sub-bands are processed with different ways for image contrast enhancement.

Usually, the speckle noise is strong and the contrast is low of the sonar image, which is not conducive to the subsequent image segmentation and target recognition. Therefore, the sonar image contrast enhancement is a necessary step in the process of sonar image. This paper intends to investigate the application of gray scale transformation, histogram equalization, unsharp masking and discrete wavelet transform method in the contrast enhancement of sonar images.

2 Image Contrast Enhancement Method

This paper introduces four methods for image contrast enhancement, which are the gray scale transformation, the histogram equalization, the unsharp masking and the discrete wavelet transform.

2.1 Gray Scale Transformation

The gray scale transformation is a simple and classic method for image contrast enhancement. This method directly adjusts the gray value of pixels in the image. The grayscale piecewise linear transformation is adopted in this paper, and the formula is:

$$y = \begin{cases} (c/a)x & 0 \le x < a \\ [(d-c)/(b-a)](x-a)+c & a \le x < b \\ [(255-d)/(255-b)](x-b)+d & b \le x \le 255 \end{cases} \tag{1}$$

where x and y represent the gray value of pixel before and after the image contrast enhancement, respectively, a, b, c and d are constants. This function is used to convert the gray value of most pixels of the original image from the interval a to b to the

interval c to d, to improve the contrast of image. The gray value of pixels of the image are centrally distributed between a and b, and the values of a and b can be obtained on the gray value histogram. In addition, c and d can be adjusted according to actual needs.

2.2 Histogram Equalization

The histogram equalization is also a common method for image contrast enhancement. This method changes the histogram of the image into a form of uniform distribution. This increases the dynamic range of pixel gray values and enhances the contrast of image.

2.3 Unsharp Masking

The unsharp masking technique [1] was first used in photography to enhance the edges and details of an image. Its basic principle is: firstly, the fuzzy image is obtained by the process of fuzzy passivation (the equivalent of using a low-pass filter) for original image, then the fuzzy image is subtracted from the original image and the result is multiplied by the correction factor, finally, the original image is added to the above results, so that to improve the image high-frequency component, and enhance the image. The formula of the unsharp masking method can be expressed as:

$$f(i, j) = x(i, j) + C \times [x(i, j) - m(i, j)] \tag{2}$$

where C represents the correction factor, $x(i, j)$ represents the original image, $m(i, j)$ represents the fuzzy image, and $f(i, j)$ represents the enhanced image.

2.4 Discrete Wavelet Transform

The wavelet basis selected is represented by $\psi(x)$, and the wavelet function $\psi_{a,b}(x)$ is obtained after translation and expansion transformation for $\psi(x)$.

$$\psi_{a,b}(x) = |a|^{-1/2}\psi\left(\frac{x-b}{a}\right) \tag{3}$$

where $a, b \in R$ and $a \neq 0$ is the scale factor, b is the translation factor. The wavelet function and the scale function are related by the double scale equation, in accordance with Eq. (4):

$$\psi(x) = \sqrt{2} \sum_{k=-\infty}^{\infty} h(k)\varphi(2x - k), k \in Z \tag{4}$$

where $\varphi(\bullet)$ is the scale function, and the filter sequence $\{h(k)\} \in L^2(Z)$, $L^2(Z)$ is a finite space. The fourier transform of the scale function has the property of low-pass filter, while the fourier transform of the wavelet function has the property of high-pass filter.

Let the low-pass filter in the discrete wavelet transform be U and the high-pass filter be T. For the one-dimensional discrete wavelet decomposition of the signal, firstly, the signal passes through the low-pass filter U and the high-pass filter T, and then the low-frequency and high-frequency components are obtained by down-sampling operation. The discrete wavelet reconstruction of signal is the inverse process of decomposition. Firstly, the up-sampling operation is carried out for the decomposed low-frequency and high-frequency components respectively, then the reconstructed signal is obtained through low-pass filter and high-pass filter.

Suppose the original signal b_j, b_{j+1} represents the low-frequency component and c_{j+1} represents the high-frequency component, then the decomposition formula of the discrete wavelet transform of the signal is as follows [6]:

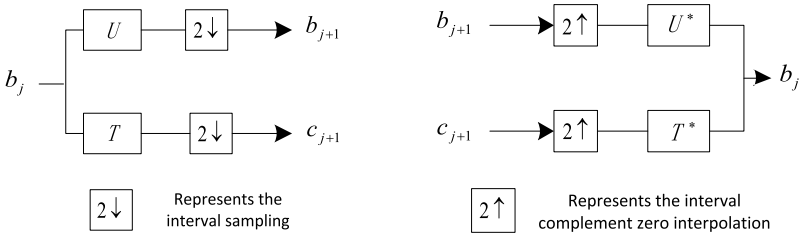
$$\begin{cases} b_{j+1} = D_\varepsilon U b_j \\ c_{j+1} = D_\varepsilon T b_j \end{cases} \tag{5}$$

where D_ε is the down-sampling operator. Then, the low-pass filter U^* and the high-pass filter T^* required in signal reconstruction are the dual operators of U and T, respectively. The discrete wavelet reconstruction formula of the signal is:

$$b_j = Z_\varepsilon U^* b_{j+1} + Z_\varepsilon T^* c_{j+1} \tag{6}$$

where Z_ε is the up-sampling operator. The decomposition and reconstruction process of the discrete wavelet transform of signals is shown in Fig. 1 [7]:

The above is the discrete wavelet transform of one-dimensional signal, while the present paper is concerned with the two-dimensional image. Just like the decomposition and reconstruction principle of the one-dimensional signal, the two-dimensional image is only divided into row direction and column direction, that is, firstly, the low-frequency and high-frequency components are obtained by the original signal passing through the low-pass filter and the high-pass filter in the row direction, and then, the obtained components in the previous step are passed through a low-pass filter and a high-pass filter in the column direction, respectively.



(a) Discrete wavelet decomposition of signals (b) Discrete wavelet reconstruction of signals

Fig. 1 Discrete wavelet decomposition and reconstruction of signals

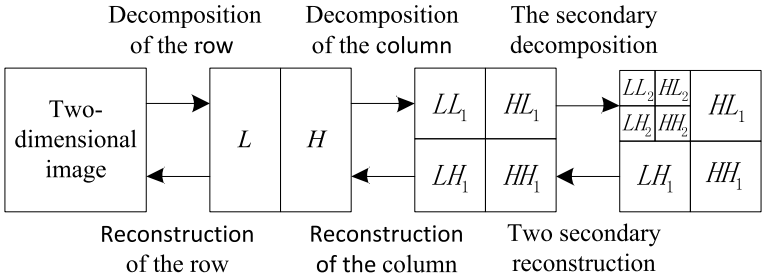


Fig. 2 Discrete wavelet decomposition and reconstruction process of a two-dimensional image

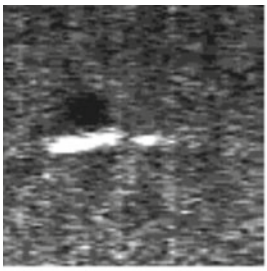
After the discrete wavelet transform, the original image will get the low-frequency sub-band LL, and the high-frequency sub-bands LH, HL and HH in the horizontal, vertical and diagonal directions. The decomposition and reconstruction process of two-dimensional images by discrete wavelet transform is shown in Fig. 2 [4].

