Advances in Intelligent Systems and Computing 536

Jeng-Shyang Pan
Jerry Chun-Wei Lin
Chia-Hung Wang
Xin Hua Jiang *Editors*

Genetic and Evolutionary Computing

Proceedings of the Tenth International Conference on Genetic and Evolutionary Computing, November 7–9, 2016 Fuzhou City, Fujian Province, China

Springer

Advances in Intelligent Systems and Computing

Volume 536

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/11156

Jeng-Shyang Pan · Jerry Chun-Wei Lin
Chia-Hung Wang · Xin Hua Jiang
Editors

Genetic and Evolutionary Computing

Proceedings of the Tenth International Conference on Genetic and Evolutionary Computing, November 7–9, 2016 Fuzhou City, Fujian Province, China

Springer

Editors
Jeng-Shyang Pan
College of Information Science and Engineering
Fujian University of Technology
Fuzhou, Fujian
China

Chia-Hung Wang
College of Information Science and Engineering
Fujian University of Technology
Fuzhou, Fujian
China

Jerry Chun-Wei Lin
School of Computer Science and Technology
Harbin Institute of Technology (Shenzhen)
Shenzhen, Guangdong
China

Xin Hua Jiang
College of Information Science and Engineering
Fujian University of Technology
Fuzhou, Fujian
China

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-48489-1 ISBN 978-3-319-48490-7 (eBook)
DOI 10.1007/978-3-319-48490-7

Library of Congress Control Number: 2016954623

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume comprises the proceedings of the Tenth International Conference on Genetic and Evolutionary Computing (ICGEC 2016), which is hosted by Fujian University of Technology and is held in Fuzhou City, China on 7–9, November, 2016. ICGEC 2016 is technically co-sponsored by Springer, University of Computer Studies, Yangon, University of Miyazaki in Japan, Kaohsiung University of Applied Science in Taiwan, VSB-Technical University of Ostrava, and Taiwan Association for Web Intelligence Consortium. It aims to bring together researchers, engineers, and policymakers to discuss the related techniques, to exchange research ideas, and to make friends.

Thirty seven excellent papers were accepted for the final proceedings. One plenary talk is kindly offered by Prof. Han-Chieh Chao (President of National Dong Hwa University, Taiwan).

We would like to thank the authors for their tremendous contributions. We would also express our sincere appreciation to the reviewers, Program Committee members and the Local Committee members for making this conference successful. Finally, we would like to express special thanks for the financial support from Immersion Co., Ltd, China in making ICGEC 2016 possible.

September 2016

Jeng-Shyang Pan
Jerry Chun-Wei Lin
Chia-Hung Wang
Xin Hua Jiang

Organizing Committee

Honorary Chair

XinHua Jiang	Fujian University of Technology, China

Advisory Committee Chairs

XiaoDong Wang	Fujian University of Technology, China
Jun Murai	Keio University, Japan
Ke-Shou Wu	Xiamen University of Technology, China
Ajith Abraham	Machine Intelligence Research Labs, USA

General Chairs

Jeng-Shyang Pan	Fujian University of Technology, China
Ngo Thanh Long	Le Quy Don University, Vietnam
Ponnuthurai Nagaratnam Suganthan	Nanyang Technological University, Singapore

Program Committee Chairs

Mie Mie Thet Thwin	University of Computer Studies, Yangon, Myanmar
Pyke Tin	University of Computer Studies, Myanmar
Chu-Sing Yang	National Cheng Kung University, Taiwan
Hui Sun	Nanchang Institute of Technology, China
Jerry Chun-Wei Lin	Harbin Institute of Technology (Shenzhen), China

Invited Session Chairs

Thi Thi Zin	University of Miyazaki, Japan
Chia-Hung Wang	Fujian University of Technology, China
Chia-Jung Lee	Fujian University of Technology, China

Local Organizing Chairs

Fu-Min Zou	Fujian University of Technology, China
Zhi-Ming Cai	Fujian University of Technology, China

Electronic Media Chair

Tien-Wen Sung	Fujian University of Technology, China

Publication Chairs

Pei-Wei Tsai	Fujian University of Technology, China
Saw Sanda Aye	University of Technology, China
Xiangwen Liao	Fuzhou University, China
Mao-Hsiung Hung	Fujian University of Technology, China

Finance Chairs

Hong Chen	Fujian University of Technology, China
Alina Wu	Fujian University of Technology, China

Program Committee Members

Amir H. Alavi	Michigan State University, USA
Yogesh Bhalerao	Maharashtra Institute of Technology, Pune, India
Chun-Hao Chen	Tamkang University, Taiwan
M.C. Deo	Indian Institute of Technology, Bombay, India
U.S. Dixit	Indian Institute of Technology Guwahati, India
Philippe Fournier-Viger	Harbin Institute of Technology (Shenzhen), China
Amir Gandomi	Michigan State University, USA
Liang Gao	Huazhong University of Science and Technology, China
Ankit Garg	Indian Institute of Technology Guwahati, India

