Computer Science & Information Technology 91

David C. Wyld
Jan Zizka (Eds)

Computer Science & Information Technology

4th International Conference on Computer Science, Engineering and Information Technology (CSITY-2018) July 28-29, 2018, Sydney, Australia

AIRCC Publishing Corporation

Volume Editors

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Jan Zizka,
Mendel University in Brno, Czech Republic
E-mail: zizka.jan@gmail.com

ISSN: 2231 - 5403
ISBN: 978-1-921987-89-2
DOI : 10.5121/csit.2018.81101 - 10.5121/csit.2018.81107

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

Preface

The 4th International Conference on Computer Science, Engineering and Information Technology (CSITY-2018) was held in Sydney, Australia during July 28~29, 2018. The 4th International Conference on Data Mining (DTMN-2018), The 4th International Conference on Networks & Communications (NWCOM-2018) and The 4th International Conference on Signal and Image Processing (SIGPRO-2018) was collocated with The 4th International Conference on Computer Science, Engineering and Information Technology (CSITY-2018). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The CSITY-2018, DTMN-2018, NWCOM-2018, SIGPRO-2018 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in computer network and communications research.

In closing, CSITY-2018, DTMN-2018, NWCOM-2018, SIGPRO-2018 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the CSITY-2018, DTMN-2018, NWCOM-2018, SIGPRO-2018.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

David C. Wyld
Jan Zizka

Organization

General Chair

David C. Wyld	Southeastern Louisisna University, USA
Jan Zizka	Mendel University in Brno, Czech Republic

Program Committee Members

Abdelhalim BOUTARFA	University of Batna , Algeria
Abderrahmane Nitaj	University of Caen Normandie, France
Aberto Magrenan	International University of La Rioja (UNITE), Spain
Agoujil Said	University of Moulay Ismail Meknes, Morocco
Ahmad Qawasmeh	The Hashemite University, Jordan
Ahmad T. Al-Taani	Yarmouk University, Jordan
Ahmed Mohamed Khedr	Sharjah University, UAE
Ali Javadi	Iran University of Science and Technology, Iran
Ali Salem	University of Sfax,Tunisia
Ameera Saleh. Jaradat	Yarmouk University, Jordan
Amel B.H.Adamou-Mitiche	University of Djelfa, Algeria
Amizah Malip	University of Malaya, Malaysia
Annamalai	Prairie View A&M University, USA
Ashok Kumar T.A.	Garden City University, India
Basar Oztaysi	Istanbul Technical University, Turkey
Bing Zhou	Sam Houston State University, USA
Bouchra Marzak	Faculty of Sciences - Hassan II University, Morocco
Carmen Martinez	University of Jaen, Spain
Chang-Hyun	Korea Marine Equipment Research Institute, Korea
Chin-Chih Chang	Chung-Hua University, Taiwan
Christophe NICOLLE	University of Bourgogne Franche, France
Da Yan	The University of Alabama at Birmingham, USA
Dabin Ding	University of Central Missouri, USA
Debabrata Datta	St. Xavier's College, India
Farhad pourfarzi	Ardabil University of Medical Sciences, Iran
Figen Balo	Firat University, Turkey
Florence SEDES	Toulouse University, France
Guoqing Xiao	Hunan University, China
Hadi Amirpour	Universidade da Beira Interior, Portugal
Haibat Jadhav	Flora Institute of Technology, India
Haibo Yi	Shenzhen Polytechnic, China
Hamid Ali Abed AL-Asadi	Basra University, Iraq
Hamzeh Khalili	Universitat Politecnica de Catalunya (UPC), Spain
Hani Bani-Salameh	Hashemite University, Jordan
Hasnaoui Salem	University Tunis El-Manar, Tunisia
Houda KHROUF	Atos Innovation Lab, France
Hung Tran Cong	Posts and Telecoms Institute of Technology, Viet Nam

Hyunsung Kim	Kyungil University, Korea
Irena Patasiene	Kaunas University of Technology, Lithuania
Isa Maleki	Islamic Azad University, Iran
Islam Atef	Alexandria University, Egypt
Issa Atoum	The World Islamic Sciences and Islamic Studies, Jordan
Iyad alazzam	Yarmouk University, Jordan
Jasmine Seng K. P	Charles Sturt University, Australia
Jatindra Kumar Deka	Indian Institute of Technology Guwahati, India
Jia Zhu	South China Normal University, China
Jonice Oliveira	Universidade Federal do Rio de Janeiro (UFRJ), Brazil
Jun Liu	University of Michigan at Dearborn, USA
Junmei Zhong	Inspur, USA
Karina Gibert	Universitat politecnica de catalunya, Spain
Khaireddine Bacha	University of Tunisia, Tunisia
Khaled Almakadmeh	Hashemite University, Jordan
Kosai Raoof	Le Mans Universite, France
Manuel Angel Serrano Martin	Universidad de Castilla, Spain
Marius CIOCA	Lucian Blaga University of Sibiu, Romania
Maryam Habibi	Humboldt-Universitat zu Berlin, Germany
Mike Turi	California State University, Fullerton
Miroslaw Kwiatkowski	AGH University of Science and Technology, Poland
Mohammad Hamdan	Heriot Watt University, UAE
Mohammadreza Balouchestani	Indiana Purdue Fort Wayne University, USA
Mohammed AL Zamil	Yarmouk University, Jordan
Mohammed Al-Mai'itah	Al-Balqa applied university, Jordan
Mohammed Falah Mohammed	Universiti Malaysia Pahang
Mohammed J. Zaki	Rensselaer Polytechnic Institute, Troy
Morteza Alinia Ahandani	University of Tabriz, Tabriz, Iran
Nadhir Ben Halima	Taibah University, Saudi Arabia
Nahlah Shatnawi	Yarmouk University, Jordan
Nasser Thabet	Szabist University, UAE
Oded Maimon	Tel Aviv university, Israel
Omar Boussaid	University of Lyon, France
Ouafa Mah	Ouargla University, Algeria
Paulo Roberto Martins de Andrade	University de Regina, Canada
Quang Hung Do	University of Transport Technology, Vietnam
Rafael Stubs Parpinelli	State University of Santa Catarina, Brazil
Razieh malekhoseini	Islamic Azad University, Iran
Ruksar Fatima	KBN College of Engineering, India
Ryan Alturki	University of technology Sydney, Australia
Saban GLC	Necmettin Erbakan University, Turkey
Saltanat Meiramova	Seifullin Kazakh Agrotechnical University, Kazakhstan
Stefano Michieletto	University of Padova, Italy
Taeghyun Kang	University of Central Missouri, United States
Tanzila Saba	Prince Sultan University, Riyadh
Yashar Deldjoo	Universita degli Studi di Milano-Bicocca Milan, Italy
Zeyu Sun	Luoyang Institute of Science and Technology, China
Zhao Liang	University of Sao Paulo, Brazil

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Networks & Communications Community (NCC)

Soft Computing Community (SCC)

Organized By

Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

4th International Conference on Computer Science, Engineering and Information Technology (CSITY-2018)

4th International Conference on Data Mining (DTMN-2018)

4th International Conference on Networks & Communications (NWCOM-2018)

4th International Conference on Signal and Image Processing (SIGPRO-2018)

THE RELATIONSHIP BETWEEN NETWORK CAPABILITIES AND INNOVATION PERFORMANCE : EVIDENCE FROM CHINESE HIGH-TECH INDUSTRY

Gang Fang[1], Chen Chouyong[2] and Qing Zhou[3]

[1,2,3]Management School, Hangzhou Dianzi University, Hangzhou, China

ABSTRACT

Firms situated within innovation networks require specific abilities to acquire, from their network partners, the knowledge and the complementary assets that facilitate their innovation performance. Drawing on the resource-based view and social network theory, this study identifies two types of network capabilities: network structural capability and network relational capability. The purpose of this study is to deepen our understanding of the precise manner in which these network capabilities affect the networked firm's innovation performance. Based on the data obtained from Chinese high-tech firms, this study's findings suggest that network structural capability has a greater positive impact on innovation performance than network relational capability does within an exploration-orientated network. However, network relational capability is more positively associated with innovation performance within an exploitation-orientated network.

KEYWORDS

Innovation Network; Network Capabilities; Innovation; Resource-based View

1. INTRODUCTION

A firm situated within a network can acquire complimentary assets and resources from its network partners[1][2][3]. In particular, the knowledge sharing and learning routines that are fostered between a firm and its network partners can contribute to the former's ability to innovate[4][5]. Previous research in strategic management theory has introduced the concept of network resources[1][6][7], which can be described as the source of a firm's competitive advantage[8][9]. However, competitive advantages cannot be generated by resources alone; rather, they are contingent on the ways through which resources are effectively exploited and deployed, and these require specific capabilities[10][11]. Consequently, it is believed that firms situated within innovation networks require specific capabilities to better exploit network resources for enhancing and improving their innovation performance.

Previous research in social network theory has suggested that because of firms' asymmetric access to resources and their differing capacities of information gathering, inter-firm networks can significantly influence a firm's performance[12]. Similarly, Gulati (1998) argued that a firm's embeddedness within a network, which includes both structural embeddedness and relational embeddedness, can either facilitate or impede the benefits that it obtains from its partners[13]. Firms that are 'better connected' to their partners can obtain more benefits from innovation

David C. Wyld et al. (Eds) : CSITY, DTMN, NWCOM, SIGPRO - 2018
pp. 01–16, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81101

networks through extensive knowledge and information sharing with each other than those that are not, thereby improving their innovation success[14][15][16][17][18].

However, there has been a long-running debate within the network literature on the kind of network configuration that enhances a firm's performance, i.e. what is the 'better connection'? Weak ties or strong ties[19][20], and sparse structure or dense structure[21] [22]? As a way of promoting this debate further, several recent studies have proposed the use of a contingency approach. For instance, some studies have argued that weak or strong ties and sparse or dense structure can each be critical for a firm's innovation performance, depending on the particular context being studied and/or the firm's specific strategic purpose [23] [24]. Such studies have shed light on our understanding of the specific conditions under which strong/weak and sparse/dense networks are positively related to firm performance [25].

Although previous studies have highlighted the need for different levels of network density or tie strength in particular contexts, substantially less attention has been focused on the differential impacts of network density and tie strength on the innovation performance of a firm with a specific strategic purpose. Especially, exploration and exploitation may require inconsistent network configurations and firm capabilities. Some recent research has already discussed the impact of exploration and exploitation on value extraction from innovation network, however, our knowledge still remains undeveloped and, at least, unsystematic[24].

Drawing on the resource-based view and social network theory, this study aims to deepen our understanding of the precise manner in which network capability affects a firm's innovation performance. Following the contingency approach, it further attempts to identify the specific capability, whether network structural or network relational, that a firm would need most when shaping its innovation network to maximize value appropriation while keeping in line with the firm's strategic focus of exploration or exploration.

2. THEORY AND HYPOTHESES

2.1 Network capabilities

Innovation network is a system of autonomous and legally equal firms connected by selective, formal and persistent relations to transfer knowledge, or to innovate cooperatively. It provides an efficient mechanism for embedded firms to acquire new knowledge from partners[2], share risk or uncertainty with partners[26], and cope with systemic innovation[27]. One major research topic in innovation network area is based on social network theory. The majority of recent studies indicate that network configurations affect a firm's success at innovating. Among these configurations, three noted by previous researchers can be identified and integrated: network structures[28], network relationships[24][29], and network positions [4][30].

Another emerging research stream has attempted to delineate the source of value in inter-firm networks. Application of the resource-based view has been expanded to incorporate the inter-firm context by identifying valuable resources and capabilities that reside within networks[31][32]. Firms can create networks of external resources to complement their own resources, thereby facilitating their performance and, especially, the achievement of their organizational goals.The network resources perspective has advanced the theory of value creation within a network context. Dyer and Singh (1998) contended that relational rents can only be enjoyed by firms that combine, exchange, and co-develop idiosyncratic resources with their partners[1]. Networked firms do not merely respond passively to their existing network relationships[34]; rather, they proactively and deliberately manage and design their own ego networks. They do so either to pursue specific network structures (e.g., widely dispersed) or to become 'better connected' with their partners (e.g. stronger ties); they may also pursue both goals in accordance with their overall business strategies by utilizing specific network capabilities[35][36]. Introducing the concept of network capabilities, which represent a firm's ability to develop and manage networks, is thus vital for discussing the value creation and appropriation of network resources.

Prior research has identified several network capabilities or competencies of firms. These relate to their network management, including network competence [37][38][39], network management capability[40], strategic network capability[41], interaction capability, relational capability [42] [43], and alliance capability[44]. For example, Ritter (1999) suggested that a networked firm requires network competence to manage its network[37]. Ritter and Gemünden (2004) empirically determined that network competence has a positive effect on a networked firm's innovation success.[39] Hagedoorn et al. (2006) also argued that strategic network capability, i.e., the specific intelligence of firms regarding their network settings and their choice of particular partners, has a significant effect on the engagement of firms in future partnering activities[41].