The original sonar image was decomposed by discrete wavelet transform to obtain four sub-bands of LL, LH, HL and HH, namely low-frequency sub-band, horizontal high-frequency sub-band, vertical high-frequency sub-band and diagonal high-frequency sub-band.

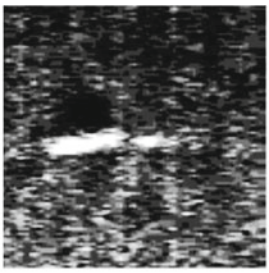
Wavelet transform is used to decompose the image into high-frequency sub-band and low-frequency sub-band. In this paper, the image contrast enhancement is realized by enhancing the low-frequency sub-band coefficients and decreasing the high-frequency sub-band coefficients. In this paper, the low-frequency sub-band coefficients is increased to 1.35 times, and the high-frequency sub-band coefficients is reduced to 0.75 times.

3 Experimental Results and Analysis

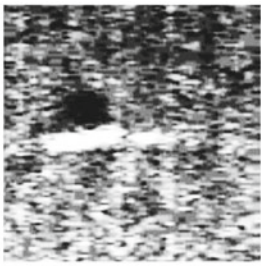
In this paper, four methods are used to enhance a sonar image: gray scale transformation, histogram equalization, unsharp masking and discrete wavelet transform.



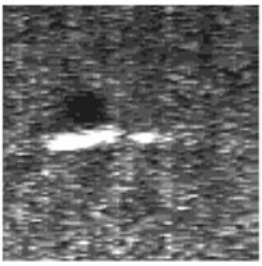
(a) Sonar image



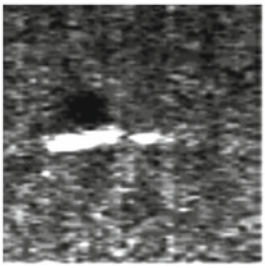
(b) Gray scale transformation



(c) Histogram equalization



(d) Unsharp masking



(e) Discrete wavelet transform

Fig. 3 Experimental results of sonar image contrast enhancement

The experimental results are shown in Fig. 3. Figure 3(a) is a sonar image of a manmade object. Figure 3(b)–(e) are the enhancement results obtained with various methods. For enhancement by the gray scale transformation, $a = 40, b = 130, c = 10, d = 200$. For enhancement by unsharp masking, the correction factor is $C = 2$. For enhancement by discrete wavelet transform, two-layer discrete wavelet transform is adopted. The Symlet wavelet (the compact-supported orthogonal wavelet with approximate symmetry) is selected.

In addition, in order to further compare the above methods, the contrast [8] and the signal to noise ratio (SNR) of image are selected as objective evaluation criteria.

The higher the contrast of image is, the more obvious the improvement degree of the contrast is. The higher the SNR of image, the less the image is affected by noise and the higher the image quality. The experimental data are shown in Table 1.

In terms of subjective evaluation, the results are shown in Fig. 3. The gray scale transformation method and histogram equalization method enhance the contrast of

Table 1 Objective evaluation indexes for the four image contrast enhancement methods

Evaluation indexes	Gray scale transformation	Histogram equalization	Unsharp masking	Discrete wavelet transform
Contrast	0.4306	0.3488	0.2972	0.2031
SNR	19.5690	21.6460	20.3337	25.8003

image obviously, but enlarge the noise in the image to a great extent. The unsharp masking method is not good at improve the contrast and suppress the noise of image. Discrete wavelet transform method is not ideal in improving contrast of image, but it is outstanding in suppressing noise. The data in Table 1 can reflect the effect of contrast enhancement to some extent.

4 Conclusion

The gray scale transformation method and histogram equalization method can not suppress the noise effectively while enhancing the contrast of sonar image. The unsharp masking method is not ideal in contrast enhancement and noise suppression. Discrete wavelet transform method is not so good at enhancing image contrast, but it can suppress the noise well, which is very helpful for sonar image processing in later stage. In the following research, the stationary wavelet transform can be used to enhance the image contrast. The stationary wavelet transform is an improvement of the discrete wavelet transform, which can effectively avoid the image edge distortion. Based on the stationary wavelet transform and combined with the corresponding image enhancement methods, the sonar image is enhanced to improve the overall visual effect of the image.

Acknowledgements We would like to thank professor Tian Tan for providing the sonar images.

This work is supported by the Institute and Local Cooperation Foundation of Sanya in China (No. 2019YD08) and the National Science Foundation of China (No. 61661038).

References

1. Schreiber, W.F.: Wirephoto quality improvement by unsharp masking. Pattern Recognit. **2**(2), 117–121 (1970)
2. Huang, J., Chen, X., Xu, W.: Research on underwater image enhancement technology based on fuzzy gray scale transformation. Appl. Sci. Technol. **45**(3), 1–6 (2018). (in Chinese)
3. Jia, Y., Ye, X., Guo, S.: A piecewise nonlinear enhancement method of side scan sonar images. In: Oceans 2019—Marseille, pp. 1–6. IEEE, Marseille, France (2019)
4. Dippel, S., Stahl, M., Wiemker, R.: Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform. IEEE Trans. Med. Imaging **21**(4), 343–353 (2002)
5. Priyadharsini, R., Sharmila, T.S., Rajendran, V.: Underwater acoustic image enhancement using wavelet and K-L transform. In: 2015 International Conference on Applied and Theoretical Computing and Communication Technology, pp. 563–567. IEEE, Davangere, India (2015)
6. Nason, G.P., Silverman, B.W.: The stationary wavelet transform and some statistical applications. Lect. Notes Statistios **103**, 281–299 (1995)

7. Gao, Q., Li, J., Xie, G.: Speckle reduction method of SAR image using stationary wavelet transform. J China Univ. Sci. Technol. **32**(5), 566–572 (2002). (in Chinese)
8. Zhan, B., Wu, Y., Ji, S.: Infrared image enhancement method based on stationary wavelet transform and Retinex. Acta Opt. Sin. **30**(10), 2788–2793 (2010). (in Chinese)

Region Contraction-Based Sparse Approach for Magnetocardiography Current Source Imaging

Lu Bing and Weiyuan Wang

Abstract Reconstructing the cardiac current source and its imaging with magnetic field data measured by the superconducting quantum interference on the surface of the human thoracic cavity is a new technique for non-invasive acquisition of information on cardiac electrical activity. Through the region contraction strategy, combined with the efficient sparse reconstruction algorithm which has been widely used in wireless communications, the accuracy of reconstruction is improved, and experimental support is provided for the subsequent development and practice of cardiac current imaging. The trajectory imaging of the reconstructed current source of magnetocardiography data is used to study the cardiac electrical activity conduction in the Q-R-S part of magnetocardiographic wave. Results for healthy people are given. The band trajectories initially reveal the conduction characteristics of ventricle electrical excitation in depolarization. This study aims to provide an applicable E-health smart service system with visual function of cardiac electrical activity.

1 Introduction

Heart disease is a major illness that harms human health. The proportion of deaths due to heart disease has also increased year by year. To this end, the early diagnosis and treatment of heart disease has been highly valued by countries around the world. The structure of the heart is very complex, and the heart movement is accompanied by changes in shape such as contraction, relaxation, and torsion [1]. Studying the quantification of cardiac electrical activity and its characteristics via magnetocardiography (MCG) has important theoretical significance and application value

L. Bing
Faculty of Business Information, Shanghai Business School, 123 Feng Pu Avenue, 201400 Shanghai, China

W. Wang (✉)
Shanghai United Imaging Healthcare Co., Ltd, 2258 Chengbei Road, 201807 Shanghai, China
e-mail: weiyuanw168@126.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_27

(E-health smart service system) for the analysis of cardiac function and diagnosis of heart disease [2, 3].