Tai Kang	Nanyang Technological University, Singapore
Jean Hok Yin Lai	Hong Kong Baptist University, China
Jasmine Siu Lee Lam	Nanyang Technological University, Singapore
Kawuu W. Lin	National Kaohsiung University of Applied Sciences, Taiwan
Jose M. Luna	University of Cordoba, Spain
S.S. Mahapatra	National Institute of Technology, India
Tay Kai Meng	Universiti Malaysia Sarawak, Malaysia
Biranchi Panda	Instituto Superior Técnico, Portugal, Europe
Amiya Patanaik	DUKE-NUS Graduate Medical School, Singapore
Jitendra Prasad	Indian Institute of Technology, Ropar, India
Nagarajan Raghavan	Singapore University of Technology and Design (SUTD), Singapore
G.P. Rangaiah	National University of Singapore, Singapore
Friedrich Recknagel	University of Adelaide, Australia
Srinivasa Prakash Regalla	Birla Institute of Technology, Pilani, Hyderabad Campus, India
Monica Mahesh Savalani	Hong Kong Polytechnic University, Hong Kong
Pravin Singru	Birla Institute of Technology, Pilani, Goa Campus, India
Suwin Sleesongsom	Chiangrai University, Thailand
Ja-Hwung Su	Cheng-Shiu University, Taiwan
Derrick I-Hsien Ting	National University of Kaohsiung, Taiwan
Sebastian Ventur	University of Cordoba, Spain
Venkatesh Vijayaraghavan	Rolls-Royce NTU Corporate Lab, Nanyang Technological University, Singapore
Bay Vo	Ho Chi Minh City University of Technology, Vietnam
Miroslav Voznak	VSB-TUO: Technical University of Ostrava, Czech Republic
Jimmy Ming-Thai Wu	University of Nevada, Las Vegas, USA
Mu-En Wu	SooChow University, Taiwan
A Yildiz	Bursa Technical University, Turkey
Ivan Zelinka	VSB-TUO: Technical University of Ostrava, Czech Republic
Sanyou Zeng	China University of Geosciences, China
XiangXiang Zheng	Xiamen University, China
Hannah Wan Zhou	University of Macau, China

Contents

Optimization Models and Techniques with Engineering Applications

A Comprehensive Evaluation Model for Traffic Rule

Lin Xiao[1,2,3](✉), Minqian Tang[2], and Jeng-Shyang Pan[1,3]

[1] School of Information Science and Engineering, Fujian University of Technology, No. 3 Xueyuan Road, University Town, Fuzhou 350108, People's Republic of China
xiaolin201@qq.com

[2] School of Information Management, Jiangxi University of Finance and Economics, No. 665, West Yuping Road, Nachang 330013 People's Republic of China

[3] Key Laboratory of Big Data Mining and Applications of Fujian Province, Fuzhou 350108, Fujian, China

Abstract. This article puts forward a model aimed at evaluating the traffic rule. We build the evaluating models to measure the traffic influencing factors, which can be divided into two kinds, the traffic flow factor and the safety factor. Analyze these factors to judge the performance of the keep-right-except-to-pass rule in light and heavy traffic. Draw the curve about time and other factors, let time be the intermediate variable, by using the figure conversion method, we get the curve in order to analyze the changing situation of each factor in light and heavy traffic. Do the comprehensive analysis of the combination figure by putting all the three curves in one coordinate system. We further set the basic lines as standards to be compared with the observing values. The result shows keep-right-except-to-pass rule performs well in the normal traffic yet badly in the extremely light traffic and the extremely heavy traffic.

Keywords: Traffic rule · Space-occupation ratio · Real traffic capacity · Vehicle headway distance

1 Introduction

Transportation is of vital importance in both ancient and modern human civilization as one of the main ways of communication. With the development of technology, approximately, the prime transportation means changes from carriage to automobile leading to the thriving and prosperous growth period of economy. However, in order to satisfy the increasing demand of economy rise and of human population, more and more automobiles are produced, which results in the high probability of the happening

The work is partially supported by Fujian Province Education Planning projects (FJJKCG15-051) and Fujian University of Technology Foundation Project (CY-Z15092) and Jiangxi Province Graduate Student Innovation Foundation Project (YC2015-B052).

J. Pan et al. (eds.), *Genetic and Evolutionary Computing*, Advances in Intelligent Systems and Computing 536, DOI 10.1007/978-3-319-48490-7_1

of the traffic jam, especially in the big cities such as New York, Beijing, and Rio. People have developed many methods which can be divided into two major different emphases (the hardware method and the software method) to solve such disturbing problems. The hardware method concerns itself mostly with the construction and designs of the roads, that is, building more roads in the city and designing some new kinds of road such as flyovers and tunnels, while the other method with the regulations and rules which are made to guide the behaviors of the drivers, such as the keep-right-except-to-pass rule, which is referred in [1, 2].

In countries (except for Great Britain, Australia, and some former British colonies) where people are used to driving automobiles on the right side of the road should adopt this rule, which requires drivers to drive in the right-most lane unless they are passing another vehicle, in which case they move one lane to the left, pass, and return to their former travel lane. No one knows exactly why most people choose to drive on the right, maybe it could satisfy the right-handers which most people are, and as a result, could guarantee the safety. However, to make sure whether this traditional rule could optimize the traffic situation leading to the best traffic flow, we are supposed to evaluate it by mathematic modeling and computing.

2 Problem Statement and Model Construction

This article puts forward a model aimed at evaluating the traffic rule. We assume the certain cross section of the roadway with two lanes as the simulate traffic situation. Two edges of this cross section (entry and exit, the function of two edges can be exchanged) both allow people to drive in and out straightly.

2.1 Quantitative Estimation Method of the Traffic Situation: Space-Occupation Ratio Model

The city's traffic situation reflects the overall operation condition of the automobiles on the road; it can be roughly divided into two ranges, the light traffic and the heavy one, which is referred in [3]. The heavy traffic happens, according to the definition made by Chicago transport agency, as long as the roadway occupancy rate of one certain road is larger than 30 %. On one certain cross section of the road, the roadway occupancy rate can be classified as R (space-occupation ratio), which equals to S_A (the area of the lane occupied by the automobiles) divided by S_R (the area of the total roadway), referred in [4]. Under the assumption of the dual-lane roadway, two lanes have the equal width, so the simplification form of the space-occupation ratio equation is f (the length of the automobiles passing the cross section during the given observing period) divided by l_r (the total length of the roadway).