These two streams of research, social network theory and strategic management, emphasized that the configurations of a network shape the performance of a networked firm. Recent results from social network theory suggest efforts that firms could make to improve benefits from the networks. Meanwhile, research of strategic management suggested that a networked firm could certainly benefit from its abilities to manage its ego network. However, to bridge the gap between these two streams of research, a new framework must be developed to explain the relationship between configuration shaping and network management. Therefore, the purpose of network capability introduced in this paper is to improve each aspect of network configuration to optimize interactions with partners and gain supernormal performance. Following Gulati (1998)[13], this study focuses on two types of network capabilities. Gulati's framework demonstrated that there are two main types of network embeddedness: structural embeddedness that focuses on the structure of the entire network and the position occupied by the firms within the network; and relational embeddedness that emphasizes the direct ties and close interactions among partners. The capabilities pertaining to the structural design of a network and the management of relationships within it are considered to have an important role in a firm's innovation performance.

2.2. Network structural capability and innovation

Following Gulati's (1998) framework[13], network structural capability refers to a focal firm's ability to improve a network's structural configuration. Structural elements may include the network's size[24], the different identities/diversity of membership within the network[45], the network's density[34], and the relative competitive position of the focal firm within the network [30]. Previous research has explored some of these network structural elements and the effects of these elements on innovation performance.

Through the identification, evaluation, and selection of potential and capable collaborators, the network structural capability may enable the focal firm to establish an innovation network that connects the partners who possess complementary knowledge and capabilities. Such a capability may also enable the focal firm to construct a high-density innovation network, which can improve information velocity, inculcate shared norms and behaviors, and increase the overall volume and speed of resource flows within the network[46]. According to Karamanos (2012)[20], a dense network structure has a positive effect on innovation performance. Ahuja (2000) also found that direct and dense connections within a network provide more resource-sharing and information-spillover benefits than indirect ones, as they result in more innovation opportunities[23].

Generally, capabilities do not automatically lead to performance improvements. However, network capabilities could optimize the network configurations, and which in turn, could impact the performance. This intermediate mechanism of network configuration is consistent with the extant literature[47]. Meanwhile, the results of a relevant case study of six Chinese high-tech firms (not presented here) suggested that there might be a positive impact of network capabilities on performance. This process and result is consistent with the suggestion by Ambrosini and Bowman (2009)[48], which contended that a fine-grained case study would help to explore the relationship between capabilities and performance. Therefore, the above arguments lead to the following:

Hypothesis 1: The higher the level of a firm's network structural capabilities, the greater the degree of innovation performance it will enjoy.

2.3. Network relational capability and innovation

Network relational capability, which is similar to the concepts of relationship-specific competence and network management capability[37][40], refers to a focal firm's ability to effectively manage relationships with its network partners. This entails fostering strong ties, engaging in frequent interaction with each partner, and maintaining long-term relationships[49] [50]. These activities enable a firm to effectively manage and mobilize resource exchange and to coordinate activities with network partners.

Network relational capability enables the focal firm to handle and exploit relationships with individual partners to maximize the benefits and complementary assets that it gains from these relationships. This contrasts with network structural capability in terms of the respective strategic foci of these two kinds of capabilities. In other words, network relational capability entails more emphasis on developing stronger ties and exploiting existing relationships, while network structural capability is more focused on the selection and exploration of new connections/members within a network. The benefits of exploiting relationships with existing partners are numerous. For instance, by effectively deploying its network relational capability, a focal firm may foster high levels of intimacy, trust, and compatibility with partners. And a trust-based and stable relationship can lead to a greater exchange of tacit knowledge[29], potentially generating higher innovation performance[51]. Thus, the second hypothesis is:

Hypothesis 2: The higher the level of a firm's network relational capabilities, the greater the degree of innovation performance it will enjoy.

2.4. Exploration-oriented and exploitation-oriented networks

March (1991) developed a framework that differentiates between explorative and exploitative modes of organizational learning[52]. Firms may alternate between explorative and exploitative learning modes, depending on their strategic purposes and environmental contexts. 'Exploration' refers to the pursuit of new knowledge or technology[53], and involves basic research, invention, the development of new capabilities, risk taking, and entry into new lines of businesses[54]. By contrast, 'exploitation' means the development and use of things that are already known, and includes improvement and refinement of existing capabilities and technologies, as well as systematic cost reduction. Extending March's (1991) framework to innovation networks leads us to the postulation that firms joining an innovation network may either be exploration oriented, with a focus on seeking new opportunities, or exploitation oriented, with a focus on exploiting existing resources and capabilities[52][55].

For this reason, firms attempting to implement radical innovations, focus on explorative learning, tend to establish or join exploration-oriented innovation networks to acquire new knowledge and ideas[56]. By contrast, firms attempting to implement incremental innovations, with a focus on exploitative learning, enter exploitation-oriented innovation networks to cooperate with partners and access complementary assets. Both types of innovation networks are beneficial for embedded firms, either because of changes in their fundamental architectures over long run or improvements within their basic structures and cost reductions in the short run.

Attempting to elucidate whether there are any advantages to be derived from network configurations for these two types of organizational learning and innovation purposes, for example, structural holes and dense connections[21][22], or weak ties[19] and strong ties[22][57], has long been at the center of a prevailing controversy in network literature. There was mixed evidence, and the findings were inconsistent originally in this research field. Some studies have shown that dense networks improve knowledge transfer[23], and thus innovation success[58], because dense ties tend to lead to the development of knowledge-sharing routines among partners[28]. However, other studies have argued that both strong and weak ties are positively associated with a firm's performance[59]. Meanwhile, Reagans and McEvily (2003) suggested that it is easier to transfer various sorts of knowledge when there is a strong tie, as opposed to a weak tie[29]. However, a weak tie is considered more efficient in transferring public or simple knowledge[60], because maintenance is less costly[50]. This debate has been resolved to some extent by certain studies' use of a contingency approach. Rowley et al. suggested that weak ties are beneficial for explorative purposes, while strong ties are positively related to the performances of firms engaged in exploitation. Likewise, Gilsing and Nooteboom (2005) argued that exploration requires higher network densities, since dense ties lead to some degree of redundancy in the types of knowledge sources, which is needed for ensuring the quality and reliability of information, and thus minimizing the uncertainty that is associated with exploration[24].

Although the contingency approach suggests that different types of networks are required for exploration and exploitation, previous studies have tended to focus on the differential benefits provided by weak or strong ties, or by sparse or dense connections. Consequently, there has been little or no attention paid to the different degrees of importance of tie strength and network density for the purposes of exploration and exploitation. This study attempts to shed some light on this issue by arguing that network structural capability and relational capability have interaction effects in relation to exploration/exploitation on a firm's innovation performance.

There are two arguments that support the significance of interaction effects. First, there are differences in the knowledge or information requirements of exploration and exploitation. As Gilsing and Nooteboom (2005) pointed out, exploration-based learning is an expansive process that involves broad searches for new knowledge[24], whereas exploitation-based learning is a deepening process that aims to refine and strengthen existing technology. Explorative learning thus focuses on redundant and diverse connections with partners. Denser networks provide more alternatives in terms of general knowledge, and improve the chances of developing all kinds of ties, including both strong and weak ties, that are effective for transferring either complex or simple knowledge[60]. The purpose of exploitative learning is to gain specific information, implying that interactions with certain technology providers become increasingly important. Strong ties enable partners to establish trust relationships and frequent interactions. These lead to enhanced mutual understanding and the development of common norms or routines. Establishing a common standard or work routine facilitates the transfer of specific knowledge[61]. Firms engaged in exploitation often focus their attention on a limited solution space[25], such as efficiency improvement or cost reduction. Stronger ties can serve better to solve these specific problems by providing tacit knowledge more efficiently [60].

Second, the different attributes of radical and incremental innovations contribute to the diversity of foci in exploration or exploitation. Firms that invest heavily in radical innovation face high environmental uncertainty, rapid changes in technology and ambiguity of direction. To receive redundant information, they require dense networks rather than repeated partnerships[62]. Diverse external collaborations can help them to obtain fresh ideas. In situations characterized by ambiguity in technological direction, dense networks enable firms to identify viable alternatives, discover the most likely future technological developments, and verify the accuracy of their knowledge. This would, in turn, increase firm's exploratory innovation performance [63]. Compared with radical innovation, incremental innovation pays closer attention to efficiency and short-term costs. Firms that are oriented toward incremental innovation typically focus on specific problems and invest in one direction. They prefer to solve specific problems jointly rather than gather general knowledge, implying that they have low tolerance for information noise. Strong ties promote the sharing of specific information and joint problem solving[64][65]. Consequently, firms within exploitation-oriented networks tend to depend more heavily on maintaining strong ties with specific information providers rather than on maintaining extensive relationships.

To recapitulate the above discussion, firms within explorative networks tend to be more heavily dependent on dense connections with diverse partners, compared with those within exploitative networks that typically prefer to maintain strong ties with specific information providers. Network structural capability improves a firm's ability to establish a dense network, while network relational capability enables a firm to create strong ties. Network structural/relational capability would, therefore, appear to yield positive interaction effects, which are associated with the type of innovation network, on a firm's innovation performance. Two further hypotheses are introduced as follows:

Hypothesis 3: The positive relationship between network structural capability and innovation performance is greater in exploration-oriented innovation networks than in exploitation-oriented innovation networks.

Hypothesis 4: The positive relationship between network relational capability and innovation performance is greater in exploitation-oriented innovation networks than in exploration-oriented innovation networks.

3. RESEARCH METHODOLOGY

3.1. DATA

The hypotheses were tested with the use of data from the survey that was administered to high-tech firms located in five provinces in eastern China. The Chinese high-tech industry was chosen for this study for two reasons. First, technological collaboration has been, and continues to be, a significant feature of this industry. Second, China's high-tech industry has developed rapidly since the 2000s, but its innovation level has remained relatively low than that found in other developed countries. The findings of this study may help practitioners and managers, especially in China, to improve their innovation activities through collaborations with network partners.

Potential participants were identified through an Internet search and interviews held with key informants. This study targeted top executives, as they were considered to be knowledgeable about their firms and inter-firm cooperation activities. A total of 1,285 questionnaires were distributed via email, or in paper format, and the final number of usable questionnaires was 211 (an effective response rate of 16.4%). Over 60 percent (66.8%) of the participating firms had less than 500 employees, and 58.3% of the firms were less than 10 years old.

3.2. MEASUREMENT SCALE

To ensure content validity of the measures used in this study, the measurement scale of the constructs was developed with the use of existing scales wherever possible, and a few items were slightly modified to fit the research setting. All items used the seven-point Likert scales.

The design of this scale followed the procedure introduced by Hinkin (1995)[66]. The format and items for each construct were initially developed based on a literature review and the combined inputs from relevant works. This effort was then complemented by field work undertaken within six Chinese high-tech firms to improve the selection of individual items. All items were then reviewed by a panel of experts within an inter-firm collaborative team composed of four professors and six managers from different firms. After conducting this review, some items that featured repeatedly, or were obscure, were eliminated or rephrased.

The resulting questionnaire was then pilot-tested. It was distributed to 325 individuals (approximately half were MBA students at a Chinese university; the remaining were employees of six Chinese high-tech firms). There were 113 responses in total, yielding a 34.8% response rate. Within this group, 84 were valid, resulting in a 25.8% effective response rate. After deleting two items with low loadings, an explorative factor analysis (EFA) was performed. This demonstrated that each variable had a loading greater than 0.5 with the expected factor. In addition, each Cronbach's α value exceeded 0.70, which indicated acceptable levels of internal consistency.

4. RESULTS

4.1. SCALE ASSESSMENT AND PRELIMINARY ANALYSES

Reliability and validity. To evaluate construct validity and internal consistency reliabilities[67], this study used principal component factor analysis. The results provided support for the validity of the constructs. In addition, this study included interviews with academic experts, and some of the measures were consistent with those used in previous research, thereby increasing the content validity of the constructs. Additionally, a confirmative factor analysis based on partial least squares (PLS) was conducted to examine discriminant validity. To obtain acceptable discriminant validity, the square root of the average variance extracted (AVE) of any variable in the model should be greater than the correlation coefficients between this value and any other variables[68][69][70]. As shown in Table 1, the results indicated good discriminant validity. Cronbach's α value for each construct was well above the cut-off value of 0.7 (Nunally,1978[71]), demonstrating adequate internal consistency of the constructs.

Table 1. Means, standard deviations, correlations and scale reliabilities

Variables	*Mean*	*SD*	*1*	*2*	*3*	*4*	*5*
1. Network Structural Capability	4.419	1.437	(.877)				
2. Network Relational Capability	4.558	1.209	.483**	(.850)			
3. Type of Innovation Network	3.995	1.809	-.152*	-.063	(.942)		
4. Innovation Performance	4.386	1.406	.754**	.604**	-.131	(.906)	
5. Age of Firm	15.84	29.509	.213**	.009	-.086	.028	n.a.
6. Size of Firm	2.422	.823	.329**	.302**	.042	.257**	.255**

Note: n = 211; Values in the diagonal cells are square roots of AVE. *p<.05. **p<.01.

Common method bias. Questionnaire, with random order of items, was separated into two parts and dispensed to different anonymous respondents, and data was collected through multiple sources. Factor analysis (Harmon's one-factor test) of all variables was conducted to check for common method variance. The results showed 4 factors with eigenvalues greater than 1.0 that accounted for 78% of the total variance, with the first factor accounting for only 29% of the total variance. These results implied that common method bias was not a significant problem in the survey responses. Additionally, as argued by Siemsen *et al.* (2010)[72], common method bias would not be a problem if the interaction hypotheses were found to be supported.