This paper focuses on the reconstruction accuracy and efficiency of sparse sources in MCG, and combines effective reconstruction algorithms to reduce the serious ill-posedness of reconstruction process. As is known, the feasible area is often used as a priori information for the reconstruction process, which helps to improve the accuracy and efficiency of reconstruction [4]. How to construct a suitable set of feasible areas becomes a challenge of MCG's inverse problem. This paper introduces a feasible region contraction strategy, so as to improve the accuracy of reconstruction. The experimental results show that the reconstruction quality and efficiency of the feasible region contraction strategy is better compared to traditional reconstruction algorithm. The method is used to analyze the cardiac magnetic data of normal people. Results indicate that this method provides spatiotemporal information regarding cardiac electrical activity during ventricular depolarization in healthy subject. This study aims to provide an applicable E-health smart service system with visual function of cardiac electrical activity.

2 Method

The linear relationship between the MCG data and the distribution of the internal target of the reconstructed object is established. The relationship between the density **Q** of the current source and the source weight matrix **L** is obtained [5]:

$$\mathbf{B} = \mathbf{L} \cdot \mathbf{Q} \tag{1}$$

Here, **B** is the detected MCG data. L_1 regularization combined with compressed sensing theory is introduced to optimize the ill-posedness of the target and (1) is transformed into:

$$\min_{\mathbf{Q}} \frac{1}{2} \|\mathbf{B} - \mathbf{L} \cdot \mathbf{Q}\|_1^2 + \lambda \|\mathbf{Q}\|_1 \tag{2}$$

where λ is the regularization parameter, which is often obtained through experience.

The inverse problem of the cardiac magnetic imaging is a mathematical problem where the number of unknown parameters is much higher than the number of equations. The solution is not unique and is susceptible to interference and noise, with serious morbidity. Setting the feasible area as the prior information of reconstruction may usually reduce the reconstruction range and the number of variables and relieve the ill-posedness of the problem to some extent. The steps of the feasible regional strategy are as follows [6]. (1) Divide the imaged objects to obtain *N* different network sizes. (2) N-group mesh obtained in step (1) is used as the inverse problem reconstruction network independently and reconstructed by fast bayesian matching pursuit

Table 1 Possible number of networks and reconstruction time for different network numbers ($N = 6$, $N = 12$, $N = 18$, $N = 24$)

N	X(m)	X(m)	X(m)	Time(s)
target	9.0	13.0	17.0	
6	(5.24,13.55)	(8.55,17.02)	(12.70,22.50)	82.5
12	(5.26,12.99)	(8.32,17.34)	(12.45,21.88)	118.9
18	(5.11,13.03)	(8.19,16.98)	(12.39,22.10)	256.7
24	(5.08,12.78)	(8.15,16.89)	(12.30,22.04)	467.2

(FBMP) algorithm which has been widely used in wireless communications [7], and the reconstruction results of N groups of different targets are obtained. The unit whose energy is greater than the set threshold is taken as the possible areas. (3) The intersection of the N feasible regions is taken as the a priori information of the multi-network statistics, and the reconstruction is performed by the FBMP algorithm $\min_{\mathbf{Q}} \frac{1}{2}\|\mathbf{B} - \mathbf{L}' \cdot \mathbf{Q}\|_1^2 + \lambda \|\mathbf{Q}\|_1$.

3 Simulation

3.1 Parameters

In theory, the more the number of splits, the more feasible is the range of feasible regions obtained by statistical multi-grid reconstruction and the reconstruction results are more accurate. However, if the number of grids is too large, it will directly lead to an increase in calculation time and reduces the feasibility. Table 1 lists the feasible areas and corresponding times after reconstruction using four different sets of N ($N = 6$, $N = 12$, $N = 18$, $N = 24$). Considering the accuracy of the feasible area and the feasibility of the reconstruction time, this paper sets the number of reconstructed meshes N to 12.

3.2 Noise Influence

We designed a set of simulation experiments to further prove the stability of the proposed reconstruction method by adding different degrees of Gaussian noise during the reconstruction process. The SNR levels are chosen to be 5, 10, 15, 20, 25, and 30 dB, respectively, to demonstrate the stability of the proposed method. The reconstruction results corresponding to different Gaussian noises are shown in Fig. 1. Under the influence of different levels of noise set, the reconstruction error remains between 1 mm and 22 mm. The results indicate that the FBMP algorithm combined

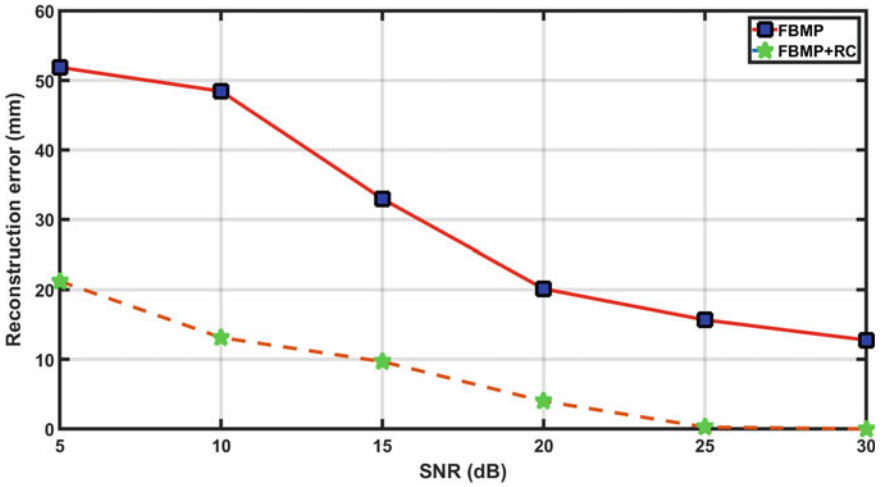


Fig. 1 Impact of SNR (levels are chosen to be 5, 10, 15, 20, 25, and 30 dB) on performance (RC: Region Contraction)

with the feasible region iterative contraction strategy is better, compared to traditional reconstruction algorithm.

4 Magnetic Imaging of Cardiac Electrical Activity

4.1 Settings

During the whole cardiac cycle, we selected 100 moments in the Q-R-S part of magnetocardiographic wave (QRS) as the analysis object. Using the trajectory imaging of the reconstructed current source, the projection of the cardiac electrical activity conduction in the QRS segment on the chest surface and the direction change of the current source in the cardiac cycle is studied. According to the statistical priori feasible region contraction strategy, the feasible regions finally obtained are $6 \leq x \leq 18$, $7 \leq y \leq 16$, $3 \leq z \leq 8$, in cm.

Assuming that the heart current source is sparse at each moment, the sparse decomposition method can be used to solve the inverse problem of linearization. The selected sparse current source adopts a set source intensity threshold, which is a dominant current source whose source intensity is not less than 90% of the reconstructed source intensity at that time and avoids the interference source caused by the noise of the measurement signal or the calculated residual. The electrical activity of the heart is studied by using the current source with great local intensity.

4.2 QRS-Band Current Source Reconstruction Results

The results of the single-time source estimation can be used to analyze the source activity at a certain point in time, but the workload is analyzed one by one, and it is not easy to find the rule. The result is combined with time to obtain a sequence of source estimates that show the evolution of the estimated source activity over a specified time period, which is shown in Fig. 2. Source activity area distribution map is shown in Fig. 3.