As for the parameter f, we cannot neglect the factor that there are many kinds of vehicles differs from size and some other factors. To solve this problem, we use the standardization method. In most countries the transport agency stipulates that one certain kind of automobile as the basic unit of the traffic situation. When counting the quantity of the passing vehicles, other kinds of vehicles should be adjusted by their

Table 1. Type of automobile and the conversion coefficients

Typical type of automobile	Conversion coefficients	Introduction
Basic vehicle	1.0	For passenger car, seat quantity ≤ 19 For freight car, weight ≤ 2[ton]
Middle-sized vehicle	1.5	For passenger car, seat quantity>19 For freight car, 2[ton] < weight < 7 [ton]
Heavy duty car	2.0	Freight car, 7[ton] ≤ weight < 14 [ton]
Trailer	3.0	Freight car, weight ≥ 14[ton]

conversion coefficients to become the quantity of the basic unit. By looking up in the references, we summarized the related information as the table below to show the classification of the vehicle types and the conversion coefficients (Table 1):

Let a, b, c, d respectively represent basic vehicle, middle-sized vehicle, heavy duty car and trailer in order. By using the standardization method which combined conversion coefficients with four different types of automobile, we can calculate the length of the automobiles passing the cross section. As a result, the former simplification form of the space-occupation ratio equation can be rewritten as the final form like:

$$R = \frac{f = \sum a + 1.5\sum b + 2\sum c + 3\sum d}{l_r} \tag{1}$$

2.2 Traffic Flow Evaluation Model: The Real Traffic Capacity Model

According to the related references, the traffic capacity refers to the maximize quantity of automobiles passing the assumed cross section of the roadway during the given observing period, in the case of the certain traffic situation. As a result, we can simplify the real traffic capacity equation turning out to be:

$$N = \frac{1000\overline{v}}{l}(\text{vehicles/h}) \tag{2}$$

where we use N as the symbol of the real traffic capacity, and $\overline{v}$ is the average speed in a certain time interval, l is the vehicle headway distance in a certain time interval.

We use four types of vehicles and the conversion coefficient to calculate the standard traffic flow. a_1, a_2, a_3, a_4 respectively represent the four types' conversion coefficient (basic vehicle = 1, middle-sized vehicle = 1.5, heavy duty car = 2, and trailer = 3) in order, and q_i, similarly, represents the quantities of the certain type of vehicle passing the cross section. As for the vehicle headway distance calculating equation, l_1 is the length of automobiles, and Δt refers to the total time of the given observing period. Under the roadway assumption, when computing l_1, to simplify, we further assume that all the vehicles are the type of basic vehicle and determine according to the relative references its value as 5 m. Then we can get the final form of the real traffic capacity calculating equation, that is:

$$N = \frac{1000\overline{v}}{l_1 + \frac{\overline{v}\Delta t}{n-1}} = \frac{1000(n-1)\overline{v}}{5(n-1) + 60\overline{v}} = \frac{1000(-1 + \sum a_i q_i)\overline{v}}{5(-1 + \sum a_i q_i) + 60\overline{v}} \tag{3}$$

2.3 Safety Evaluation Model (1): The Vehicle Headway Distance Model

Under the previous roadway assumption, the two-lane roadway with the straight driving orientation, there are two main factors which can possibly cause the traffic accident, the vehicle headway distance and the passing sight distance. In this part we first consider the vehicle headway distance model.

The vehicle headway time distance D_t is the difference between two adjacent vehicles' time of before and after passing the cross section of the assumed roadway. It is closely related to the traffic environment and vehicle performance, and also by the impact of traffic control. The vehicle headway distance D_h is the distance between the two near vehicles' head. Where T_f and T_l respectively represent the time used to pass the cross section of the front car and of the later car, and l_1 refers to the length of cars, as we have explained its value is 5 m. Thus the final equation form is:

$$D_h = \left(T_f - T_l\right) \times \overline{v} + 5 \tag{4}$$

2.4 Safety Evaluation Model (2): The Overtaking Sight Distance Model

We know the driving characteristics of dual-lane road are only one lane in the same direction, and a variety of different types of the vehicle running on the road in different speed. As a result, when overtaking, drivers often have to occupy the opposite lane. The whole overtaking process is divided into three stages, namely, the lane changing, passing, and lane returning, referred in [5]. We consider the simplest situation, namely, only one overtaking car and one overtaken car, just as Fig. 1 shows:

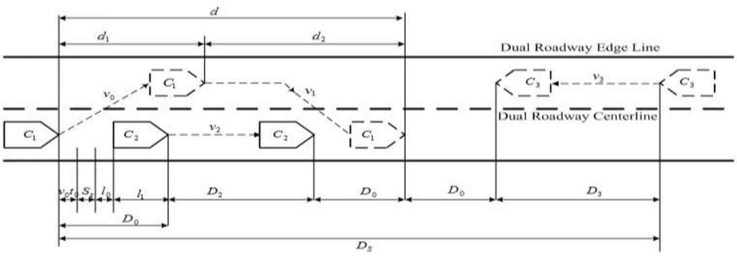

Fig. 1. Overtaking process in details

We call the overtaking sight distance D_S, that is:

$$D_S = 3D_0 + (v_2 + v_3)t = \frac{3v_1 - v_2 + 2v_3}{v_1 - v_2} D_0 + \frac{(v_1 - v_0)^2 (v_2 - v_3)}{2a_a (v_1 - v_2)} \tag{5}$$

This model shows that the overtaking sight distance is determined by the speed of the overtaking car when returning back to its original lane v_1, the speed of the overtaken car v_2, the speed of the coming car in opposite orientation v_3, the average accelerate speed a_a, and the distance between heads of the overtaking and overtaken car D_0. It also indicates that D_S is mostly determined by the value of $(v_1 - v_2)$. When speed of the overtaking and overtaken car is equal, the value of D_S is infinite. This result fits the practical situation, which the overtaking car is unable to overtake.