Multicollinearity. The variance inflation factor (VIF) was used to assess the degree of collinearity that existed within the regression models. All VIF values were found to be below 2.0, except for that of network constructing capability (VIF=2.096). These results indicated that substantial multicollinearity was not a serious issue in the study.

4.2. REGRESSION ANALYSIS

This study treated the size and age of firms as control variables and analysed the data with the use of hierarchical multiple regression. Table 2 presents the results of the regression analysis.

In Model I, firm size was positively related to innovation performance ($p< 0.001$). The effect of age was not significant. When the two network capabilities were included in Model II, the R^2 value increased significantly from 0.066 to 0.653. F-test revealed that adding the two network capabilities contributed significantly to the explanation of the dependent variable ($p< 0.001$). The results of Model II showed that the coefficients for each network capability were positive and significant ($p< 0.001$), indicating that either network structural capability or network relational capability contributed to innovation performance. Hypotheses 1 and 2 were supported.

Table 2. Results of regression analysis: moderating effects of the type of network

Variables	**Model I**	**Model II**	**Model III**	**Model IV**	**Model V**	**Model VI**
Constant	3.332^{***}	4.685^{***}	4.678^{***}	4.697^{***}	4.652^{***}	4.665^{***}
Age of Firm	-.010	$-.158^{*}$	$-.165^{*}$	$-.142^{*}$	$-.166^{*}$	-.129
Size of Firm	$.444^{***}$	.017	.027	.005	.035	.007
Network Structural Capability		$.617^{***}$	$.613^{***}$	$.601^{***}$	$.617^{***}$	$.601^{***}$
Network Relational Capability		$.364^{***}$	$.364^{***}$	$.369^{***}$	$.355^{***}$	$.357^{***}$
Type of Innovation Network			-.023	-.026	-.025	-.031
Network Structural Capability × Type of Innovation Network				$.048^{*}$		$.077^{***}$
Network Relational Capability × Type of Innovation Network					$-.054^{*}$	$-.090^{***}$

R^2	.066	.653	.654	.663	.662	.682
Adjusted R^2	.057	.647	.646	.654	.652	.671
⊿R^2	.066	.587	.001	.009	.008	.028
⊿F	7.331 ***	97.105 ***	.493	5.536*	4.753*	8.822** *

Note: n = 211; Dependent variable: Innovation Performance; * $p< 0.05$, ** $p< 0.01$, ***$p< 0.001$

Model VI, containing all of the variables, was considerably improved in comparison with Models III, IV, and V; the change in R^2 from Model III (0.654) to Model VI (0.682) was also significant (⊿R^2 = 0.028, ⊿F = 8.822, $p< 0.001$). This demonstrated the superior ability of Model VI to explain the moderating effect of the type of innovation network on the relationship between network capability and innovation performance. The regression coefficient of the interaction term, Network Structural Capability × Type of Innovation Network, was positive and significant (β = 0.077, $p< 0.001$). This implies that when an innovation network is oriented toward exploration, network structural capability will have a greater impact on innovation performance, thus supporting Hypothesis 3. The regression coefficient of the interaction term, Network Relational Capability × Type of Innovation Network, was negative and significant (β = -0.090, $p< 0.001$). This implies that when an innovation network is oriented toward exploitation (i.e., less explorative), network relational capability will have a greater impact on innovation performance, thus supporting Hypothesis 4.

To better understand the effects of the interactions discussed above, the interaction effects were plotted in graphs, as shown in Figure 1, with the use of one standard deviation above and below the mean to capture the high and low levels of the type of innovation network. These results provided further support of Hypotheses 3 and 4.

5. Discussion and Conclusion

In an innovation network, a firm’s network capabilities serve as enablers of value appropriation from a network. The empirical results of this study show that each type of network capability has a positive impact on a firm’s innovation performance. Previous studies, drawing from both social network theory and strategic management theory, have argued that interconnected firms are superior to independent firms. By integrating these two theoretical areas, and identifying the precise source of a networked firm’s competitive advantage, the concept of network resources corroborates this argument[1][6][7]. Moreover, the results of the current study further extend this insight by suggesting that network capability enables firms to generate rents that are latent within network resources. The finding of this study is consistent with Ritter (1999) [37]and Ritter *et al.* (2002)[73], which suggested that possessing network management capabilities improves a firm’s innovation performance.

More specifically, this study assesses the role of the type of innovation network as a critical mechanism underlying the innovation benefits derived from network capabilities. This study provides empirical support for these findings by focusing attention on the different types of innovation networks. First, the results suggest that a firm with higher levels of network structural and relational capabilities will evidence superior innovation performance, regardless of whether it is in an explorative or an exploitative network. This finding is at odds with the arguments of Granovetter (1973) [19]and Burt (1992)[14] on “weak ties” and “the structural hole,” respectively. It also contrasts with the argument made by Rowley *et al.* (2000)[25]. Based on their empirical study of American networked firms in the steel (exploitative) and semiconductor (explorative) industries, they contended that a combination of dense and strong ties provided few additional benefits, since creating and maintaining these ties incurred high costs. This study alternatively suggests that firms within the Chinese high-tech industry require high levels of both network structural capability and network relational capability to establish dense and strong ties with their partners. This, in turn, would improve their innovation performance.

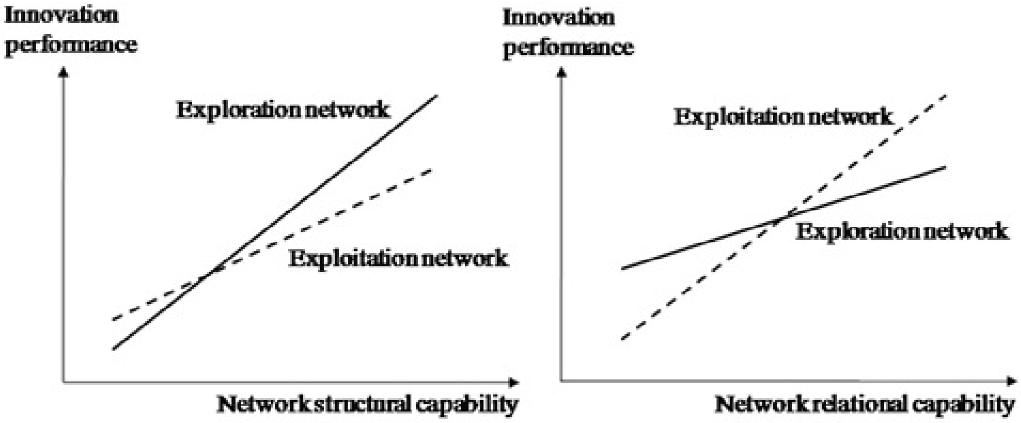


Figure 1. Interaction results[1]

This argument, however, is consistent with that of Coleman (1988)[22] concerning the benefits accrued from both dense and strong ties. Some recent research has also made similar suggestions. For example, Krackhardt (1992) contended that strong ties are more accessible and willing to be helpful[57], and so strong ties lead to greater knowledge exchange. Based on their empirical research, Reagans and McEvily (2003)[29] further suggested that the transfer of different types of knowledge through strong ties is relatively easier than the transfer through weak ties. This indicates that strong ties are more beneficial than weak ones with respect to a firm's innovation activities. This can be reasonably applied to Chinese high-tech firms, given that most of them are relatively young, and the level of interaction among firms is quite low. Many Chinese high-tech firms are now at a point where they are more interested in improving inter-firm cooperation and coordination than considering the cost of maintaining these ties.

This study suggested that the impact of network capabilities on performance via an intermediate effect of capabilities on network configurations. It conforms to the extant research[47]. Based on a literature review, Niesten and Jolink (2015) [47] unveiled a same explanatory mechanism for the impact of network management capabilities on performance. In addition, the results of a relevant study, which identified some antecedents of network capabilities[74], would also alleviate the possibility of presence of reverse causality.

Second, the empirical results from the Chinese high-tech industry further suggest that the positive effects of network structural capability (which leads to a dense network) are connected to a firm's particular purpose. When a focal firm faces an uncertain environment and focuses on explorative innovation, network structural capability is closely related to superior innovation performance. When the level of network structural capability increases, the performance of firms within an exploration-oriented network improves more rapidly than the performance of those located within an exploitation-oriented network (indicated by a slope of 0.740 vs. 0.468, see Figure 1(a)). The most plausible explanation for this is that it is indeed more important for exploration-oriented firms than for exploitation-oriented firms to obtain new knowledge and ideas and additional

[1] To illustrate the direction and magnitude of effects, the mean values of network relational capability in (a) and the mean value of network structural capability in (b) were used.

opportunities through the seeking of new partnerships/connections within the network. A dense innovation network that results from a firm's high level of structural capability is the best option for providing these inputs. This conclusion is consistent with that of Gilsing and Nooteboom (2005)[24], who argued that a higher network density and range would be more effective in improving performance for explorative learning than for exploitative learning.

Third, previous studies have suggested that weak ties promote the transfer of codified information or explicit knowledge, while strong ties are better suited for the transfer of non-codified information or tacit knowledge (Hansen,1999[60]; Uzzi and Lancaster,2003[50]). This study provides new insight into this issue. The empirical findings presented here suggest that the extent of the positive relationship between network relational capability and innovation performance depends on the focal firm's standpoint regarding innovation. When a focal firm focuses on exploitative innovation, this positive relationship becomes more significant. The line increases more sharply for exploitation, with a slope of 0.509, compared with that of 0.191 for exploration (see Figure 1(b)). This finding indicates that high-level relational capability is more important for exploitation-oriented firms than exploration-oriented firms. To engage in exploitative innovation, firms need strong and long-enduring ties for transferring existing knowledge and technologies. This is because exploitative learning focuses on the specific information being transferred through close and stable relationships.

In conclusion, this study offers a theoretical contribution to strategic management theory by highlighted the ways in which a networked firm creates and appropriates value from an innovation network according to its strategic purpose, and thus providing a more dynamic perspective for understanding performance differences across firms situated within the same network. The implications of this study - that a firm can enhance the value of its ego network by shaping and adjusting network configurations, rather than by passively reaping the benefits from existing relationships or ties with partners - may also contribute to social network theory. And this empirical study on innovation activities in the firms of China would be useful to contribute to, as Ambrosini and Bowman (2009)[48] suggested, a contingency approach to dynamic capabilities. Although high levels of both network structural capability and relational capability are beneficial, a full and meaningful understanding can only be attained if they are studied in conjunction with the type of innovation network under consideration. Within the existing literature, studies have found that different types of impact are produced by dense and sparse network structures (structural embeddedness) when firms are situated within explorative or exploitative networks (e.g., Rowley *et al.*2000[25]; Gilsing and Nooteboom 2005[24]). Researchers have also suggested that weak and strong ties (relational embeddedness) provide diverse benefits according to the changes of context. These arguments are challenging responses to those of 'the structural hole' (Burt,1992[21]) and 'the strength of weak ties' (Granovetter, 1973[19]). This study advances the contingent approach by comparing the different degrees of the importance of a dense structure (structural embeddedness) and strong ties (relational embeddedness). Consequently, it offers a new, general complementary perspective, as well as new evidence in support of the contingency-based argument within social network research.

REFERENCES

[1] Dyer, J. H., and H. Singh, (1998) "The relational view: cooperative strategy and sources of interorganizational competitive advantage", Academy of management review, Vol. 23, No. 4,pp660-679.

[2] Kale, P., H. Singh, and H. Perlmutter, (2000) "Learning and protection of proprietary assets in strategic alliances: Building relational capital", Strategic Management Journal, Vol. 21, pp217-237.

[3] Levin, D. Z., and R. Cross, (2004) “The strength of weak ties you can trust : The mediating role of trust in effective knowledge transfer”, Management Science, Vol. 50 , No. 11, pp1477-90.

[4] Tsai, W, (2001) “Knowledge transfer in intraorganizational networks: Effects of network positionand absorptive capacity on business unit innovation and performance”, Academy of Management Journal , Vol. 44, No. 5, pp996-1004.

[5] Cooke, P, (2006) “Global bioregions: knowledge domains, capabilities and innovation system networks”, Industry and Innovation, Vol. 13, No.4, pp437-458.

[6] Gulati, R, (1999) “Network location and learning:The influence of network resources and firm capabilities on alliance formation”, Strategic Management Journal , Vol. 20, No. 5, pp397-420.

[7] Gulati, R., N. Nohria, and A. Zaheer, (2000) “Strategic networks”, Strategic Management Journal ,Vol. 21, No. 3, pp203.

[8] Barney, J. B, (1992) “Integrating organizational behaviour and strategy formulation research:Aresource based analysis”, Advances in Strategic Management, Vol. 8, No. 1, PP39-61.

[9] Madhok, A., and S. B. Tallman, (1998) “Resources, transactions and rents: managing val-ue through interfirm collaborative relationships”, Organization Science,Vol. 9, No. 3, pp326-339.

[10] Grant, R. M, (1991) “The resource-based theory of competitive advantage:implications for strat-egy formulation”, Knowledge and Strategy, Ed. M. Zack, pp3-23.

[11] Amit, R., and P. J. Schoemaker, (1993) “Strategic assets and organizational rent”, Strate-gic Management Journal, Vol. 14, No. 1, pp33-46.

[12] Granovetter, M, (1983) “The strength of weak ties:A network theory revisited”, Sociologic-altheory, Vol. 1, No. 1, pp201-233.