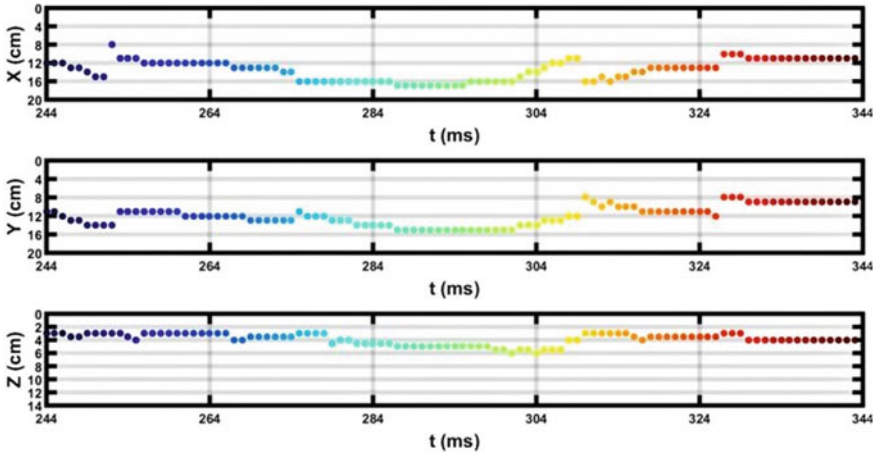
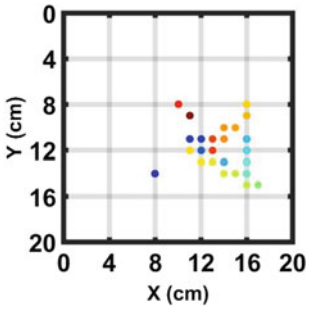


Fig. 2 Activity area time curve of QRS complex, where the horizontal axis represents time, and the vertical axis represents the components of the moving track in the *X*-, *Y*-, and *Z*-directions

Fig. 3 Source activity area distribution map



5 Conclusion

In this paper, a method for improving estimation of distributed sources is proposed for the reconstruction of cardiac current source and cardiac electrical activity. The simulation results show that the proposed method is stable for different parameters and different levels of Gaussian noise. The trajectory imaging of the reconstructed current source of MCG data is used to study the cardiac electrical activity conduction in the QRS segment. How to extract the difference features from the images that reflect the spatiotemporal distribution and intensity changes of the reconstructed current source and give the cardiac function diagnosis indicators is a key issue that needs to be studied in depth for better application in E-health smart service system.

Acknowledgements The authors would like to express their gratitude to Prof. Shiqin Jiang, Ph.D. Dafang Zhou, Ph.D. Chen Zhao for academic exchange and great support during this project. The authors are also grateful to Ph.D. Shulin Zhang, Prof. Xiaoming Xie, et al. from Shanghai Institute of Microsystem and Information Technology, for kindly providing the MCG data. This work is supported by the Natural Science Foundation of Shanghai (Grant No. 18ZR1427400) and the Shanghai Business School 'Phosphor' Science Foundation.

References

1. He, B.: Modeling and Imaging of Bioelectrical Activity: Principles and Applications, 2nd edn. Kluwer Academic, Plenum Press, Minnesota (2003)
2. Ghasemi-Roudsari, S., Al-Shimary, A., Varcoe, B., Byrom, R., Kearney, M.: A portable prototype magnetometer to differentiate ischemic and non-ischemic heart disease in patients with chest pain. PLoS ONE **13**(1), e0191241 (2018)
3. Sun, W., Kobayashi, K.: Estimation of magnetocardiography current sources using reconstructed magnetic field data. IEEE Trans. Magn. **53**(11), 1–4 (2017)
4. Hyde, D.E., Duffy, F.H., Warfield, S.K.: Voxel-based dipole orientation constraints for distributed current estimation. IEEE Trans. Biomed. Eng. **61**(7), 2028–2040 (2014)
5. Handiru, V.S., Vinod, A.P., Guan, C.: EEG source imaging of movement decoding: the state of the art and future directions. IEEE Sys. Man Cybern. Mag. **4**(2), 14–23 (2018)
6. Yi, H., Qu, X., Sun, Y., Peng, J., Hou, Y., He, X.: A permissible region extraction based on a knowledge priori for X-ray luminescence computed tomography. Multimedia Sys. **1**(6), 1–8 (2017)
7. Masood, M., Al-Naffouri, T.Y.: Sparse reconstruction using distribution agnostic bayesian matching pursuit. IEEE Trans. Signal Process. **61**(21), 5298–5309 (2013)

The Differential Properties of the Metric Projector Over a Class of Closed Convex Cones

Shiyun Wang and Jinyang Zhao

Abstract In this paper, we study the differential properties of the projector over the intersection of a closed half-space and a variable box: the directional derivative and B-subdifferential.

1 Introduction

Since the differential properties of metric projectors over convex closed cones are important for researching the augmented Lagrangian algorithm and sensitivity analysis, we will study the differential properties of metric projector over a class of closed convex cones. We carry out our work on the intersection of a closed half-space and a variable box:

$$K^n := \left\{(y, y_0)\, \epsilon R^n \times R : e^T y \leq k y_0, 0 \leq y \leq y_0 e\right\}, \tag{1}$$

where $k > 0$ satisfying $0 < k < n$. For national simplicity without confusion, we simplified K^n as K. In [1], we obtained the closed form of metric projector over K, denoted Π_K. In [2], we computed the geometric properties of K. Based on the closed form of the metric projector [1] and the critical cone of K [2], we unfold our work with the differential properties of metric projectors over K. Based on the closed form of the metric projector [1] and the critical cone of K [2], we unfold our work with the differential properties of metric projectors over K.

There are two main motivations for our works. Firstly, the closed convex cone K usually plays an important role in the optimization problems related to the epigraph of the vector/matrix k-norm functions. K also appears in finding the fastest mixing Markov chain on a graph [3, 4] and structured low rank matrix approximation [5], etc. These applications of K have been introduced in [1, 2].

S. Wang (✉) · J. Zhao
School of Science, Shenyang Aerospace University, Shenyang 110136, China
e-mail: wsy0902@163.com

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5_28

Secondly, the differential properties of metric operators, B-subdifferential and the Clarke's generalized Jacobian are connected with sensitivity analysis [6], semismooth/smoothing Newton method [7] and augmented Lagrangian method [8], etc. However, as far as we know, the differential properties of Π_K have not been found in the previous works. Encouraged by the wide applications of K, this paper aims at researching the differential properties of Π_K. We hope the results obtained in this paper can support further research into studying the optimization problems related to K.

We know that for given point (x, x_o), $\Pi_K(x, x_o)$ is the unique optimal solution of the following optimization problem (2)

$$\min\left\{\frac{1}{2}\left(\mathrm{y}-x^2+(y_0-x_0)\right)^2 : (y, y_0)\in K\right\}. \tag{2}$$

Let $\Pi'_K((x, x_o); (h, h_o))$ denote the directional derivative of $\Pi_K(x, x_o)$ along the given direction (h, h_o). Since K is a polyhedral convex set, the directional derivative $\Pi'_K((x, x_o); (h, h_o))$ can be characterized by (one can refer to [9, 10])

$$\prod\nolimits'_K((x, x_o); (h, h_o)) = \Pi_{\overline{K}}(h, h_o). \tag{3}$$

where $\overline{K}$ is the critical cone related to the problem (2).