3 Numerical Computation and Curve Analysis

3.1 Explain the Source of Data

We find a related paper with the similar simulate traffic environment, referred in [6], so we just take the data from it as reference. In the paper the total observing time is 400 s, and the time interval is fixed as 2 s. We summarize the observing values in four large tables, with the all the needed parameters available, namely, the real traffic capacity N, the space-occupation ratio R, the vehicle headway distance D_h, and the average interval velocity V_i, just as Table 2 shows.

Table 2. Observing data of R, N, D_h and V_i

Parameter \ Time(s)	0	2	4	6	8	10	12	…	388	390	392	394	396	398	400
R(%)	0	7	14	14	7	7	7	…	0	0	0	0	0	0	0
N(vehicle/h)	0	14	29	29	14	14	14	…	0	0	0	0	0	0	0
D_h (m/vehicle)	70	70	35	35	70	70	70	…	70	70	70	70	70	70	70
V_i (km/h)	100	90	70	60	55	52	53	…	59	60	62	61	63	65	65

3.2 Figure Conversion

Based on the four separate data tables, we use matlab to find the result respectively in Figs. 2, 3, 4 and 5 which are shown below in order.

We can find that the common independent variable of the four curves is the time. As a result, we can use the method of figure conversion to combine two curves based on the time. By using matlab, $R - N$ curve, $R - V_i$ curve, $R - D_h$ curve are obtained base on the new coordinates, with the ability to analyze the relationship between the two parameters, in other word, from light to heavy traffic, the changes of the real traffic capacity, the average

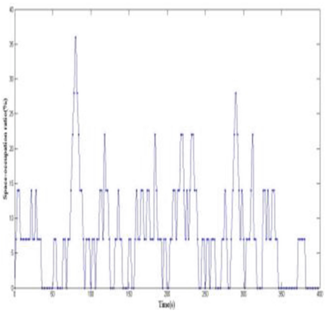

Fig. 2. Time-the space occupation ratio curve

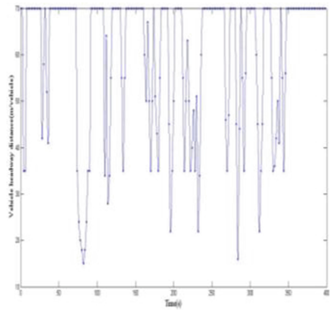

Fig. 3. Time-vehicle headway distance curve

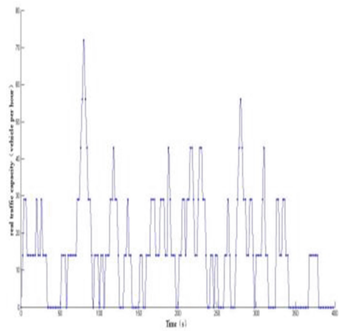

Fig. 4. Time-real traffic capacity curve

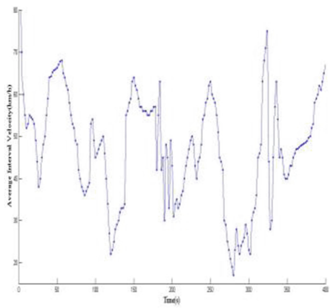

Fig. 5. Time-average interval velocity curve

Table 3. Selected coordinates of the four parameters based on the same time

R(%)	0	7	14	22	28	30	36	38	45	50	67	70	84	90	94	100
V_i(km/h)	100	90	57	37	30	25	23	20	17	14	11	9	7	5	3	2
N(vehicle/h)	0	14	29	43	56	88	72	70	64	60	49	45	32	25	10	0
D_h(m/vehicle)	70	58	42	37	31	24	22	21	14	13	10	9	5	4	3	2

interval and the vehicle headway distance. The required new-form coordinates' data is summarized in Table 3.

By using matlab, we put the $R-N$, $R-V_i$ and $R-D_h$ curve in one figure, in order to evaluate the performance of the keep-right-except-to-pass rule in light and heavy traffic. The combination figure is shown in Fig. 6.

3.3 Curve Analysis

First we analyze the $R-N$ curve by setting the basic line named the general capacity line. According to the related references, the general real traffic capacity is

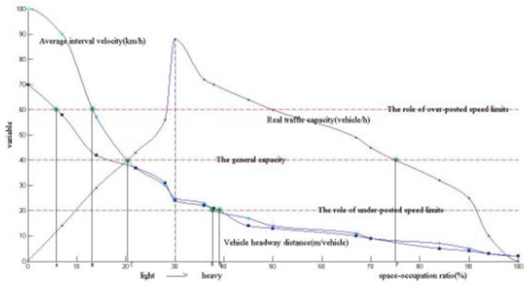

Fig. 6. Combination figure

40 vehicles/h. We can see in the figure that the basic line and the $R-N$ curve have two intersections, the left one and the right one. When $R \in [21\,\%, 76\,\%]$, the $R-N$ curve is above the basic line, which means that the real traffic capacity is better than the general condition. When $R \in [0, 21\,\%)$ and $R \in (76\,\%, 100\,\%]$, however, the $R-N$ curve is below the basic line, which indicates the bad traffic situation. To be more specific, the keep-right-except-to-pass rule performs well in the normal traffic condition and bad in the extremely light traffic and the extremely heavy traffic.

Then we judge the $R-V_i$ curve with the consideration of the role of under- or over-posted speed limits, which are determined as 20 km/h and 60 km/h. These two basic lines intersect the $R-V_i$ curve at two points. When $R \in [12\,\%, 38\,\%]$, the average interval velocity is in the proper speed range, which shows that the keep-right-except-to-pass rule fits well. However, when $R \in [0, 12\,\%)$ and $R \in (38\,\%, 100\,\%]$ the $R-V_i$ curve is out of the proper speed range. Consequently, the rule performs badly in the extremely light traffic and the extremely heavy traffic. In the former condition, the vehicles' speed is too fast to guarantee the safety, while in the latter condition, too slow to guarantee the traffic flow.