[13] Gulati, R, (1998) “Alliances and networks”, Strategic Management Journal,Vol. 19, No.4, pp293-317.

[14] Burt, R. S, (2000) “The network structure of social capital”, Research in organizational behavior, Vol. 22, pp345-423.

[15] Tsai, W, (2002) “Social structure of “coopetition” within a multiunit organization: Coordination, competition, and interorganizational knowledge sharing”, Organization Science , Vol. 13, No. 2, pp179-190.

[16] Bellamy, M. A., S. Ghosh, and M. Hora, (2014) “The influence of supply network structure on firm innovation.”, Journal of Operations Management, Vol. 32, No. 6, pp357-73.

[17] Owen-Smith, J., and W. W. Powell, (2004) “Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community”, Organization Science, Vol. 15, No. 1, pp5-21.

[18] Salman, N., and A. L. Saives, (2005) “Indirect networks: an intangible resource for biotechnology innovation”, R&D management, Vol. 35, No. 2, pp203-215.

[19] Granovetter, M. S, (1973) “The strength of weak ties”, American journal of sociology, pp1360-1380.

[20] Karamanos, A. G, (2012) “Leveraging micro- and macro-structures of embeddedness in alliance networks for exploratory innovation in biotechnology”, R&D Management, Vol. 42, No. 1, pp71-89

[21] Burt, R. S, (1992) “Structural hole”, Harvard Business School Press, Cambridge, MA.

[22] Coleman, J. S, (1988) “Social capital in the creation of human capital”, American journal of sociology, ppS95-S120.

[23] Ahuja, G, (2000) “Collaboration networks, structural holes, and innovation: A longitudinal study”, Administrative Science Quarterly, Vol. 45, No. 3, pp425-55.

[24] Gilsing, V., and B. Nooteboom, (2005) “Density and strength of ties in innovation networks: an analysis of multimedia and biotechnology”, European Management Review, Vol. 2, No. 3, pp179-197.

[25] Rowley, T., D. Behrens, and D. Krackhardt, (2000) “Redundant governance structures: An analysis of structural and relational embeddedness in the steel and semiconductor industries”, Strategic Management Journal, Vol. 21, No. 3, pp369-386.

[26] Bleeke, J., and Ernst, D, (1991) “The way to win in cross-border alliances”, Harvard Business Review, Vol. 69, No. 6, pp127-135.

[27] Freeman, C, (1991) “Networks of innovators: a synthesis of research issues”, Research Policy, Vol. 20, pp499-514.

[28] Walker, G., B. Kogut, and W. Shan, (1997) “Social capital, structural holes and the formation of an industry network”, Organization Science, Vol. 8, No. 2, pp109-25.

[29] Reagans, R., and B. McEvily, (2003) “Network structure and knowledge transfer: The effects of cohesion and range”, Administrative Science Quarterly, Vol. 48, No. 2, pp240-267.

[30] Bell, G. G, (2005) “Clusters, networks, and firm innovativeness”, Strategic Management Journal, Vol. 26, No. 3, pp287-95.

[31] Wernerfelt, B, (1984) “A resource-based view of the firm”, Strategic Management Journal, Vol. 5 No. 2, pp171-180.

[32] Barney, J, (1991) “Firm resources and sustained competitive advantage”, Journal of Management , Vol. 17, No. 1, pp99-120.

[33] Cunningham, M. T, (1995) “Competitive strategies and organizational networks in new-technology markets”, Business marketing: an interaction and network perspective, pp336.

[34] Dhanaraj, C., and A. Parkhe, (2006) “Orchestrating innovation networks”, Academy of management review, Vol. 31, No. 3, pp659-669.

[35] Hambrick, D. C, (1984) “Taxonomic approaches to studying strategy: some conceptual and methodological issues”, Journal of Management, Vol. 10, No. 1, pp27-41.

[36] Miller, D, (1987) “The genesis of configuration”, Academy of management review, Vol. 12, No. 4, pp686-701.

[37] Ritter, T, (1999) “The networking company: antecedents for coping with relationships and networks effectively”, Industrial Marketing Management, Vol. 28, No. 5, pp467-479.

[38] Ritter, T., and H. G. Gemünden,(2003) “Network competence: its impact on innovation success and its antecedents”, Journal of Business Research, Vol. 56, No. 9, pp745-755.

[39] Ritter, T., and H. G. Gemünden, (2004) “The impact of a company's business strategy on its technological competence, network competence and innovation success”, Journal of Business Research, Vol. 57, No. 5, pp548-556.

[40] Möller, K. K., and A. Halinen, (1999) “Business Relationships and Networks:Managerial Challenge of Network Era”, Industrial Marketing Management , Vol. 28, No. 5, pp413-427.

[41] Hagedoorn, J., N. Roijakkers, and H. Kranenburg, (2006) “Inter□Firm R&D Networks: the Importance of Strategic Network Capabilities for High□Tech Partnership Formation1”, British Journal of Management , Vol. 17, No. 1, pp39-53.

[42] Lorenzoni, G., and A. Lipparini, (1999) “The leveraging of interfirm relationships as a distinctive organizational capability: a longitudinal study”, Strategic Management Journal, Vol. 20,No. 4, pp317-338.

[43] Collins, J. D., and M. A. Hitt, (2006) “Leveraging tacit knowledge in alliances: The importance of using relational capabilities to build and leverage relational capital”, Journal of Engineering and Technology Management, Vol. 23, No. 3, pp147-167.

[44] Kale, P., and H. Singh, (2007) “Building firm capabilities through learning: the role of the alliance learning process in alliance capability and firm□level alliance success”, Strategic Management Journal, Vol. 28, No. 10, pp981-1000.

[45] Cummings, J. N, (2004) “Work groups, structural diversity, and knowledge sharing in a global organization”, Management Science, Vol. 50, No. 3,pp352-364.

[46] Gnyawali, D. R., and R. Madhavan, (2001) “Cooperative networks and competitive dynamics: A structural embeddedness perspective”, Academy of management review, Vol. 26, No. 3. pp431-445.

[47] Niesten, E. and A. Jolink, (2015) “The impact of alliance management capabilities on alliance attributes and performance: a literature review”, International Journal of Management Reviews, Vol. 17, pp69-100.

[48] Ambrosini, V. and Bowman, C, (2009) “What are dynamic capabilities and are they a useful construct in strategic management?”, International Journal of Management Reviews, Vol. 11, No. 1, pp 29-49.

[49] Dyer, J. H., and K. Nobeoka, (2000) “Creating and managing a high□performance knowledge□sharing network: the Toyota case”, Strategic Management Journal, Vol. 21, No. 3, pp345-367.

[50] Uzzi, B., and R. Lancaster, (2003) “Relational embeddedness and learning: The case of bank loan managers and their clients”, Management Science, Vol. 49, No. 4, pp383-399.

[51] Rese, A., and Baier, D, (2011) “Success factors for innovation management in networks of small and medium enterprises”, R&D Management, Vol. 41, No. 2, pp138-155.

[52] March, J. G, (1991) “Exploration and exploitation in organizational learning”, Organization Science, Vol. 2, No. 1, pp71-87.

[53] Levinthal, D. A., and J. G. March, (1993) “The myopia of learning”, Strategic Management Journal, Vol. 14, No. S2, pp95-112.

[54] Koza, M. P., and A. Y. Lewin, (1998) “The co-evolution of strategic alliances”, Organization Science, Vol. 9, No. 3, pp255-264.

[55] Rothaermel, F. T, (2001) “Incumbent's advantage through exploiting complementary assets via interfirm cooperation”, Strategic Management Journal, Vol. 22, No. 6□7, pp687-699.

[56] Ettlie, J. E., W. P. Bridges, and R. D. O'keefe, (1984) “Organization strategy and structural differences for radical versus incremental innovation”, Management Science , Vol. 30 , No. 6, pp682-695.

[57] Krackhardt,D, (1992) “The strength of strong ties: the importance of philos in organizations”, N.Nohria, R. Eccles, eds. Network and Organizations: Structure, Form and Action. Harvard Business School Press, MA, pp216-239

[58] Obstfeld, D, (2002), “Knowledge creation, social networks and innovation: an integrative study”, Paper presented at the Academy of Management Proceedings.

[59] Uzzi, B, (1997)“Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness”, Administrative Science Quarterly , Vol. 42 ,No. 1, pp35-67.

[60] Hansen, M. T, (1999) “The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits”, Administrative Science Quarterly , Vol. 44 , No. 1, pp82-111.

[61] Mowery, D. C., J. E. Oxley, and B. S. Silverman, (1996) “Strategic alliances and interfirm knowledge transfer”, Strategic Management Journal, Vol. 17 , No. S2, pp77-91.

[62] Goerzen, A, (2007) “Alliance networks and firm performance: The impact of repeated partnerships”, Strategic Management Journal , Vol. 28 , No. 5, pp487-509.

[63] Phelps, C. C(2010) “A longitudinal study of the influence of alliance network structure and composition on firm exploratory innovation”, Academy of Management Journal, Vol. 53, No. 4, pp 890-913.

[64] Uzzi, B, (1996) “The sources and consequences of embeddedness for the economic performance of organizations: The network effect”, American sociological review, pp674-698.

[65] McEvily, B., and A. Marcus, (2005) “Embedded ties and the acquisition of competitive capabilities”, Strategic Management Journal, Vol. 26, No. 11, pp1033-1055.

[66] Hinkin, T. R, (1995) “A review of scale development practices in the study of organizations”, Journal of Management , Vol. 21, No. 5, pp967-988.

[67] Gerbing, D. W., and J. C. Anderson, (1988) “An updated paradigm for scale development incorporating unidimensionality and its assessment”, Journal of Marketing research, pp186-192.

[68] Chin, W. W, (1998) “The partial least squares approach to structural equation modeling”, Modern methods for business research, Vol. 295, No. 2, pp295-336.

[69] Fornell, C., and D. F. Larcker, (1981) “Structural equation models with unobservable variables and measurement error: Algebra and statistics”, Journal of Marketing research, pp382-388.

[70] Hulland, J, (1999) “Use of partial least squares (PLS) in strategic management research: a review of four recent studies”, Strategic Management Journal, Vol. 20 , No. 2, pp195-204.

[71] Nunally, J. 1978. Psychometric methods, McGraw-Hill, New York.

[72] Siemsen, E., A. Roth, and P. Oliveira., (2010) “Common method bias in regression models with linear, quadratic, and interaction effects”, Organizational Research Methods, Vol. 13 , No. 3, PP456-476.

[73] Ritter, T., I. F. Wilkinson, and W. J. Johnston., (2002) “Measuring network competence: some international evidence”, Journal of Business & Industrial Marketing ,Vol. 17 ,No. 2/3, PP119-138.

[74] Fang, G., X. Ma, L. Ren, and Q. Zhou, (2014) “Antecedents of Network Capability and Their Effects on Innovation Performance: an Empirical Test of Hi-tech Firms in China”, Creativity and Innovation Management, Vol. 23, No. 4, pp36-452.

AUTHORS

Gang Fang is an Associate Professor at Management School, Hangzhou Dianzi University, China. He got his PhD degree from the University of Lausanne, Switzerland in March 2009. His research interests include knowledge management and innovation networks.

Chen Chouyong is a Professor at Management School, Hangzhou Dianzi University, China. He is a professor who enjoy the State Council Special Allowance. His recent research interests focus on IT and its impact to the development of economy and enterprises.

Qing Zhou is a Professor at Management School, Hangzhou Dianzi University, China. He is the head of Institute of Management Decision and Innovation, HDU. He has been doing post- doctoral work at Beijing University of Technology since 2013. His recent research interests focus on innovation management and technology innovation alliance.

GEOGRAPHICAL COUPLING OF THE LOCATION'S ACTIVITY IN CYBERSPACE AND AGGLOMERATION OF BUSINESS SERVICE ENTERPRISES

Jian Wu[1] and Sheng Qian[2]

[1,2]Management School, Hangzhou Dianzi University, Hangzhou, China

ABSTRACT

Cyberspace is the new geographical space formed in the virtual digital world which is different from the traditional geographical system. There are some debates that whether Cyberspace lead to further spatial divergence of economic activities or not. Based on existing related research achievements this paper studies the Cyberspace's influence on the spatial distribution of business services, puts forward some hypothesizes about the relationship between the Cyberspace active location and the agglomeration of business services enterprises. By some empirical method, it measures the spatial coupling of location's activity in Cyberspace and the business service enterprises' agglomeration, and the econometric model is established to test the hypotheses. It finds that there is a high coupling relationship of location's Cyberspace activity degree and spatial agglomeration of business services.

KEYWORDS

IT, Cyberspace, Active Location, Agglomeration, Business Service Enterprise

1. INTRODUCTION

With the development of information technology, a new space is formed in the digital world different from the traditional geographic space[1][2][3][4], which is called Cyberspace in the literature research, and it is referred to as Cyberspace, Cyberspace[5][6], or network information space[7][8]. With the rapid development of information technologies, our daily life has become deeply dependent on cyberspace[9]. In the perspective of economic geography, the biggest difference between the geographic space and Cyberspace is that the traditional concept of geographical space is based on physical distance, transport infrastructure and tangible material flow. Distance is the most important factor which influences spatial distribution; the Cyberspace is based on information infrastructure, information resources, and intangible information flows, and in many respects the decisive role of distance from the spatial distribution of economic activities has been severely weakened[10].With the decline of the importance of physical location factors such as distance in Cyberspace, a series of changes are bound to occur in enterprises' location selection. Compared with the manufacturing industry, the service industry is less affected by physical location factors such as transportation costs, and its location choice is more diversified and flexible, and the exploration of service industry location theory and its model is imminent[11]. The most important change concerning cyberspace has been its permanent and instant availability to users through broadband services[12].This article focuses on the relationship between Cyberspace and location of business service industry, and studies the relevance of location's activity in Cyberspace and spatial location of business service industry.