2 Symbols and Preliminaries

For $x \in R^p$ and $Q\{1, 2, \ldots, p\}$, $|Q|$ denotes the cardinal number of Q and x_Q denotes the sub-vector of x by removing all the components of x not in Q. Let $\pi : [1 : p] \to [1 : p]$ be a one-to-one transformation such that $x_{\pi(i)} \geq x_{\pi(i+1)}$. For given point $k \in R$, we denote $k_- := \min\{k, 0\}$ and $k_+ := \max\{k, 0\}$.

In this section, we shall briefly introduce some relevant concepts. Let S be a given closed convex cone in the finite dimensional Euclidean space $R^n \times R$. The interior of S is denoted by $int(S)$, and the boundary of S is denoted by $bd(S)$. The dual S^* and polar S^o are defined, respectively, by

$$S^* = \left\{(x, x_0) \in R^n \times R : \langle (x, x_0), (y, y_0)\rangle \geq 0, \forall (y, y_0) \in S\right\}$$

and

$$S^o := \left\{(x, x_0) \in R^n \times R : \langle (x, x_0), (y, y_0)\rangle \geq 0, \forall (y, y_0) \in S\right\}.$$

Let Π_S denote the metric projector over S. For a given point $(x, x_0) \in R^n \times R$, if Π_S is F-differentiable at (x, x_0), we use $\prod'_s(x, x_0)$ to denote the F-derivative. If Π_S

is not F-differentiable at (s, s_0), the B-subdifferential is defined by

$$\partial_B \Pi_S(x, x_0) = \left\{ \lim_{k\to\infty} \prod\nolimits_S' \left(x^k, x_0^k\right) : \left(x^k, x_0^k\right) \to (x, x_0), \right.$$
$$\left. \Pi_S \text{ is F-differentiable at } \left(x^k, x_0^k\right) \right\}.$$

We first introduce the metric projector over K. By [1], for any point $(x, x_0) \in R^n \times R$, let $\sigma^* \geq 0$ be the parameter relative to (x, x_0). The projection Π_S at (x, x_0) is denoted by $(\bar{x}(\sigma^*), \bar{x}_0(\sigma^*))$. If there is no confusion, $(\bar{x}(\sigma^*), \bar{x}_0(\sigma^*))$ is simplified as $(\bar{x}, \bar{x}_0)$.

Let m be the cardinal number of the set $\{x_i : x_i > \sigma^*\}$ and λ be the smallest integer $i \in \{0, 1, \ldots, m-1\}$ such that $x_{i+1}^{\downarrow} - \sigma^* < \frac{x_0 + \sum_{k=1}^{i} x_k^{\downarrow} + (K-i)\sigma^*}{1+i} \leq x_i^{\downarrow} - \sigma^*$. If this inequality is not satisfied, let $\lambda = m$. Denote $\theta(x, x_0, \sigma^*) = \frac{x_0 + \sum_{k=1}^{i} x_k^{\downarrow} + (K-i)\sigma^*}{1+\lambda}$.

Then, the metric projector Π_K at (x, x_0) can be computed as $(\bar{x}, \bar{x}_0)$:

$$\bar{x}_0 = \max\left\{0, \theta\left(x, x_0, \sigma^*\right)\right\} \tag{4}$$

and for $i = 1, \ldots, n$,

$$\bar{x}_i = \begin{cases} 0, & x_i \leq \sigma^* \\ x_i - \sigma^*, & \sigma^* < x_i < \bar{x}_0 + \sigma^* \\ \bar{x}_0, & x_i \geq \bar{x}_0 + \sigma^* \end{cases} \tag{5}$$

For the sake of simplicity of description, we denote

$$\alpha := \left\{i \in [1:n] : x_i \leq \sigma^*\right\}, \gamma := \left\{i \in [1:n] : x_i \geq \bar{x}_0 + \sigma^*\right\}, \beta := [1:n](\alpha \cup \beta) \tag{6}$$

And we denote

$$\alpha := \left\{i \in \alpha : x_i = \sigma^*\right\}, \alpha_{\neq} := \alpha \backslash \alpha_{=}, \gamma_{=} := \left\{i \in \gamma : x_i = \bar{x}_0 + \sigma^*\right\}, \gamma_{\neq} := \gamma \backslash \gamma_{=}. \tag{7}$$

In order to introduce the critical cone $\overline{K}$, let p, l be the given real positive numbers and I_1, I_2, I_3, I_4 be the disjoint subsets of $[1:n]$. we define the following sets:

$$K_1^{I_1,I_2,I_3,p,l} = \left\{(z, z_0) \in R^n \times R : z_{I_1} = 0, z_{I_2} - z_0 e = 0, 0 \leq z_{I_3} \leq z_0 e, e^T z_{I_3} \leq l z_0\right\},$$
$$K_{1=}^{I_1,I_2,I_3,p,l} = \left\{(z, z_0) \in R^n \times R : z_{I_1} = 0, z_{I_2} - z_0 e = 0, 0 \leq z_{I_3} \leq z_0 e, e^T z_{I_3} = l z_0\right\},$$
$$K_2^{I_1,I_2,I_3,I_4,p} = \left\{(z, z_0) \in R^n \times R : z_{I_1} = 0, z_{I_2} - z_0 e = 0, 0 \leq z_{I_3}, z_{I_4} = z_0 e\right\},$$
$$K_3^{I_1,I_2,I_3,I_4,p,l} = \left\{(z, z_0) \in R^n \times R : z_{I_1} = 0, z_{I_2} - z_0 e = 0, 0 \leq z_{I_3}, z_{I_4} = z_0 e, e^T z \leq l z_0\right\},$$
$$K_{3=}^{I_1,I_2,I_3,I_4,p,l} = \left\{(z, z_0) \in R^n \times R : z_{I_1} = 0, z_{I_2} - z_0 e = 0, 0 \leq z_{I_3}, z_{I_4} = z_0 e, e^T z = l z_0\right\}.$$

If there is no confusion, the three sets $K_1^{I_1,I_2,I_3,p,l}$, $K_2^{I_1,I_2,I_3,I_4,p}$, $K_3^{I_1,I_2,I_3,I_4,p,l}$ are simplified by K_1, K_2 and K_3, respectively.