Finally we examine the safety by $R-D_h$ curve. From the figure we can find that a large part of the two curves (the $R-D_h$ curve and the $R-V_i$ curve) are almost the same, except for the part of $R \in [0, 23\,\%]$ which we regard as calculation error. So, it's appropriate to assume that the values of two parameters are similar. It's easy to understand that if the speed is big, then the vehicle headway distance is big in consideration of safety. By using the same comparing way as what we use in $R-V_i$, we set two basic line (the upper-limit and the lower-limit), their position are approximately the same as the speed limit line, and the similar conclusion will be shown.

4 Conclusions

To conclude, with the comprehensive analysis of the combination of safety (by D_h), the role of under- or over-posted speed limits and the traffic flow (by N), the keep-right-except-to-pass rule performs well in the normal traffic yet badly in the extremely light traffic and the extremely heavy traffic. So there do exist the optimization opportunity of the rule, we need to discuss further how to design a new rule in order to perform better than the keep-right-except-to-pass rule.

References

1. Khoury, J.E., Hobeika, A.G.: Integrated stochastic approach for risk and service estimation: passing sight distance application. J. Transp. Eng. **138**, 571–579 (2012)
2. Bai, W., Cun-Jun, L.I.: Overtaking model based on different limiting speed. J. Transp. Syst. Eng. Inf. Technol. **13**, 56–63 (2013)
3. Wang, R.Q., Zhou, Y.J., Xiao, C.E.: Calculation method of overtaking sight distance for dual-lane highway. J. Traffic Transp. Eng. **11**, 67–68 (2011)
4. Yue, Y., Luo, S., Luo, T.: Micro-simulation model of two-lane freeway vehicles for obtaining traffic flow characteristics including safety condition. J. Mod. Transp. **11**, 1–9 (2016)
5. Li, S., Zhao, X.J., Gao, Y., et al.: The determination of vehicular trajectories based on video image processing technology. In: 10th International Conference on Electric Technology and Civil Engineering, pp. 3086–3090. IEEE Computer Society, New York (2012)
6. Yao, R.H., Jing, J.: Chain reaction models for vehicle queuing on single-lane road section. Jilin Daxue Xuebao **42**, 892–898 (2012)
7. Bo, Y., Shenghua, X.: The evolutionary game simulation analysis of the multi-agent modeling's virtual enterprise knowledge transfer. Intell. Mag. **29**(5), 20–25 (2010)
8. Jing, Z., Li, X.: Research on topology control system for mesh networks. J. Chin. Comput. Syst. **34**(1), 140–144 (2013)

Composite Probe and Signal Recovery of Compressed Sensing Microarray

Zhenhua Gan[1,2,4], Baoping Xiong[1,3], Fumin Zou[1,4(✉)], Yueming Gao[3], and Min Du[2,3]

[1] College of Information Science and Engineering, Fujian University of Technology, No. 3 Xueyuan Road, University Town, Minhou, Fuzhou, Fujian, China
{ganzh, fmzou}@fjut.edu.cn, xiongbp@qq.com
[2] College of Electrical Engineering and Automation, Fuzhou University, No. 2 Xueyuan Road, University Town, Minhou, Fuzhou, Fujian, China
dm_dj90@163.com
[3] College of Physics and Information Engineering, Fuzhou University, No. 2 Xueyuan Road, University Town, Minhou, Fuzhou, Fujian, China
fzugym@yahoo.com.cn
[4] Key Lab of Automotive Electronics and Electric Drive Technology of Fujian Province, No. 3 Xueyuan Road, University Town, Minhou, Fuzhou, Fujian, China

Abstract. Due to the large number of uncertain factors in hybridization, image capture and processing of the microarray, multiple probes were generally arranged to improve the reliability of the measurement. However, the small area limited the number of probes that were allowed to be added on, so a composite probe would be the better choice. A composite probe contained the linear combination of a variety of gene fragments. It was used so that the microarray could easily realize the repeated gene fragments within a limited region. The number of composite probes would rapidly dwindle when it compared to a traditional microarray. At the same time, since the sparse characteristics of biological gene mutation, the compressed sensing idea is adopted to recovery the gene variation in the composite probes. The 96 fragments can be used with the 48 × 96 sparse random matrix to construct the 48 composite probes when the sparsest level K is no more than 12. Simulation results show that compressed sensing can accurately recover the gene mutation by using the Orthogonal Matching Pursuit (OMP) algorithm.

Keywords: Compressed sensing · Microarray · Composite probe · Sparse random matrix · OMP

1 Introductions

Microarray is a newly technology for high-throughput and quantitative detection in the biology science area. The abilities of microarray to express of thousands of genes simultaneously in a single detection have allowed the application in wide variety of fields, such as molecular biology, genetics, agriculture, disease diagnosis, medical treatment,

J. Pan et al. (eds.), *Genetic and Evolutionary Computing*, Advances in Intelligent Systems and Computing 536, DOI 10.1007/978-3-319-48490-7_2

food safety supervision, and judicial identification [1]. In a traditional microarray, each of the probe represents a complementary gene segment to be used to detect the corresponding gene information [2].

For the measurement noises, multiple probes were usually arranged to improve the reliability of the determination. The same probe was an effective way to avoid the information losses due to the interference of noises, but the repeated arrangement of probes resulted in an increase in the number of probes on the microarray. Thereupon, the weak fluorescence and the small size of probe were producing adverse effects while the density was increased, which also had caused serious irreparable damage for the ability to obtain reliable expression of the probes.