David C. Wyld et al. (Eds) : CSITY, DTMN, NWCOM, SIGPRO - 2018
pp. 17–30, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81102

2. LITERATURE REVIEW

Some scholars tried to define the concept of Cyberspace. Gibson (1984) first proposed the concept of network space[13]. He took that Cyberspace as a huge network based on network connections. It consists of geographically dispersed multiple independent computers which are interconnected by communication lines, essentially it is a computer network information system. Jiang & Ormeling (1997) defines it as a computer-generated landscape; it is the virtual space of the global computer network; it connects all people, computers and various information resources in the world through a network[14]. Zeng Guoping (1998) believed that Cyberspace is a virtual space where the Internet realizes global computer networking and creates social life and communication through the transportation, storage, and processing of digital information[15]. Li Jiang and Duan Jie (2002) put forward that based on the comprehensive application of computer technology, modern communication network technology, and virtual reality technology, Cyberspace is a new space for people to socialize and communicate[16]. He Guosong (2006) defines Cyberspace as a computer-generated landscape, a virtual space in a global computer network that connects all people, computers, and sources of information in the world[17]. Ning(2018) points out Cyberspace is the digital world created based on traditional physical, social, and thinking spaces (PST) but in turn makes a great difference on PST[18]. In summary, existing researches have not yet formed a clear definition of Cyberspace, it basically agrees with its connotation that the Cyberspace is based on the use of information technology and forms a virtual non-material world through the long-distance transmission method of the network. It is a new form of space.

Some scholars studied the structural characteristics of Cyberspace. Batty's (1997) research on virtual geography revealed the geographical attributes of Cyberspace, and pointed out that the space and location in computer nodes and computer networks are formed by the cyclic development relationship of real place, computer space, internet and network location. [19] Cai et al. (l999) divided Cyberspace into four layers: physical layer, network layer, application layer, and knowledge and behavior layer[20]. Shiode (2000) also stratified the Cyberspace to the real world as the basis layer; above the basis layer is he infrastructure layer for electronic communications which consists by the material infrastructure of information technology; and the third layer is metaphorical space layer, such as multimedia content and network super links; the top layer is the three-dimensional virtual city and virtual places[21]. Feng Zhen (2004) believed that the development of information technology forms a virtual space that corresponds to the real space[22]. Graham(2013) focused on the usage of the ‘cyberspace’ metaphor and outlines why the reliance by contemporary policy makers on this inherently geographic metaphor matters[23]. The Cyberspace is largely rooted in physical space and location. The two often merge with each other, but there are still significant differences. From the above analysis of the layers of Cyberspace structure, the Cyberspace is a fusion of digital space and geospatial space, including communication facilities and internetworks responsible for information production and transmission, as well as the invisible digital space made up by intangible information that flows between information nodes. Therefore, Cyberspace is a hybrid space composed of information infrastructure and information resources distributed geographically, and intangible information flows in virtual space.

Some scholars studied the elements of Cyberspace. Feng Zhen (2004) compared the constituent elements of the industrialization era and the information era economy[24]. In the era of information economy, the continuous advancement of information technology has given new meanings to the elements of the original spatial structure and created new spatial combination’s mode. Hai Jiao (2008) combined the elements of point, line and surface in spatial structure, and presented the specific spatial economic agglomeration type for each combination[25]. Kitchin (2015) argued that geography continues to matter, both off- and online[26]. Based on the analysis of existing literature, this paper summarizes the constituent elements of geographic space and Cyberspace, and compares the sources of competitive advantage of the constituent elements, as shown in the table.

Table 1. Elements of Geographic Space and Cyberspace

	Geospatial components	**Competitive Advantage**	**Cyberspace component**	**Competitive Advantage**
point	Economic centre	Economic resources gathering	information Centre	Information resources gathering
line	Traffic network	Geographic reachability	Communication network	Information accessibility
surface	Economic belt	Economic radiation ability	The virtual space formed by information flow	Information flow radiation capabilities

There is some debate in academic research about the spatial effects of Cyberspace. Some scholars are pessimistic about the prospects of the city and believe that information technology indicate the role of city as a transportation node gradually disappear, and it will bring about the end of cities and agglomeration[27]. And they believed that the first three advantages in Cyberspace are weakened by the important factors that determine the location of enterprises and economic activities in traditional location theory, such as natural advantages, agglomeration economies, and transportation costs[28]. Jungyul, Tschangho & Hewings (2002) found that information technology is attractive and has spillover effect, and concentration dominant in Chicago area[29]. Kumar (2007) believed that the Internet reduces transportation costs and enables organizations in far geographically distant to share information at any time, and it leads distance to death and economic activity further decentralized[30]. Other scholars pointed out that the Internet is similar to a freeway. The "Siphon effect" of highways has made the underdeveloped rural economy a dead-living region. The same logic may also apply to the Internet[31].Information technology actually enhances the city's function. The effect makes the city more important and more powerful[32]. Hall (1999) explained why the high-end service industry would gather at the core of a few large cities through models and conducted empirical research[33]. The results showed that financial and business services, supranational organizations, and multinational corporate headquarters (originating and controlling functions) , creative and cultural industries and tourism are mainly concentrated in the world cities. Some other scholars believe that the Internet is not only a supplement to the city but also a substitute for the city[34]. It can be seen that the impact of Cyberspace on the service industry has not yet been finalized, and the conclusions drawn from studies on different types of service industries are not the same. It is necessary to specifically analyse the impact of Cyberspace on the location choice of the service industry based on industry characteristics for the subdivided industries.

Digital space methods are used to study Cyberspace inter-location relationships. Devriendt (2008) distinguishes two kinds of Cyberspace research methods: content-based analysis (CBA) and structure-based analysis (SBA)[1]. The CBA uses the information found on the World Wide Web to examine the links between cities. The hyperlinked data in search engines are used by CBA researchers to study urban spatial relationships[35]. Although the search engine is far from perfect, it still provides the largest, real-time updated, comprehensive, real-time database that meets research goals[36]. Brunn (2003) used hyperlink data from search engines to test the connections between cities in Asia and Europe[37]. And Moscow, Istanbul, Tehran, and Beijing became the centers of digital connections between Asia and Europe. Williams & Brunn (2005) classified 197 cities based on information derived from search engines and portrayed the connections between the largest cities in Asia [38]. Devriendt (2008) uses hyperlinks to analyse information networks between European city pairs[35]. London, Paris, and Berlin are the most digitally connected cities. SBA regards the level of website links to each other as a Cyberspace structure. The analysis of website structure is a common method used by SBA researchers to measure digital connections in Cyberspace. Park (2005) obtained broadband connection data from 63 countries and website data from 47 countries, examined the structure of the Internet, and found that the United States is a linking centre country[39]. Brunn (2005) used Google's search engine's URL site data to search and rank 199 world capitals and divide them into five sequences[40]. Huiping Zhang (2011) applied network content analysis method, link analysis method and network influence factor theory to propose a set of information measurement indicators for evaluating the websites of municipal governments[41]. Guangyong Zhai (2010) used nine indicators such as the total number of links and the number of external links to explore the widespread phenomenon of "information isolated islands" and its causes[42]. Mingfeng Wang (2007) established a system of indicators for evaluating the development of the Internet in cities based on the data of several important portals, and also classified the Internet use scale, the level of development and extroversion of Internet use in Chinese cities above prefecture-level cities[43][44]. It can be seen that the study of Cyberspace examined the structure of invisible virtual worlds, such as Internet hyperlinks, the structure of engine search, and the amount of email. Cyberspace research does not focus on physical infrastructure. It focuses on the production, launch, transportation, and consumption of vast intangible information between locations, interested in determine the flow of information between locations. It also analyses quantitatively and qualitatively what kind of information flows and how it flows.

From the results of research on the spatial effects of Cyberspace, the focus of attention is mainly on the spatial attributes of information flow and information resources. Some achievements have also been made in the interaction between Cyberspace and geospatial space, but so far the Cyberspace has become more economically active, the influence of spatial distribution is still rarely involved. This paper attempts to study the influence of Cyberspace on the geographical distribution of economic activities, especially the spatial distribution of business service industries. On the basis of literature review, the relevance of location activity in Cyberspace and location business service industry agglomeration is measured. It is believed that in the Cyberspace, the more digital connections between space units and other units, the more active the unit is in Cyberspace. Therefore, this paper defines the location's activity of Cyberspace as the degree of digital connection with other spatial units.

3. HYPOTHESIS

The impact of information technology brings changes in the function and shape of geospatial points and lines. In traditional geospatial space, nodes often assume the role of growth poles in the entire spatial structure system, and they are the growth centres of the region. Their scale depends on the size of the hinterland. The development of information technology has enabled cities to gradually realize the transition from traditional transportation hubs to information hubs. In the Cyberspace, information flow reshapes the urban system, prompting the transition of regional nodes from the production and consumption centres of traditional industrial products to the centre of information production and transmission and knowledge innovation. The node plays the role of regional knowledge base, information base, and source of innovation. The role and function of the node depends more on the type and quantity of inflow and outflow of information and knowledge. The information flow determines the size and importance of the city. Therefore, the nodes in the traditional geographical space are regional growth centres and are the gathering places of regional economic resources. In the modern Cyberspace, nodes are the regional information centres and the gathering places of regional information resources. In the traditional geographical space, the transportation network is the line connecting spatial nodes. A good transportation network improves the regional investment environment, changes the industrial layout and industrial structure along the line, integrates various economic resources in the region, and it has a very important role in promoting the construction of the regional economy. The Internet is the world's most important information flow carrier and infrastructure in the information age. It has constructed a new social economic model and a cyber virtual space. In Cyberspace, the Internet is the line that connects the nodes of Cyberspace. Information flow has become an important element of economic and social systems. The emergence of the information economy has reconstructed the city's competitive advantages in many areas and has become an important driving force for regional development. In these regions, the dominant infrastructure is not freeways, ports, railroads, or airports, but fibre optic networks that connect the world. Different levels of connectivity have created a global polarization of information flow and Internet connectivity, forming the archipelago economy: Cities with higher access to information gain enormous economic benefits from nodes that act as information flows, logistics, and people flow, and many regions lack an access to information and they become the edge of the information world economy. Therefore, the lines of traditional geospatial space are constituted by transportation networks, and geographic accessibility is an important competitive advantage of location; the lines of modern Cyberspace are constituted by the Internet, and information accessibility is an important competitive advantage of location.

As mentioned before, the extensive application of the Cyberspace breaks through the space-time constraints of traditional geography, and it makes geographical research into an open field full of information flow, and in particular a new standard for the study of information geography, resulting in changes in geographical thinking and philosophy. While the Cyberspace weakens the importance of the original geographic elements, it also adds a series of new elements to geography, gives new meaning, and provides new research content. In the current global network structure of the Cyberspace, regions in the information flow will gain advantages and create value and gain wealth, and areas that are out of the flow of information will lose their advantages. In this process, the region’s position in Cyberspace and the resulting concentration and proliferation of the flow of production factors directly determine the advantages of a region. At the same time, the flow of space also enhances the advantages and information flow in certain specific areas, resulting in new regional advantages. The decentralization of economic activities caused by globalization has greatly increased the necessity of strengthening the control and management functions of the centre. As a result, the high-level producer services, such as finance, accounting, advertising, and consulting, which are responsible for the central control functions, will be highly concentrated in global cities. It is an era of fierce competition. The speed of dissemination of knowledge and information is really fast. The productive service industry needs to maintain its leading edge in information resources, and the storage and updating of its own information is of utmost importance. The advantaged areas with high degree of activity in the Cyberspace are the distribution centres of knowledge and information. The smooth flow of information and knowledge can help the enterprise information and knowledge to be updated in a timely manner and help the enterprise to remain competitive. It can be inferred that the business services industry tends to gather in large cities with big information flow, and it is easy to control the decentralized economy in the context of globalization. Based on the above analysis and discussion on the influence of the Cyberspace activity on the spatial distribution mechanism of the business service industry, this paper proposes:

Hypothesis 1：The higher degree of Cyberspace activity has a positive effect on the number of business service enterprises agglomerate in space units.

At the same time, it is necessary to realize that there are many types of business services, and the information sensitivities of companies with different types of main business are different. In order to deeply analyse the influence of the location’s activity in Cyberspace on the business service enterprises agglomeration, and according to different business types of the company，this paper divides the business service enterprises into management consulting enterprises (including business management, certification, legal, consulting, investment, registration, training, testing, insurance and program enterprises, etc), advertising design enterprises (including advertising, design, exhibition, promotion, image, packaging, etc.) and agency enterprises (including branding, labour, translation, headhunting, travel, leasing, outsourcing, etc). Here are the following assumptions:

Hypothesis 2: A higher degree of Cyberspace activity has a positive effect on the number of management consulting business service companies agglomerate in space unit.