Lemma 2.1 Wang [2] Let $(x, x_0) \in \mathrm{R}^n \times \mathrm{R}$ be a given point and $\Pi K(x, x_0) = (\bar{x}, \bar{x}_0)$. K denotes the largest number which is less than K. The critical cone $\overline{K} R^n \times R$ of K at (x, x_0) associated with Problem (2) can be described by the following rule:

(a) if $(x, x_0) \in K$, then we have: when $(x, x_0) = (0, 0)$, $\overline{K} = K$; when $(x, x_0) \in int(K)$, $\overline{K} = R^n \times R$;
when $(x, x_0) \in bd(K)\backslash\{(0, 0)\}$, $\overline{K} = \begin{cases} K_2^{\emptyset,\gamma,\alpha,\emptyset,n}, & e^T x \neq kt. \\ K_3^{\emptyset,\gamma,\alpha,\emptyset,n,k}, & e^T x = kt; \end{cases}$

(b) if $(x, x_0) \in int\left(K^0\right)$, then $\overline{K} = \{(0, 0)\}$;

(c) if $(x, x_0) \in bd\left(K^0\right)\backslash\{(0, 0)\}$, we define $= \left\{x \in R^n : 0 \leq x \leq e, e^T \leq k\right\}$, then $\max\left\{x^T z : z \in\right\} + x_0 = 0, t \leq 0, \beta = \emptyset, \sigma^* = \left(x_{K+1}^{\downarrow}\right)_+$ and

$$\overline{K} = \begin{cases} K_{1}^{\alpha_{\neq},\gamma_{\neq,\alpha_{=},n,K-\left|\alpha_{\neq}\right|}}, & x_{K+1}^{\downarrow} \leq 0 \\ K_{1=}^{\alpha_{\neq},\gamma_{\neq,\alpha_{=},n,K-\left|\alpha_{\neq}\right|}}, & x_{K+1}^{\downarrow} > 0 \end{cases} \tag{8}$$

(d) otherwise, $\overline{K}$ is can be computed by

$$\overline{K} = \begin{cases} K_{3=}^{\alpha_{\neq},\gamma_{\neq,\alpha_{=},\gamma\neq n,K}}, \sigma^* \neq 0, \text{i.e., } K\bar{x}_0 - e^T\bar{x} = 0. \\ K_{2}^{\alpha_{\neq},\gamma_{\neq,\alpha_{=},\gamma\neq n}}, \sigma^* = 0, K\bar{x}_0 - e^T\bar{x} > 0. \\ K_{3}^{\alpha_{\neq},\gamma_{\neq,\alpha_{=},\gamma\neq n,K}}, \sigma^* = 0, K\bar{x}_0 - e^T\bar{x} = 0; \end{cases} \tag{9}$$

3 Metric Projector on the Critical Cone

Proposition 3.1 For given $(z, z_0) \in R^n \times R$, denote $\Pi K_1(z, z_0) = (\bar{z}, \bar{z}_0)$ then $(\bar{z}, \bar{z}_0)$ is comped as $\bar{z}_{I_1} = 0, \bar{z}_{I_2} = \bar{z}_0 e$, $\left(\bar{z}_{I_3}, \bar{z}_0\right) = \Pi K^{|I_3|}\left(\bar{z}_{I_3}, \bar{z}_0\right)$ and $\bar{z}_i = z_i$ for $i \in [1 : p\backslash](I_1 \cup I_2 \cup I_3)$, where $K^{|I_3|}$ is defined in (1).

Proposition 3.2 Assume that $(z, z_0) \in R^p \times R$ is given. Let $z_{I_2}^{\downarrow}$ be the nondecreasing reshape of z_{I_2} and $\overline{k}$ be the smallest integer $k \in [0 : |I_2|]$ such that $z_{k+1}^{\downarrow} \leq \frac{\sum_{i=1}^{k} z_i^{\downarrow} + z_0 + \sum_{i\in I_4} z_i}{1+k+|I_4|} < z_k^{\downarrow}$.

Then, $\Pi_{K_2}(z, z_0) = (z, z_0)$ *can be computed as*

$$\bar{z}_0 \leq \frac{\sum_{i=1}^{k} z_i^{\downarrow} + z_0 + \sum_{i\in I_4} z_i}{1 + |I_4| + k} \tag{10}$$

And *for* $i \in [i : p]$,

$$\bar{z}_i := \begin{cases} \bar{z}_0, \ i \in I_2 \text{and } z_i > \bar{z}_0; \text{ or } i \in I_4 \\ 0, \ \ i \in I_1; \text{ or } i \in I_3 \text{ and } z_i < 0, \\ z_i, \ \text{otherwise} \end{cases} \tag{11}$$

Proof The projection $\Pi_{K_2}(z, z_0)$ is the unique solution of the problem:

$$\min\left\{\frac{1}{2}\left(y - z^2 + (y_0 - z_0)^2\right) : (y, y_0) \in K_2\right\}.$$

The Lagrangian function can be written as

$$L\left(y, y_0, u_{I_1}, \omega_{I_2}, \lambda_{I_3}, V_{I_4}\right) = \frac{1}{2}\left(y - z^2 + (y_0 - z_0)^2\right) - y_{I_1} - u_{I_1} \\ - y_0 e - y_{I_2}, \omega_{I_2} - y_{I_3}, \lambda_{I_3} - y_0 e - y_{I_4}, V_{I_4}$$

Suppose that $\bar{z}.$ and $\bar{z}_0$ are defined by (11) and (12), respectively. Let $\bar{u}_{I_1} = -z_{I_1}$, $\bar{\omega}_{I_2} = \max\{0, z_{I_2} - \bar{z}_0 e\}$, $\bar{\lambda}_{I_3} = \max\{0, -z_{I_3}\}$ and $\bar{V}_{I_4} = -\bar{z}_0 e + z_{I_4}$. Set $I_5 = [1 : p]\backslash(I_1 \cup I_2 \cup I_3 \cup I_4)$. We will show that $\left(\bar{z}, \bar{z}_0, \bar{u}_{I_1}, \bar{\omega}_{I_2}, \bar{\lambda}_{I_3}, \bar{V}_{I_4}\right)$ satisfies the following KKT conditions of problem:

$$0 = y_{I_1} - z_{I_1} - u_{I_1}; \tag{12.1}$$

$$0 = y_{I_2} - z_{I_2} - \omega_{I_2}; \tag{12.2}$$

$$0 = y_{I_3} - z_{I_3} - \lambda_{I_3}; \tag{12.3}$$

$$0 = y_{I_4} - z_{I_4} - V_{I_4}; \tag{12.4}$$

$$0 = y_{I_5} - z_{I_5}; \tag{12.5}$$

$$0 = y_{I_1}; \tag{12.6}$$

$$0 = y_{I_4} - y_0 e; \tag{12.7}$$

$$0 = y_0 - z_0 - e^T \omega_{I_2} - e^T V_{I_4}; \tag{12.8}$$

$$0 \leq \lambda_{I_3} y_{I_3} \geq 0; \tag{12.9}$$

$$0 \le \omega_{I_2}\left(y_0 e - y_{I_2}\right) \ge 0. \tag{12.10}$$

It is easy to see that $\left(\bar{z}, \bar{z}_0, \bar{u}_{I_1}, \bar{\omega}_{I_2}, \bar{\lambda}_{I_3}, \bar{V}_{I_4}\right)$ satisfies the conditions in (12.1)–(12.7) and (12.9)–(12.10). Now, we only need to verify that $\left(\bar{z}, \bar{z}_0, \bar{u}_{I_1}, \bar{\omega}_{I_2}, \bar{\lambda}_{I_3}, \bar{V}_{I_4}\right)$ satisfies (12.8). Let π^{-1} be the inverse of π, and we denote $I' = \{i \in I_2 : z_i > \bar{z}_0\} = \left\{\pi^{-1}(1), \ldots, \pi^{-1}(k)\right\}$. By the expression of z_0, we have

$$\begin{aligned} e^T \bar{\omega}_{I_2} + e^T \bar{V}_{I_4} &= \sum_{i \in I'} (z_i - \bar{z}_0) - |I_4|\bar{z}_0 + \sum_{i \in I_4} z_i = \sum_{i=1}^{\bar{k}} z_i^{\downarrow} - \bar{k}\bar{z}_0 - |I_4|\bar{z}_0 + \sum_{i \in I_4} z_i \\ &= \sum_{i=1}^{\bar{k}} z_i^{\downarrow} + \bar{z}_0 + \sum_{i \in I_4} z_i - \left(\bar{k} + |I_4| + 1\right)\bar{z}_0 \\ &= \sum_{i=1}^{\bar{k}} z_i^{\downarrow} + \bar{z}_0 + \sum_{i \in I_4} z_i - \left(\sum_{i=1}^{\bar{k}} z_i^{\downarrow} + z_0 + \sum_{i \in I_4} z_i\right) = \bar{z}_0 - z_0 \end{aligned}$$

Then, we complete the proof.