A more efficient method for solving the above problems was to use the composite probes. In this way each composite probe located in a spot was designed to detect the expression of multiple gene fragments simultaneously. The microarray scanner read the intensity of linear combination information from the composite probe, and the message of each gene probe would be obtained via the appropriate recovery algorithm [3].

Traditional cDNA gene sequencing probes produced a large number of mostly useless information, due to the fact that differences in the sequence between the reference sample and test sample were sparse. Because of the sparse characteristics of biological gene mutation, the compressed sensing idea was adopted to recover the gene variation in the composite probes. The compressed sensing theory had provided a strong support for the accurate recovery of the sparse signals, and it had been widely used in biological sensing, radar detection, data compression, image processing, and pattern recognition [4]. The compressed composite probes were constructed based on the compressed sensing ideas. The difference gene sequencing signals could be recovered by observing a small amount of the composite probes [5, 6].

The application of the composite probes on microarray was confirmed by [3]. And a composite probe method for constructing the compressed sensing microarray was proposed in [6]. A sparse low density parity check code (LDPC) as the measurement matrix to construct a compressed sensing microarray, and the recovery algorithm for the gene difference information were also proposed in [6]. For more information, the sparse random matrix in the recovery algorithm had the advantage of being a simple structure, low computational complexity, and easy to update and store in [7].

2 Design of Composite Probe for Compressed Sensing Microarray

2.1 Compressed Sensing

Compressed sensing is a sampling and reconstruction theory for sparse singles. Signal or the signal after a special transformation, with sparse or compressible characteristics, is the premise of compressed sensing [8, 9]. Considering a discrete digital signal $x \in R^{N}$ that has $K << N$ non-zero elements, the signal x is K sparse and N-K elements in the signal x will be 0 or close to 0. Since the signal x is generally not directly measured, we could design an $M \times N$ measurement matrix A to observe M linear combinations of the x, where $K << M << N$.

$$y_{M\times 1} = A_{M\times N}x_{N\times 1} \qquad K << M << N \tag{1}$$

Although the Eq. (1) is a underdetermined system, we also could reconstruct the signal x for K sparse by solving the constrained l_0 minimization,

$$\hat{x} = \arg\min \|x\|_0 \quad s.t. \quad y = Ax \tag{2}$$

where $\|x\|_0$ denotes the l_0-norm.

Unfortunately, solving the l_0 minimization is known as NP-hard. In order to solve this problem, it is usually converted into minimizing the l_1 with the optimization constraints. As long as the measure matrix A satisfies the restricted isometric property (RIP), the Eq. (1) agree with the following constraints,

$$\hat{x} = \arg\min \|x\|_1 \qquad s.t. \quad y = Ax \tag{3}$$

where $\|x\|_1$ denotes the l_1-norm [10, 11].

2.2 Composite Probe for Compressed Sensing Biological Microarray

The biological microarray uses the principle of molecular hybridization, which the gene to bind specific complementary sequences in the microarray probes. Since fluorescent labeling has been achieved already, we can get the fluorescent signal by light excitation. The information of the corresponding gene fragments from the resulting fluorescence signals can also be analyzed.

A typical cDNA microarray is fixed with a large number of probe spots located on the surface, but each probe consists of the single gene fragment, which can only detect specific complementary sequence segments. The detection principle of traditional cDNA is shown in Fig. 1.

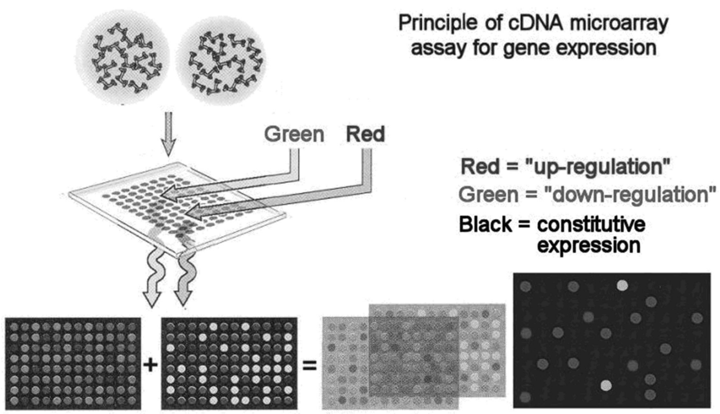

Fig. 1. The principle of traditional cDNA detection

The fluorescence intensity is at its most when the probes are matched normally on the microarray, and the intensity is at its weakest when the probes are mismatched. When the probes are not paired, there is little to no fluorescence intensity. The fluorescence intensity generated by match pairs is 5 times to 35 times more intense than that of a single or two bases mismatch in the probe's sequence. So the accurate determination of the fluorescent intensity is the basis of the specific detection of the biological sequence of microarray probe [12].

The composite probe fluorescence intensity is reflected the cumulative number of fluorescent molecules in various biological fragments fixed in the probe's spot. Literature [7] uses similar techniques as literature [3], which the design of the composite probe is realized by mixing the existing probe molecules according to the linear relationship of the measurement matrix A. This method can be used concurrently with the existing cDNA processing technology.

In particular, there are only a small fraction of the genes to be in a state of mutation. We are considering the difference that the gene expression of test sample is compared with the reference sample. And the difference of the signals which produced by two samples is nature sparse.

In order to construct a compressed sensing microarray with M composite probes, an M × N measurement matrix A with M < < N must be designed for N gene fragments. And we design the measurement matrix A with binary 0/1 elements only to simplify the construction difficulty of the compressed probes.

In two-color microarray of cDNA experiments, the reference sample is labeled by Cy3 while the test sample is labeled by Cy5 [13]. We are comparing two channel's sample by data vectors x_{cy3} and x_{cy5}, and interesting the difference expression of $x = x_{cy3} - x_{cy5}$.