Hypothesis 3: The high degree of Cyberspace activity has a positive impact on the number of advertising design business service companies agglomerate in space unit.

Hypothesis 4: The high degree of Cyberspace activity has a positive effect on the number of mediation agency business service companies agglomerate in space unit.

4. EMPIRICAL RESEARCH

4.1. Measurement Method and Evaluation Index Selection

This paper measures the degree of geographic coupling between the location's activity of Cyberspace and the agglomeration of business services enterprises, and then establishes the empirical econometric model and makes empirical study to examine the influence of Cyberspace location's activity on the concentration of business services enterprises. In general, the smaller the spatial unit is, the easier it is to reflect the regularity and difference of the inner space of the city. China's industry and population census are usually based on blocks (townships, towns) as the smallest unit of statistics, with blocks as the basic unit, can make cities comparable, and the accuracy of the data is also high, so it is most commonly used. Therefore, this paper uses blocks as the microscopic spatial unit for statistical analysis. Regression analysis variables mainly include explanatory variables the distribution of business service industry, and the explanatory variables the Cyberspace location's activity. At the same time, introduces the control variables of economic variables as local markets, location accessibility, knowledge spillovers, industrial chains, and government.

The explanatory variable of the regression analysis is the Shanghai business service enterprises' agglomeration. In this paper, the total number of statistical data from the business administration government department is used to select valid business service enterprises data which spatial location can clearly been defined. Because a large amount of data address information is not standardized, it is impossible to directly determine its spatial location. Therefore, this part of the work has taken a lot of time and energy. The final statistics obtained 5,775 samples of business service enterprises. The samples were classified, and statistics were made based on the sample addresses. A sample space attribute database was established. Find the address of each business service company and categorize it by its block.

The main explanatory variable is the Cyberspace location's activity. The literature on the distribution of Cyberspace within the city is still very few. Only a few scholars have begun to get involved[44]. Zook (1998, 2000) has illustrated the spatial distribution of business domain names in metropolitan areas in the United States (such as New York City and San Francisco)[45][46]. The results show that these domain names are obviously concentrated in the city's central business district. The study by Malecki (2000) has similar conclusions[47]. The research team headed by Moss (1999) designed an evaluation framework for urban comprehensive portals to compare the space design of major cities in the United States and proposed optimization measures for urban Cyberspace[48]. After summarizing several models of future Internet development, Moss (1999) discussed the construction of the community-level website, finally put forward the government information strategy for urban development [49]. Since many portal sites on the micro-level spatial unit blocks have not been established, most of the websites established belong to the district and county domain names, causing the website structure analysis method to fail. This paper uses the hyperlink method of search engine to search for the number of hyperlinks of Shanghai's 253 blocks.

Control variables includes economic level, local market, location accessibility, knowledge spillover, industry chain and government. The measurement index and data sources for each variable are as follows.

Table 2: Variable measurement index and Sources of Data

variable	Measurement index	Data Sources
business services distribution	space unit business service industry concentration degree	Number of business services companies in each unit of space
Cyberspace activity	Spatial unit digital connections	hyperlinks number of spatial unit
economic level	land price	According to the “2010 Shanghai Standard Land Price Correction System”, the statistics of the land unit’s land price level from 0 to 9
local market	population density	spatial unit’s population per unit area
location accessibility	number of transport hubs within the spatial unit	According to the “Layout Plan of Shanghai's Comprehensive Passenger Transport Hubs”, 145 traffic hubs in Shanghai are assigned to five levels from 4 to 0, and the number of traffic hubs per spatial unit is calculated.
Knowledge spillover	university units in spatial unit	According to 101 universities and branch campuses, calculates the universities and their campuses within each spatial unit
Industry chain	park within spatial unit	Check the location of 63 economic development zones in Shanghai, and calculates the number of economic parks and development zones in each spatial unit
government	government agencies within the spatial unit	Whether there is an urban county government within each spatial unit

4.2. Selection of statistical and econometric models

The geographic interrelation coefficient reflects the location relationship between the two economic factors in geographical distribution. Judging the consistency of the spatial distribution of the two factors by the similarity degree in spatial structure, the formula is:

$$G = 100 - \frac{1}{2}\sum_{i=1}^{n}\left|S_i - P_i\right|$$

In the formula, G is the geographic interrelation coefficient, n is the number of blocks, and S_i、P_i are the percentage of each economic factor in each block. It indicates that the geographic interrelation of the two economic factors is relatively close, and the geographic distribution is relatively uniform when G is large; when the G value is small, it indicates that the two economic factors are not closely related to each other, and the geographical distribution is relatively different. This paper uses geographic interrelation coefficient to measure the relationship between Shanghai's business service industry agglomeration and Cyberspace location’s activity.

When the dependent variable is a discrete integer, it is a counting variable, its value is small or sometimes zero, meanwhile the explanatory variables are mostly qualitative variables, the application of the counting model should be considered. The Poisson model is widely used in the counting model[50]. Wu (1999) used the model to study the selection of foreign-owned enterprises within Guangzhou city[51]; Figueiredo (2002) studied the selection of US manufacturing units in county-level space units[52]; Hua Zhang et al. (2007) applied this model to study the relationship between location selection and accessibility of foreign-funded enterprises in Beijing[53]. However, Zhang Hua did not further verify whether the Poisson model was suitable, and did not verify whether the harsh conditions such as the conditional mean and the conditional variance were meet. Under the condition of applying the counting model, it is necessary to choose between the Poisson model and the negative binomial distribution model. In general, the conditions of the Poisson model are difficult to meet. This paper uses the negative binomial distribution model for estimation. Weiguo Lu and Wen Chen (2009) used blocks (established towns) as spatial units to analyze the location choices of Nanjing manufacturing companies using the negative binomial distribution model[54]. The log-likelihood function of the negative binomial distribution is:

$$L(\beta,\eta)=\sum_{i=1}^{N}\left\{y_i\ln\left[\eta^2\lambda_i\right]-(y+1/\eta^2)\ln\left(1+\eta^2\lambda_i\right)+\ln\Gamma(y_i+1/\eta^2)-\ln(y_i!)-\ln\Gamma(1/\eta^2)\right\}$$

T

he above parameters η^2 are estimated together with the parameter β. When the degree of dispersion of the data is so large that the conditional variance is greater than the conditional mean, a negative binomial distribution model is usually used. In this way, the conditional variance is greater than the conditional mean, and the following moment conditions hold:

$$E(y_i|X_i,\beta)=\lambda_i$$

$$\operatorname{var}(y_i|X_i,\beta)=\lambda_i(\eta^2\lambda_i)$$

Among them, η^2 measures the degree of conditional variance exceeds the conditional mean.

If the distribution of dependent variables cannot be assumed to be Poisson, then quasi-maximum likelihood estimation (QML) is performed under other distribution assumptions. Even if the distributions are incorrectly set, these quasi-maximum likelihood estimators can produce a consistent estimate of the parameters for which the conditional mean is set correctly. For these QML models, the requirement for consistency is that the conditional mean is set correctly[45].Therefore, the parameters of the QML estimation of the negative binomial distribution are used to estimate the parameters. For the fixed η^2 parameter, the quasi-maximum likelihood estimation of the parameter β can be obtained. If set correctly, the quasi-maximum likelihood estimator remains consistent even if the conditional distribution of y disobeys the negative binomial distribution.

4.3. Empirical test results and analysis

According to the statistical results, the spatial concentration of business services is still concentrated in the central of city. The largest number of business service companies are Xujiahui Block and Tianlin Xincun Block in Xuhui District, Tianmu West Road Block in Zhabei District, and Lujiazui Block in Pudong New Area. As shown in Figure 1 below.

In this paper, it use the hyperlink method of search engine to search for the number of hyperlinks of 253 blocks in Shanghai and find out the Cyberspace in Shanghai. The result is shown in Figure 1. As can be seen from the figure, Shanghai's more active location in Cyberspace centers are concentrated in several central districts such as Huangpu District, Yangpu District and Jing'an District. Chongming County, Pudong New Area, Songjiang District, Qingpu District and other suburban districts and counties are less active in Cyberspace.

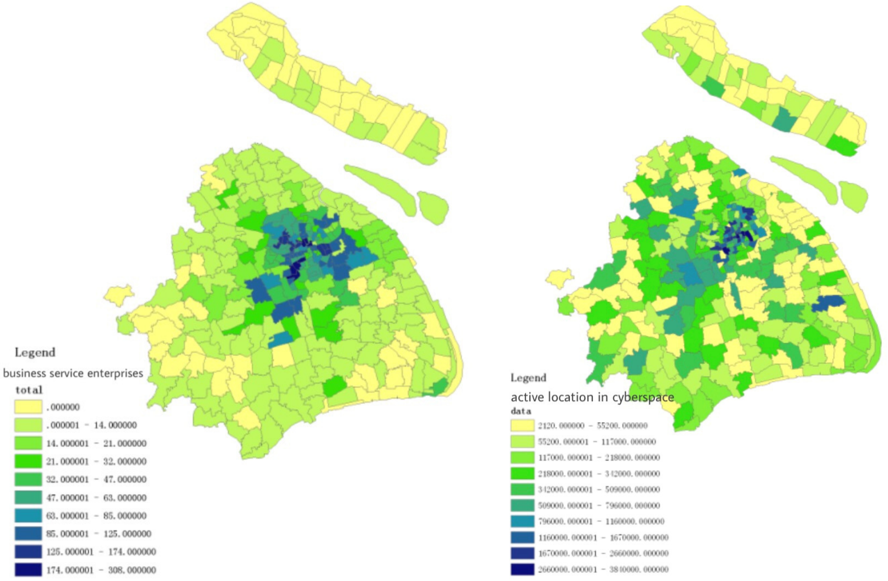


Figure 1. Overall spatial distribution of business services and Cyberspace in Shanghai

In this paper, the geographic interrelation coefficient of Shanghai's business service industry and Cyberspace activity is used to measure the geographic relationship, and the geographic interrelation coefficient of business service industry clustering and location’s activity is shown in Table 3. On the whole, the geographic interrelation coefficient between spatial concentration of business services and activity of Cyberspace is relatively high, indicating that there is a positive correlation between geographical distribution of them. For different types, the geographic interrelation coefficient of advertising design and information service industries are relatively low, and management consulting is relatively high.

Table 3: Geographical Interrelation Coefficient of Unit Business Service Agglomeration and Location’s Activity in Cyberspace

	Geographical contact rate with location activity
All business services	75.94611
Management consulting	69.30087
Advertising design	74.91884
Agency Agents	72.38692

Use Stata software to perform negative binomial regression analysis with the data. The analysis results are shown in Table 4. From the regression analysis, the location’s activity in Cyberspace significantly positively affects the overall agglomeration of business services in Shanghai and the sub-industry agglomeration. As far as business types are concerned, the degree of location's activity has less impact on management consulting and intermediary agency companies, and has a greater impact on advertising design companies. Through the regression analysis results, the following conclusions can be drawn: there are different influence of location's activity in Cyberspace when company types are different. The spatial distribution of management consulting business service companies is the least affected by the location's activity in Cyberspace, and intermediary agency companies are most affected.

Table 4: Negative Binomial Regression Results

	overall	**Management consulting**	**Advertising Design**	**Agents**
Cyberspace activity	2.82e-07***	2.22e-07**	2.62e-07**	2.33e-07**
	(2.67)	(2.11)	(2.52)	(2.14)
Local market	-1.90e-07	1.03e-07	-7.15e-07	-6.97e-07
	(-0.19)	(0.10)	(-0.67)	(-0.50)
economic level	.352379***	.3244841***	.3376388***	.3412439****
	(10.85)	(9.76)	(10.43)	(9.79)
Location accessibility	.3524257***	.3403458***	.3096915***	.3498424***
	(11.35)	(10.69)	(9.72)	(10.94)
Knowledge spillover	-.0135859	-.0332696	-.0537594	-.0025649
	(-0.22)	(-0.56)	(-0.88)	(-0.04)
Industry chain	.2940056**	.0601844	.3994586***	.2719905*
	(2.29)	(0.44)	(3.16)	(1.94)
government	.5148045**	.6646881***	.2805462	.5259932**
	(2.44)	(3.09)	(1.35)	(2.42)
constant	.8400997	-.008621	.1476751	-.4340789
	(6.86)	(-7.02)	(3.16)	(-2.98)
Number of samples	253	253	253	253
LR	306.49	284.90	264.02	270.53
Significant	0.0000	0.0000	0.0000	0.0000
P- R^2	0.1109	0.1346	0.1204	0.1410

The Z-statistics are in parentheses, *, **, and *** are 1%, 5%, and 10%, respectively, and P-R2 is Pseudo R2.

The business services industry is knowledge-intensive, this is the biggest difference between traditional service industries such as catering service and commerce. Business services industry mainly use professional talents and professional knowledge as the main producing factors, and production, application, and dissemination knowledge are the main service processes. It has the feature of talents concentration and innovation. Because of its knowledge-intensive characteristics, the business service industry is in a high-end position in the value chain, and it has a lot of links and a wide range highly value-added. Compared with the traditional manufacturing industry who requires a large amount of land resources, the business service industry is not highly dependent on land and has less requirements on environmental carrying capacity, but it has a strong dependence on urban functions. Its information sensitivity, knowledge intensity and the characteristics of innovative services have determined that their main agglomerations are center cities and core urban areas. The business service industry has a higher demand for talent density and information flow, and the strength of the urban function directly affects the flow and concentration of talent and information. Therefore, business service industry does not spread with the development of information technology, in contrast it has concentrated in cities' core where have a high degree of location's activity in the Cyberspace.