To compute $\prod_{K_3}(\cdot, \cdot)$, inspired by [1], we consider the problem

$$\min\left\{\frac{1}{2}\left(y - x^2 + (y_0 - x_0)^2\right) - \hat{\sigma}\left(l y_0 - e^T y\right) : (y, y_0) \in K_2\right\} \tag{13}$$

where $\hat{\sigma} \ge 0$ is a parameter. If $\hat{\sigma}$ satisfies $0 \le \hat{\sigma}\left(l\bar{z}_0(\hat{\sigma}) - e^T \bar{z}(\hat{\sigma})\right) \ge 0$, then the optimal solution of (12) is just the optimal solution of (13), and the value of $\hat{\sigma}$ can be determined by the same method in [1]. By using the results of Proposition 3.2, we can immediately obtain the following result.

Proposition 3.3 Assume that $(z, z_0) \in R^p \times R$. Let $z_{I_2}^{\downarrow}$ be the nondecreasing reshape of z_{I_2} and $\bar{k}$ be the smallest integer $k \in [0 : |I_2|]$ satisfying:

$$z_{k+1}^{\downarrow} - \hat{\sigma} \le \frac{\sum_{j=1}^{k}\left(z_j^{\downarrow} - \hat{\sigma}\right) + (z_0 + l\hat{\sigma}) + \sum_{i \in I_4}(z_i - \hat{\sigma})}{1 + k + |I_4|} < z_k^{\downarrow} - \hat{\sigma}. \tag{14}$$

Then, the projection at (z, z_0) is computed by $\prod_{K_3}(z, z_0) = \left(\bar{z}(\hat{\sigma}), \bar{z}_0(\hat{\sigma})\right)$. Where

$$\bar{z}_0(\hat{\sigma}) := \frac{\sum_{j=1}^{\bar{k}} z_j^{\downarrow} + z_0 + \sum_{i \in I_4} z_i + \left(l - |I_4| - \bar{k}\right)\hat{\sigma}}{1 + |I_4| + \bar{k}},$$

and for $i \in [1 : p]$,

$$\bar{z}_i(\hat{\sigma}) := \begin{cases} \bar{z}_0(\hat{\sigma}), & i \in I_2 \text{and } z_i > \bar{z}_0(\hat{\sigma}) + \hat{\sigma}; \text{ or } i \in I_4 \\ 0, & i \in I_1; \quad \text{or} \quad i \in I_3 \text{ and } z_i < \hat{\sigma}, \\ z_i - \hat{\sigma}, & \text{otherwise,} \end{cases}$$

4 Differential Properties

Noting that α, β and γ be the index sets relative to (x, x_0) defined by (7), let $P_{\alpha,\beta,\gamma}$ be the permutation metric such that $P_{\alpha,\beta,\gamma}(x, x_0) = (x_\alpha, x_\beta, x_\gamma, x_0)$.

Theorem 4.1 Assume that $(x, x_0) \in R^n \times R$ is given. Let $(\bar{h}, \bar{h}_0)$ be the directional derivative of $\Pi_K(\cdot, \cdot)$ at (x, x_0) along the direction $(h, h_0) \in R^n \times R$, i.e. $\prod'_K((x, x_0); (h, h_0)) = (\bar{h}, \bar{h}_0)$. Then

$$(\bar{h}_\alpha, \bar{h}_\beta, \bar{h}_\gamma, \bar{h}_0) = \prod\nolimits'_{\overline{K}}(h_\alpha, h_\beta, h_\gamma, h_0).$$

Theorem 4.2 The metric projector Π_K is differential at $(x, x_0) \in R^n \times R$ if and only if (x, x_0) satisfies one of the following conditions:

(i) $(x, x_0) \in \text{int(K)}$, and then $\prod'_{\text{K}}(x, x_0) = I_{n+1}$;
(ii) $(x, x_0) \in \text{int}(K^o)$, and then $\prod'_K(x, x_0) = 0_{(n+1)\times(n+1)}$;
(iii) $(x, x_0) \in R^n \times R(K \cup K^o)$, $\alpha_= = \gamma_= = \emptyset$ and $(\sigma^*)^2 + \left(K\bar{x}_0 - e^T\bar{x}\right)^2 \neq 0$,

then $\prod'_K(x, x_0) = P^T_{\alpha,\beta,\gamma} V P_{\alpha,\beta,\gamma}$, where $P_{\alpha,\beta,\gamma}$ is the permutation matrix defined by (17) and V is given by:

$$V = \begin{bmatrix} 0_{|\alpha|\times|\alpha|} & 0_{|\alpha|\times|\beta|} & 0_{|\alpha|\times(|\gamma|+1)} \\ 0_{|\beta|\times|\alpha|} & I_{|\beta|} - uE_{|\beta|\times|\beta|} & -\delta E_{|\beta|\times(|\gamma|+1)} \\ 0_{(|\gamma|+1)\times|\alpha|} & -\delta E_{(|\gamma|+1)\times|\beta|} & \frac{1+(k-|\gamma|)\delta}{1+|\gamma|} E_{(|\gamma|+1)\times(|\gamma|+1)} \end{bmatrix} \tag{15}$$

where

$$\delta = \begin{cases} 0, & \sigma^* = 0, \\ \frac{|\gamma|-K}{(|\gamma|+1)|\beta|+(K-|\gamma|)^2}, & \sigma^* > 0; \end{cases} u = \begin{cases} \frac{1}{|\beta|}, & |\gamma| = K, |\beta| > 0 \\ \frac{(1+|\gamma|)\delta}{|\gamma|-K}, & |\gamma| \neq K. \end{cases}$$

Proof If $(x, x_0) \in \text{int}(K)$, we know $\prod_K(x, x_0) = (x, x_0)$, and if $(x, x_0) \in \text{int}(K^0)$, we know $\prod_K(x, x_0) = (0, 0)$. Hence, (i) and (ii) hold.

For (iii), we have $\bar{x}_0 \neq 0$, and σ^* is given by (8). If $\alpha_= = \gamma_= = \emptyset$, when $\sigma^* \neq 0$ or $K\bar{x}_0 - e^T\bar{x} \neq 0$ at least one holds, we can see that $\overline{K}$ is a linear space. Moreover, $\prod'_K((x, x_0); (h, h_0))$ is linear relative to (h, h_0), which implies $\prod_K(x, x_0)$ is differentiable. The derivative $\prod'_K(x, x_0)$ can be directly obtained by (4)–(5).