Since there are differences in the small number of gene segments, the distribution of the x is sparse. The compressed sensing idea is relevant to the applications of DNA microarrays in the gene variation. Figure 2 illustrates the structure of the composite probe.

Each row of the matrix A represents a linear combination of the gene fragments. The m-th composite probe is determined by the positions of the gene fragment in the m-th row of matrix A. The combination structure of a composite probe is shown as the following,

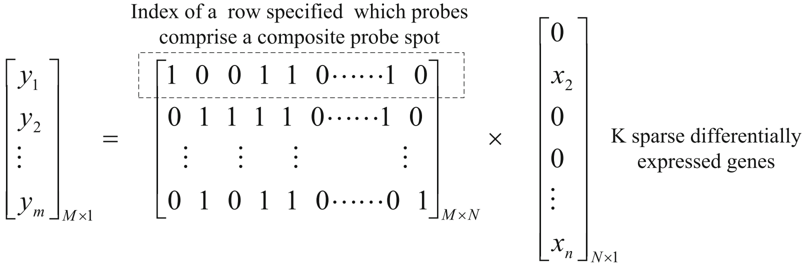

Fig. 2. Illustration of the compressed microarray

$$y_j = \sum_{i=1}^{N} a_{ji} x_i, \qquad j = 1, 2, \cdots M \tag{4}$$

where M < < N. Additionally, if the number of nonzero elements is different in each row, the actual mixed solution of probes should be diluted to the specified volume to ensure the consistency of the dilution.

3 Composite Probe Recovery Using Compressed Sensing

3.1 Sparse Random Measurement Matrices

Each column of the random sparse M × N matrix contains only uM non-zero elements with independent and identical distribution [14]. Literature [15, 16] also have pointed out that the recovery effect of sparse random measurement matrix is consistent with the gauss random measurement matrix. Moreover, the literature [14] have further proved that the sparse random matrix satisfies the RIP.

Due to the each row of the matrix represents a linear combination of a probe spot. We limit the elements of the random sparse matrix into binary 1/0 for the sake of constructing simplicity. The configuration process for sparse random matrix is as follows,

(1) Production M × N matrix of zeros;
(2) The position of each column elements is randomly selected according to the sparse coefficient u of the matrix, and these elements would be set to 1.

3.2 Recovery of Variation Gene from Composite Probe

In two-color microarry of cDNA, we are comparing two channel sample by x_{cy3} and x_{cy5}, and interesting in the difference expression of $x = x_{cy3} - x_{cy5}$. By sparse random matrix, the normalized observation value of the composite probe is defined as $y = y_{cy3} - y_{cy5}$.

If the compressed sensing recovery x is obtained directly by the combination method, which is a NP-hard as well known. Formula (3) is an l_1-norm optimization problem, compared to time-consuming convex optimization, the classical sparse approximation methods, such as the Orthogonal Matching Pursuit (OMP) algorithm, would be very suitable.

In the OMP algorithm, the residual vector r, which is the error of approximation vector y, is smaller and smaller after several iterations [17].

Let $x_k = \arg\min_x \|y - A_k x\|_2,\ r_k = y - A_k x,\ A_k = [A_{k-1}\ a_k]$ be a sub-matrix which selected in step k. Then the OMP algorithm process as follows [18, 19].

Input: compressed sampling matrix A, measured value y, the sparsity level K.

Output: reconstruction of the signal ^x, estimated support I.

Initialization: $x_0 = 0$, $r_0 = y$, $k = 0$, estimated support $I = \emptyset$.

(1) $k \leftarrow k + 1$;
(2) the index that is the best match with the residual vector r_{k-1}, and $\lambda k \leftarrow \text{argmax}_j\{| < r_{k-1}, a_j >|\}$;
(3) update the index $I_k = [I_{k-1}\ \lambda_k]$, and $A_k = [A_{k-1}\ a_k]$;
(4) reconstruction $\hat{}x \leftarrow [A_k]^{-1}\ y$;
(5) update the residual vector as $r_k \leftarrow y - A_k\ (\hat{}x)$;
(6) If $k \leq K$, then execute step (1), otherwise stop at $k > K$.

4 Simulation Results and Analysis

We have designed N = 96 cDNA microarray simulation probes with the idea of array-based comparative genomic hybridization (aCGH). The difference between the reference probes and the test probes, i.e., the sparsity level is K = 12. In the simulation experiments, the differences between the reference probes and the test probes have subjected to random distribution, and the locations of these different composite probes are also subjected to random distribution.

Figure 3a illustrates the reference probe x_{cy3}, and Fig. 3b demonstrates the probe x_{cy5}. Then, the differences between them, i.e., $x = x_{cy3} - x_{cy5}$ are shown in Fig. 3c.

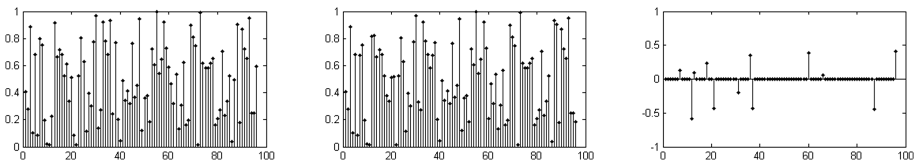

Fig. 3. **a**. The probe *xcy3,* **b**. The probe *xcy5,* **c**. $x = x_{cy3} - x_{cy5}$

We also have designed the sparse random matrix as compressed sensing measurement matrix A and let the elements sparsity coefficient $u = 0.25$. And M = 48 composite probes of compressed sensing microarray are constructed from N = 96 gene fragments by matrix A in the mixed method.

The observations of the composite probes are shown in Fig. 4a and Fig. 4b, while the differences between them, i.e., $y = y_{cy3} - y_{cy5}$ are shown in Fig. 4c.