4. CONCLUSIONS

Hyperlink analysis method was used to evaluate the location's activity of Cyberspace in Shanghai. The results shows that the most active blocks are concentrated in several central areas such as Huangpu District, Yangpu District and Jing'an District. The location's activity in the suburban districts and counties such as Chongming County, Pudong New Area, Songjiang District and Qingpu District are relatively weak. Negative binomial regression is used to empirically investigate the influence of block's activity in Cyberspace on the spatial distribution of business service industry. It is found that the location's activity of Cyberspace has a positive effect on the spatial agglomeration of business services; however, there are different influences due to different types of business. The spatial agglomeration of management consulting business service companies ia the least affected by the location's activity of Cyberspace, and intermediary agency companies is most affected by the activity. It can be inferred that the location's activity in Cyberspace will promote the spatial gathering of the business service industry, but compared with other types of business service industry, the intermediary service business service industry is more likely to gather in the area with more active location. It shows that the business service industry does not spread with the development of information technology, but has been more concentrated in the core urban area with a high degree of activity in the Cyberspace.

REFERENCES

[1] Devriendt.L, Derudder.B, Witlox.F. Cyberplace and Cyberspace,(2008) " Two Approaches to Analyzing Digital Intercity Linkages", Journal of Urban Technology, Vol. 15,No. 2, pp5-32.

[2] Devriendt.L, Boulton.A, Brunn.S, Derudder.B, Witlox.F, (2011) " Searching for Cyberspace: The Position of Major Cities in the Information Age", Journal of Urban Technology, Vol. 18, No. 1, pp73-92.

[3] Dodge M . Guesteditorial, (2001) "Cybergeography", Environment and Planning B: Planning and Design, Vol. 28, No. 1, pp1-2.

[4] Dodge M , (1999) "The Geographies of Cyberspace", Centre for Advanced Spatial Analysis working paper series, paper8. University College London,1999.

[5] Luquan Jiang,Zhiren Zou,Rongzeng Liu,Feng Zhen,(2002), " Development of Cyber Geography in Foreign Countries", World Geography Research, Vol. 11, No. 3, pp92-98.

[6] Heli Lu, Guifang Liu, (2005) "Research on Geographical Distribution of Cyberspace", Scientia Geographica Sinica,Vol. 25, No. 3, pp317-321.

[7] Zhongwei Sun, Zi Lu, Yang Wang, (2007) "Review and prospect of geography research in network information space", Advance in Earth Sciences, Vol. 22,No. 10, pp1005-1011.

[8] Jie Zhang, Chaolin Gu , Jinkang Du , Yikang Zhou , (2000) "Research progress and prospects of human geography in computer network information space", Scientia Geographica Sinica, Vol. 20, No. 4, pp368-374.

[9] Li F, Li Z, (2017)"Cyberspace-Oriented Access Control: Model and Policies", IEEE Second International Conference on Data Science in Cyberspace. 26-29.

[10] Zhongwei Sun , Yang Wang, (2011) " Progress and Prospects of Research on Information and Communication Geography in China", Advances in Geography, Vol. 30, No. 2, pp149-156

[11] Malecki, E.J,(2002) "The Economic Geography of the Internet's Infrastructure", Economic Geography, Vol. 78, No. 4, pp399-424.

[12] Kellerman A, (2015)"Mobile broadband services and the availability of instant access to cyberspace". Environment & Planning A. Vol.42, No.12, pp2990-3005.

[13] O'Brien R,(1992)Global Financial Integration: The End of Geography, New York: Royal Institute of International Affairs.

[14] Cairncross F, (1997) The Death of Distance. Cambridge,MA : Harvard Business School Press.

[15] Fotheringham A. S, (1997) "Trends in quantitative methods I: Stressing the local", Progress in Human Geography, Vol. 21, No. 1, pp88-96.

[16] Fotheringham A S,Brunsdon C,(1999) "Local forms of spatial analysis", Geographical Analysis, Vol. 31, No. 4, pp340-358.

[17] Fujii,T, Hartshorn,R.P., (1995) " The changing metropolitan structure of Atlanta,GA: Locations of functions and regional structure in a multinucleated urban area", Urban Geography, Vol. 16, No. 1, pp680-707.

[18] Ning H, (2018)"General Cyberspace: Cyberspace and Cyber-enabled Spaces" , EEE Internet of Things Journal, Accepted.

[19] Bingham,R.D., Kimble,D., (1995) "Industrial composition of edge cities and downtowns", Economic Development Quarterly, Vol. 9, pp259 -272.

[20] Jian Feng , Yixing Zhou, (2003) "Spatial Structure and Evolution of Urban Societies in Beijing (1982-2000)", Geography Research, Vol. 22, No. 4, pp 465-483.

[21] Zhou Xingxing, (1996) "The suburbanization of Beijing and the thinking it caused", Geographical Sciences, Vol. 16, No. 3,pp198-205.

[22] Zongqing Zong, (2007) Research on spatial structure of commercial activities in Beijing. Ph.D. thesis of Peking University.

[23] Graham M,(2013)"Geography/internet: ethereal alternate dimensions of cyberspace or grounded augmented realities?" . Geographical Journal, Vol.179, No.2, pp177-182.

[24] Jianbin,Chen, Jing Liu, (2007) "Research on Evaluation Method of Enterprise IT Capability Based on Resource Theory", Management Science, No. 20, Suppl PP: 63-66.

[25] Hao Jiao , Aiqi Zhai , Yang Zhang, (2008) " Measurement and Efficacy of Enterprise IT Capability: Construction and Empirical Research of Local Models", Journal of Science Studies, Vol. 26, No. 3, pp596-603.

[26] Kitchin R, Dodge M, (2015)"'Placing' Cyberspace: Geography, Community and Identity", Information Technology, Vol.1, No.2, pp25-46.

[27] Moss,M.L., Townsend, A, (1997) "Tracking the Net: Using Domain Names to Measure the Growth of the Internet in U.S. Cities", Journal of Urban Technology, Vol. 4, No. 3, pp47-60.

[28] Pan Jianping, Gu Guanqun, Gong Jian,(1997) " Research on the network operation s of the CERNet backbone", Journal of Southeast University (English Edit ion), Vol. 13, No. 1, pp5-11.

[29] Zhang Nannan, Gu Chaolin, (2002) "From the geographical space to the compound space: the urban space under the influence of information networks", Human Geography, Vol. 17, No. 14, pp20-24.

[30] Liu Weidong,(2002) "Discuss About the Development of China's Internet and Its PotentialSpatialInfluence", Geographical Research, Vol. 21, No. 3, pp347 - 356.

[31] Wang Mingfeng, Ning Yuemin, (2006) "Network Advantages of Cities: An Analysis of China's Internet Backbone Network Structure and Node Accessibility", Geography Research,Vol. 25, No. 2, pp193-203.

[32] Zhang Pingyu, Liu Wenxin, Ma Yanji, (2006) "Spatial differences and changes in the development of Internet in Liaoning Province", Economic Geography, Vol. 3, pp447-450.

[33] Batty,M, (1997) "Virtual geography" Futures, Vol. 29, No. 4-5, pp337-352.

[34] Lu Zi, Shu Fang, Wang Ran, (2008) " Comparison of Realistic Geographic Space and Virtual Network Space in China", Geographical Sciences., Vol. 28, No. 5, pp601-606.

[35] Boulton,A, Lomme Devriendt, Brunn,L.S., Derudder,B., Witlox,F, (2001) ICT's for mobile and ubiquitous urban infrastructures : surveillance, locative media and global networks. Hershey, PA, IGI Global.

[36] Devriendt, L., Boulton, A., Brunn, S., Derudder, B. & Witlox, F., (2009) Major cities in the Information World: monitoring cyberspace in real-time. GaWC Research Bulletin 308.

[37] Brunn,S.(2003) "A note on the hyperlinks of major Eurasian cities. Eurasian Geography and Economics", Vol. 44, No. 4, pp321-324.

[38] Brunn,S., Williams, J., (2005) "Cybercities of Asia: measuring globalization using hyperlinks", Asian Geographer,Vol. 23, No. 1/2, pp121-147.

[39] Park,H.W. ,(2008) "Thelwall,M.Link analysis: Hyperlink patterns and social structure on politicians' Web sites in South Korea", Quality & Quantity. Vol. 42, No. 5, pp687-697.

[40] Brunn,S.D.,(2005) An E-Classification of the World's Capital Cities: URL References to Web Sites. Idea Group Inc.

[41] Zhang Huiping,(2005) " Research on Information Metrology Index of Prefecture-level Government Websites", Journal of University of Electronic Science and Technology of China, Vol. 06, pp53-56.

[42] Zhai Guangyong., (2010) "An Empirical Study of Government Websites in the Perspective of Webometrics: An Empirical Study on Information Isolated Islands", Journalism and Communication Studies, Vol. 6, pp39-44.

[43] Wang Mingfeng, Ning Yuemin, Hu Ping, (2007) "Types and Spatial Differences of Internet Development in Chinese Cities", Urban Planning, Vol. 10, pp19-22.

[44] Wang Mingfeng, Ning Yuemin, (2002) "Urban geography in cyberspace: a review and prospect", Advances in Earth Science, Vol. 17, No. 6, pp855-863.

[45] Zook,M.A , (1998) The Web of Consumption:The Spatial Organization of the Internet Industry in the United States. The Association of Collegiate Schools of Planning 1998 Conference, Pasadena,CA , 1998.

[46] Zook,M.A, (2000) “The web of production：the economic geography of commercial Internet content production in the United States”, Environment and Planning A , Vol. 32, pp411-426.

[47] Malecki, E J, (2000) The Internet: A preliminary analysis of its evolving economic geography. Global Economic Geography Conference. Singapore.

[48] Moss,M.L.,Wade,C.,Wong,J.L.,(1999)Municipal government Online:How NYC can Become the Internet City[R]. Prepared for the Office of the Public Advocate for New York and the Accountability Project Inc,New York:Taub Urban Research Center，New York University，.

[49] Moss,M.L.,Wardrip-Fruin,N.,Harrigan,P. (1999)New York City Web Guides：An In-depth Analysis of New York City’s Web Presence. New York:Taub Urban Research Center,New York University.

[50] Gao Tiemei. (2009)Econometric analysis methods and modeling. Beijing: Tsinghua University Press (second edition)

[51] Wu F, (1999) “Intrametropolitan FDI firm location in Guangzhou, China: A Poisson and negative binomial analysis”, Annals of Regional Science, Vol. 33, No. 4, PP535-555.

[52] Figueiredo O., Guimaraes P., Woodward D, (2002) Modeling industrial location decisions in U.S. counties. ERSA conference papers, European Regional Science Association

[53] Zhang Hua, He Canfei, (2007) “Location Accessibility and Location Selection of Foreign-funded Enterprises in Beijing”, Geographical Research, Vol. 26, No. 5, pp984-994.

[54] Lu Weiguo, Chen Wen., (2009) “Location Selection of Manufacturing Enterprises and Reconstruction of Urban Space in Nanjing”, Acta Geographica Sinica, No. 2, pp142-152.

AUTHORS

Wu Jian is an Associate Professor at Management School, Hangzhou Dianzi University, China. He got his PhD degree from the University of Tongji University in 2012. His research interests include IT and its impact to the development of economy and enterprises.

Qian Sheng is a Professor at Management School, Hangzhou Dianzi University, China. His research interests include IT and its impact to the development of economy and enterprises.

GENETIC ALGORITHM FOR TESTING WEB APPLICATIONS

Nashat Mansour, Ramzi Haraty, and Hratch Zeitunlian

Department of Computer Science and Mathematics,
Lebanese American University, Lebanon

ABSTRACT

We present a metaheuristic algorithm for testing software, especially web applications, that can be modelled as a state transition diagram. We formulate the testing problem as an optimization problem and use a genetic algorithm to generate test cases as sequences of events. This algorithm evolves solutions by maximizing a fitness function that is based on testing objectives such as the coverage of events, diversity of events, and continuity of events. The proposed approach includes weights that can be assigned to events. These events would lead to important features or web pages in order to ensure that test cases will be generated to cover these features. The effectiveness of the genetic algorithm is compared with that of other algorithms, namely simulated annealing and a greedy algorithm. Our experimental results show that the proposed genetic algorithm demonstrates serious promise for testing state-based software, especially web applications.

KEYWORDS

Genetic Algorithm, Metaheuristics, Search Based Software Engineering, State-Based Testing, Testing Web Applications

1. INTRODUCTION

Web applications, including Web 2.0, have evolved significantly and are no longer represented as static pages. The web is approached as a platform, and software applications are built upon it [1]. Web 2.0 applications are built around several technologies that can be executed within webpages, etc…. The new technologies introduce additional challenges for testing web application such as those of the dynamic user interface elements and states [2]. Hence, existing web testing methods [3, 4, 5] are not sufficient and new approaches are required.