Theorem 4.3 Assume that $(x, x_0) \in R^n \times R$ is given. Denote $\prod_K(x, x_0)$ is the projection over K and σ^* is the relative parameter. Suppose that $\Pi \mathrm{K}(\cdot, \cdot)$ is not differentiable at (x, x_0) and α, β, γ are the index sets defined as (6). Let $\left\{\alpha_=^1, a_=^2\right\}, \left\{\gamma_=^1, \gamma_=^2\right\}$ be the partitions of $\alpha_=$and $\gamma_=$, respectively, such that

$$\bar{\alpha} = \alpha_{\neq} \cup \alpha_=^1, \bar{\beta} = \alpha_=^2 \cup \beta \cup \gamma_=^2, \bar{\gamma} = \gamma_{\neq} \cup \gamma_=^1$$

Define

$$V = \begin{bmatrix} 0_{Q_1 \times Q_1} & 0_{Q_1 \times Q_2} & 0_{Q_1 \times (Q_3+1)} \\ 0_{Q_2 \times Q_1} & I_{Q_2} - \bar{u} E_{Q_2 \times Q_2} & -\bar{\delta} E_{Q_2 \times (Q_3+1)} \\ 0_{(Q_3+1) \times Q_1} & -\bar{\delta} E_{(Q_3+1) \times Q_2} & \frac{1+(k-Q_3)\bar{\delta}}{1+Q_3} E_{(Q_3+1) \times (Q_3+1)} \end{bmatrix} \tag{16}$$

where $Q_1 := |\bar{\alpha}|$, $Q_2 := \left|\bar{\beta}\right|$, $Q_3 := |\bar{\gamma}|$ and

$$\bar{\delta} := \begin{cases} 0 \text{ or } \frac{Q_3 - K}{Q_2(Q_3+1)+(K-Q_3)^2}, & \sigma^* = 0, \\ \frac{Q_3 - K}{Q_2(Q_3+1)+(K-Q_3)^2}, & \sigma^* > 0; \end{cases} \quad \bar{u} = \begin{cases} \frac{(1+Q_3)\bar{\delta}}{Q_3 - k}, & Q_3 \neq K, \\ \frac{1}{Q_2}, & Q_3 = K, \bar{\beta} \neq \emptyset. \end{cases}$$

Then, we have the B-subdifferential of $\prod K(\cdot, \cdot)$ at (x, x_0) can be expressed as follows:

(i) if $(x, x_0) \in bd(K)\{(0, 0)\}$, then $\partial_B \Pi K(x, x_0) = \left\{P_{\bar{\alpha},\bar{\beta},\bar{\gamma}}^T V P_{\bar{\alpha},\bar{\beta},\bar{\gamma}} : V = I_{n+1} \text{ or } (16)\right\}$;

(ii) if $(x, x_0) \in bd\left(K^0\right)\{(0, 0)\}$, then $\partial_B \Pi K(x, x_0) = \left\{P_{\bar{\alpha},\bar{\beta},\bar{\gamma}}^T V P_{\bar{\alpha},\bar{\beta},\bar{\gamma}} : V = 0_{(n+1)\times(n+1)} \text{ or } (16)\right\}$;

(iii) if $(x, x_0) = (0, 0)$, then $\partial_B \Pi K(x, x_0) = \left\{P_{\bar{\alpha},\bar{\beta},\bar{\gamma}}^T V P_{\bar{\alpha},\bar{\beta},\bar{\gamma}} : V = 0_{(n+1)\times(n+1)}, I_{n+1} \text{ or } (156)\right\}$;

(iv) otherwise, $\partial_B \Pi K(x, x_0) = \left\{P_{\bar{\alpha},\bar{\beta},\bar{\gamma}}^T V P_{\bar{\alpha},\bar{\beta},\bar{\gamma}} : V \text{ is given by } (16)\right\}$.

Proof Given $(x, x_0) \in R^n \times R$, we suppose that $\prod K(\cdot, \cdot)$ is not differentiable at (x, x_0). Let $\mathrm{W} \in \partial_B \Pi K(x, x_0)$. Then, there exists a sequence $\left(x^j, x_0^j\right)$ which is differentiable such that

$$\lim_{j\to}\left(x^j, x_0^j\right) = (x, x_0) \quad \text{and} \quad \mathrm{W} = \lim_{j\to} \prod\nolimits_K' \left(x^j, x_0^j\right).$$

For notational simplicity, we denote $W^j = \prod_K' \left(x^j, x_0^j\right)$, then $W^j = P_{\alpha,\beta,\gamma}^T V^j P_{\alpha,\beta,\gamma}$ where V^j is given by (15). Then, $W = \lim_{j\to} W^j$. Let $\alpha^j, \beta^j, \gamma^j$ stand for the index set relative to $\left(x^j, x_0^j\right)$ defined by (6). Since $\Pi K(\cdot, \cdot)$ is not differentiable at (x, x_0), with no confusion, we suppose that $\alpha_= \neq \emptyset$ and $\gamma_= \neq \emptyset$. Then, we

know that, for any partition $\left\{\alpha_{=}^{1}, \alpha_{=}^{2}\right\}$ and $\left\{\gamma_{=}^{1}, \gamma_{=}^{2}\right\}$ of $\alpha_{=}$ and $\gamma_{=}$, respectively, there must exist a sequence $\left\{\left(x^{j}, x_{0}^{j}\right)\right\}_{j=1}$ such that

$$\lim_{j\to} x_{\alpha^j} = x_{\bar{\alpha}}, \lim_{j\to} x_{\beta^j} = x_{\bar{\beta}}, \lim_{j\to} x_{\gamma^j} = x_{\bar{\gamma}}, \lim_{j\to} x_0^j = x_0.$$

Taking limits, we can complete the proof by Theorem 4.2.

5 Conclusions

This paper studies the differential properties of the projector over the intersection of a closed half-space and a variable box K: the directional derivative and B-subdifferential. K also appears in finding the fastest mixing Markov chain on a graph which has a wide application in channel capacity of wireless communications. In the future work, we will go on studying the applications of the projector's differential properties in wireless communications.

Acknowledgements This work is supported by the National Natural Science Foundation of China (Grant No. 11701390).

References

1. Liu, Y.J., Wang, S.Y., Sun, J.H.: Finding the projection onto the intersection of a closed half-space and a variable box. Oper. Res. Lett. **14**, 259–264 (2013)
2. Wang, S.Y.: The geometry properties of a class of nonsymmetric cone, Indian journal of pure and applied mathematics, Accepted
3. Boyd, S., Diaconis, P., Xiao, L.: Fasted mixing Markov chain on a graph. SIAM Riev. **46**, 667–689 (2004)
4. Boyd, S., Diaconis, P., Parrilo, P.A., Xiao, L.: Fasted mixing Markov chain on graphs with symmetries. SIAM J. Optim. **20**, 792–891 (2009)
5. Chu, M.T., Funderlic, R.E., Plemmons, R.J.: Structured low rank approximation. Linear Anal. Appl. **336**, 157–172 (2003)
6. Sun, D.F.: The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Math. Oper. Res. **31**, 761–776 (2006)
7. Chen, Y.D., Gao, Y., Liu, Y.J.: An inexact SQP newton method for convex SC1 minimization problems. J. Optim. Theor. Appl. **146**, 33–49 (2010)
8. Zhao, X.Y., Sun, D.F., Toh, K.C.: A Newton-CG augmented Lagrangian method for semidefinite programming. SIAM J. Optim. **20**, 1737–1765 (2010)
9. Haraux, A.: How to differentiate the projection on a convex set in Hilbert space. J. Math. Soc. Japan **29**, 615–631 (1977)
10. Pang, J.S.: Newton's method for B-differentiable equations. Math. Oper. Res. **5**, 149–160 (1990)

Author Index

R. Kountchev et al. (eds.), *Advances in Wireless Communications and Applications*,
Smart Innovation, Systems and Technologies 190,
https://doi.org/10.1007/978-981-15-5697-5