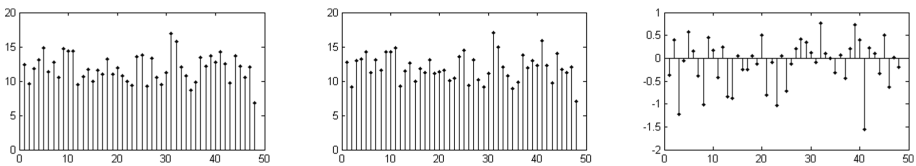

Fig. 4. **a**. The composite probes ycy3, **b**. The composite probes ycy5, **c**. $y = y_{cy3} - y_{cy5}$

We have used the OMP recovery algorithm to successfully reconstruct the gene different vector x, at N = 96, M = 48, K = 24, u = 0.25. As shown in Fig. 5, the recovery is so accurate that the relative error is $e = 4.4016 \times 10^{-15}$.

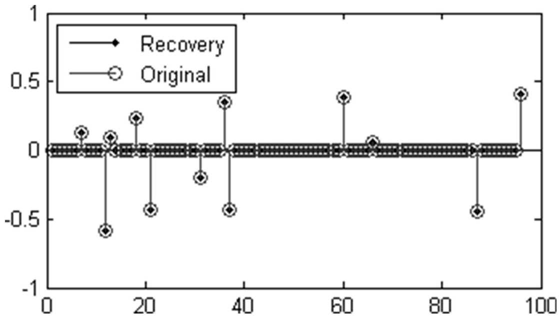

Fig. 5. The recovery of $\hat{x} = x_{cy3} - x_{cy5}$ with $e = 4.4016 \times 10^{-15}$

The structural parameters of cDNA simulation microarray have remained unchanged at N = 96, M = 0.5 N and the sparsity coefficient u = 0.25 for matrix A, and sparse K has been changed from zero to M. We still have used the OMP algorithm to recover the vector x. The accurate reconstruction ratios of the simulation signals are shown in Fig. 6.

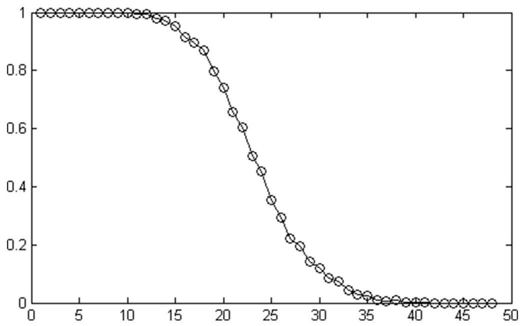

Fig. 6. The accurate reconstruction ratio of $\hat{x} = x_{cy3} - x_{cy5}$ for M = 48

As shown in Fig. 6, the compressed sensing algorithm recovers the probe's difference signals with high accuracy, at N = 96, M = 0.5 N, u = 0.25 and K ≤ 12.

Figure 7 demonstrates the accurate reconstruction ratio of simulation probes under the OMP recovery algorithm, when only the number of composite probe, i.e., M has been changed from zero to N.

It is shown in Fig. 7, compressed sensing algorithm achieves high accurate recovery for difference signals between the reference probes and the sample probes when the sparse random measurement matrix A is used at N = 96, K = 12, M ≥ 48, u = 0.25.

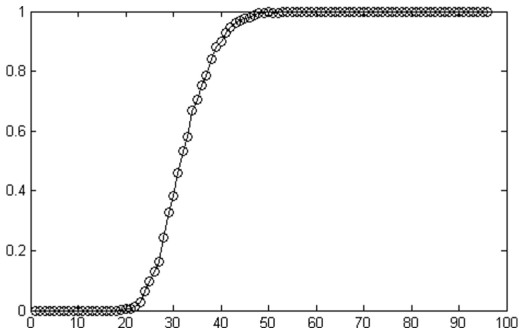

Fig. 7. The accurate reconstruction ratio of $\hat{x} = x_{cy3} - x_{cy5}$ for K = 12

5 Summary and Conclusions

There are a large number of uncertain factors in hybridization, image capture and processing of the microarray. In order to improve the reliability of the measurement, multiple probes are generally arranged to carry out repeated measurements. With a composite probe, a single spot of the compressed sensing microarray can easily and simultaneously measures many gene fragments, so that the repeated measurements of gene fragments can be realized with a limited number of spots. Considering the randomness and sparsity of genetic mutation, the total number of the composite probes installed in the compressed sensing microarray can be sharply reduced compared to that in the traditional microarray. Simulation experiment results show that, by using composite probes with gene fragment at N = 96, M = 0.5 N, and sparse random measurement matrix sparsity coefficient $u = 0.25$, when difference of cDNA probes K ≤ 12, based on OMP algorithm for compressed sensing, the high accuracy recovery of the difference signal of cDNA can be realized.

Acknowledgments. This work is partially supported by the National Natural Foundation Project (61304199), the Ministry of Science and Technology projects for TaiWan, HongKong and Maco (2012DFM30040), the Major projects in Fujian Province (2013HZ0002-1,2013YZ0002, 2014YZ0001), the Science and Technology project in Fujian Province Education Department (JB13140/GY-Z13088), and the Scientific Fund project in Fujian University of Technology (GY-Z13005,GY-Z13125).

References

1. Ping, L.Y.: Biological sensors and biological chips: the field of biological macromolecules. Chin. J. Lab. Diagn. **9**(4), 645–648 (2005)
2. Guolian, H., Chen, D., Shukuanl, X., et al.: Novel detection system of microbe chip and its application. Acta Optica Sinica **27**(3), 499–504 (2007)
3. Shmulevich, I., Astola, J., Cogdell, D., et al.: Data extraction from composite oligonucleotide microarrays. Nucleic Acids Res. **31**(7), 431–439 (2003)