In order to reduce testing costs and improve software quality, several web application testing tools have been proposed [6]. An Extended Finite State Machine (EFSM) can often be viewed as a compressed notation of an FSM. Petrenko and Boroday [7] call the state of unfolded EFSM as “configuration” and investigate the problem of constructing a configuration of sequences from an EFSM model. Memon and Pollack [8] worked on artificial intelligence planning to manage the state-space explosion by eliminating the need for explicit states. In their work, the GUI description is manually created by a tester in the form of planning operators, which model the preconditions and post-conditions of each GUI event. The planner automatically generates test cases using pairs of initial and destination transitional states. Liu et al. [9] propose a formal technique that models web application components as objects and generates test cases based on data flow between these objects. Ricca and Tonella [10] present a test generation model based on the Unified Modelling Language. These techniques extend traditional path-based test generation

David C. Wyld et al. (Eds) : CSITY, DTMN, NWCOM, SIGPRO - 2018
pp. 31–38, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81103

and use forms of model-based testing. They can be classified as “white-box” testing techniques since the testing models are generated from the web application code.

Not much research has been reported on testing web applications with dynamic features using state transition diagrams. Marchetto et al. [2] proposed a state-based testing technique designed to address the new features of Web 2.0 applications. In this technique, the DOM manipulated by AJAX code is abstracted into a state model where call-back executions triggered by asynchronous messages received from the web server, are associated with state transitions. The test cases are generated from the state model based on the notion of semantically interacting events. However, this technique generates a very large number of test cases that could limit the usefulness of the test suites. Another proposal by Marchetto et al. [11] proposed a search-based approach based on a hill-climbing algorithm to generate test sequences while keeping the test suite size reasonably small. In order to preserve a fault revealing power comparable to that of exhaustive test suite, they aimed to maximize the diversity of the test cases. The industry also proposed several functional testing tools for testing web application. Some tools rely on discovering and systematically exploring website execution paths that can be followed by a user in a web application [12]. Further approaches to functional testing are based on user session data to produce test suites [13]. Others are based on HttpUnits where the application is divided to HttpUnits and tested by mimicking web browser behaviour [14].

In this paper, we propose an effective state-based testing method, which can be used to handle the complexity of web applications. We model web applications by associating features or web pages with states and events that represent state transitions. Then, metaheuristics, namely a genetic algorithm is designed to generate a controlled number of test cases with maximum diversity and coverage. The genetic algorithm is population-based and evolves results over many generations of candidate solutions using nature-inspired genetic operators. This metaheuristic algorithm has demonstrated its effectiveness for classical software testing [15]. Also, it evolves solutions by minimizing a function that represents the testing objectives. Moreover, although the proposed method is presented for and applied to examples of web applications, it is appropriate for software applications that can be modelled by a state graph.

In the next section, we present the specificities pertaining to testing web applications, including how to build a state-based model for web applications. In section 3, we present the metaheuristic, genetic algorithm. Experimental results are discussed in section 4. Section 5 concludes the paper.

2. TESTING WEB APPLICATIONS AND STATE GRAPH MODELING

Testing is an essential part of the software development cycle. Web applications differ from traditional software development where they follow the agile software development model, which has shorter development time. Because of the short development time, web applications usually lack necessary documents during the development and the user requirements often change. Testing and maintaining web applications becomes a more complex task compared to traditional software. During the past decade radical changes were introduced to the development of web applications and even the concept of the web. The web is approached as a platform where software applications are built upon; thus, the emergence of a new generation of web applications and web systems known as Web 2.0. Web 2.0 applications are based on highly dynamic web pages, build around AJAX technologies, which through the asynchronous server calls, enable the users to interact and affect the business logic on the servers. The dynamic features of Web 2.0 add more complexity to the hard task of testing web application [16].

We propose a state based testing strategy that will dynamically generate a finite state machine from a web application by extracting interacting events [11] that produce state changes in the user interface. From the inferred graph, test cases will be generated as a sequence of events. However, generating test case sequences from the finite state machine can lead to a very large number of test cases in the test suites. This is why Marchetto et al. [11] suggested a search based approach to generate long sequences of events while keeping the test suite size reasonably large using a hill-climbing algorithm. The problem with this algorithm is that the solution will be a local optimum rather than being a global optimum.

The objective of our research is to develop a more effective state based testing for a Web 2.0 application that will cover its dynamic features. This testing approach is based on metaheuristic algorithms rather than exact graph algorithms for traversing the events in the state-based graph model. That is, the objective is to develop optimal or good suboptimal test suite that reduces the number of test cases and not merely a set of sequences of events/edges in the graph. Other considerations can be found in [16].

In addition, extracting a state graph from a Web 2.0 application is not a direct and simple task. The challenges are described in [16]. Our testing mechanism reconstructs the user interface states, and generate static pages having navigation paths each with a unique URL. These static pages will be used to conduct state-based testing [2]. To attain the static-like pages we need a tool that will execute client side code, and identify clickable elements which may change the state HTML/DOM within the browser. From these state changes, we will build our state graph that captures the states of the user interface, and the possible transitions between the states. The definition of the state graph is found in [16]. Furthermore, two issues are to be considered while building the state graph. First, we need to detect the event-driven elements; next, we need to identify the state changes. The state graph is created incrementally; initially, the state graph contains only the root state. Additional states are appended to the graph as event-driven elements are traced/invoked in the application and state changes are analyzed.

3. GENETIC ALGORITHM

Genetic Algorithms [17] simulate the natural phenomenon of populations' reproduction and selection operations in order to achieve optimal results. Through artificial evolution, successive generations search for fitter adaptations. Each generation consists of a population of chromosomes, also called individuals, and each chromosome represents a candidate solution. The Darwinian principle of reproduction and survival of the fittest and the genetic operations of recombination (crossover) and mutation are used to create a new offspring population from the current population. The process is repeated for many generations with the aim of maximizing the fitness of the individuals. In the following subsections, we describe how we generate test sequences of semantically interacting events using the genetic algorithm; an outline of the genetic algorithm is given in Fig. 1.

```
Random generation of initial population, size POP;
Evaluate fitness of individuals;
repeat
Rank individuals and allocate reproduction trials;
for i = 1 to POP step 2
        Randomly select two parents from list of reproduction trials;
                Apply crossover and mutation;
endfor
Evaluate fitness of offspring;
Save_best_so_far();
until  convergence;
```

Figure 1. Outline of the genetic algorithm

3.1. Chromosomal representation and fitness function

GA's population is an array of POP individuals (candidate solutions). An individual in the population is implemented as a vector of variable-length test cases. Each test case is represented by a sequence of a maximum of *K* events derived from the state graph. The vector's length of is *K*N*, where *N* is the maximum number of test cases required in the candidate solution. To allow variable length of test cases, we introduce a random number of fake edges into our set of valid events. These fake edges, called "No Edge", will play the role of space holder in the array.

The fitness function is composed of three weighted factors, continuity, diversity and coverage, which are explained in [16]. Briefly, testing a continuous set of events for event-based applications is likely to reveal faults. We aim to eliminate or minimize the discontinuity (*DC*) of events in a test case. Diversity guarantees that test cases will cover events from the entire scope of the web application. The lack of diversity (*LDiv*) is minimized by calculating the average frequency of events in the entire test suite. In Web 2.0 applications, end users and third parties can change the content of a web page dynamically and some events would have higher importance/weight than others. The weighted coverage (WC) is the sum of weighted coverage of events. The fitness function (to be maximized) is represented as the reciprocal of

$$\mathrm{E} = \alpha \times \frac{1}{WC} + \beta \times LDiv + \gamma \times DC$$

where α*, β, and* γ are user-defined weights which will allow flexibility in using our proposed algorithm to suit the user's particular choices.

3.2. Reproduction scheme and convergence

The whole population is considered a single reproduction unit within which random selection is performed. The reproduction scheme is based on ranking, followed by random selection of mates from the list of reproduction trials assigned to the ranked individuals. In the ranking scheme for the population, the individuals are sorted by fitness values, and ranks are assigned to individuals based on a scale of equidistant values. The ranks assigned to the fittest and least-fit individuals are 1.25 and 0.75, respectively. Individuals with ranks greater than 1 are first assigned single copies. Then, the fractional part of their ranks and the ranks of the lower half of individuals are treated as probabilities for random assignment of copies. Elitism is used to exploit good building blocks and to ensure that good candidate solutions are preserved. This is done by replacing the least-fit individual with the best-so-far individual if the latter is better than the current-fittest. Convergence is detected when the best-so-far candidate solution does not change its fitness value for 10 generations.

3.3. Genetic operators

The genetic operators are 2-point crossover and mutation at the rates 0.7 and 0.01, respectively. To apply the operators, we start with random selection of pairs of chromosomes from the mating pool, at the rate of 0.7. Each pair of these chromosomes undergoes crossover, with the condition that the randomly chosen crossover points will be aligned with the starting positions of the test cases that are represented by a sequence of *K* events (refer to section 4.1). Thus, all genes between the crossover points are swapped to create two new chromosomes. After that, mutation is applied to randomly selected genes/events, at a rate of 0.01, by randomly changing an event within a proposed test case. At this stage, the new population of offspring replaces that of the parents. We note that the genetic operators enhance the exploration feature of the algorithm in the large search space (also the fitness landscape). However, crossover and mutation may produce chromosomes that represent infeasible solutions by violating the definition of continuity. This concern is addressed by the fitness function formula, where the discontinuity term (*DC*) is assigned a large weight so that the infeasibility will be severely penalized and, thus, reduces the likelihood of survival of infeasible chromosomes.

4. EXPERIMENTAL RESULTS

The results of generating test cases using a genetic algorithm are presented in this section. These results are compared with those of simulated annealing and greedy algorithms. The simulated annealing algorithm makes use of the same objective/energy function (*E*). However, at the end of each iteration, the event frequency, coverage, and diversity matrices are saved, to be used by the energy function on the next iteration. The greedy algorithm is designed to accept only the changes that decrease the energy/objective function value, and not to allow any uphill moves. It deals with the entire test suite instead of generating a single test case after each iteration.

The three algorithms are applied on a state graph with 270 events as shown in Fig. 2. A test suite of 40 test cases is developed; each test case having a maximum of *K* test events. To deal with variable-length test case size, we append fake edges ("No Edge") to the list of events. The three algorithms were run with different test case sizes *K* = 10, 15, 20 and the execution of each algorithm was repeated ten times.

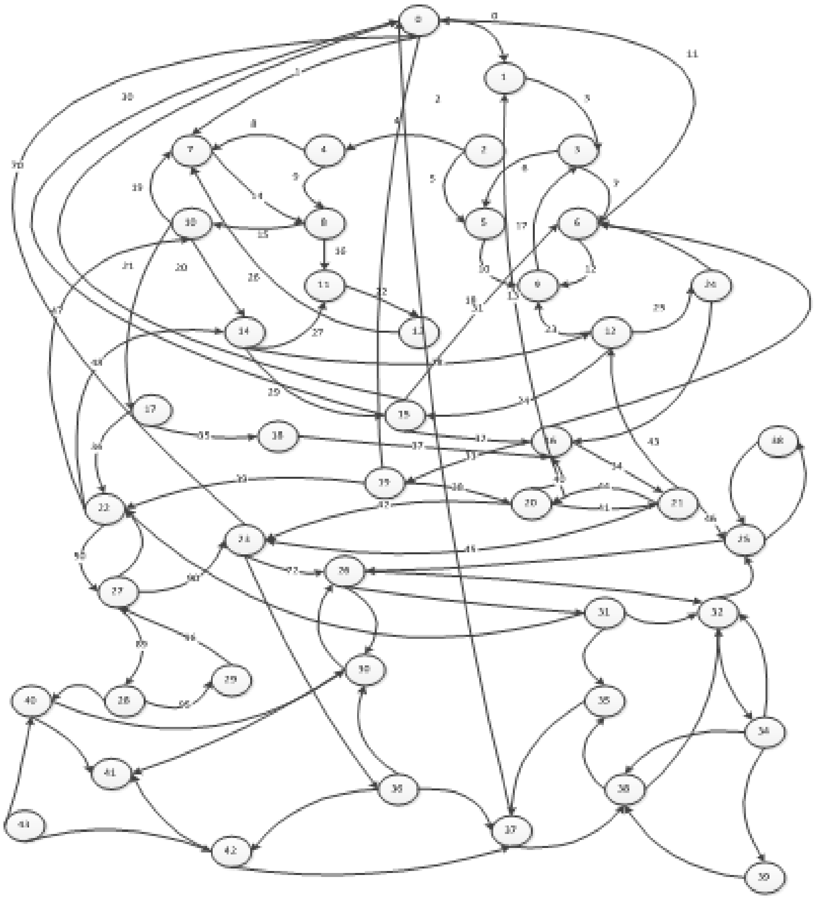


Figure 2. State graph of a web application.

The genetic and annealing algorithms successfully generated 40 useful test cases, consisting of a continuous sequence of events. The results demonstrated that the two algorithms successfully covered all the events in the application. They were also able to generate diverse suite of test cases that efficiently test different parts of the Web 2.0 application. Furthermore, the overall effectiveness of the genetic algorithm was close to that of the simulated annealing algorithm, although the genetic algorithm showed slight advantage. But, the greedy algorithm resulted in an un-optimized test suite. The greedy algorithm failed to generate continuous sequences of events in the test cases. Its objective function converged fast, within the early iterations, and no further improvements were attained. Tables 1-2 show a comparison between the best and average energy function for the three algorithms with different event numbers (maximum K) in a test case. Table 3 presents the standard deviation of the objective function values. It also shows that the genetic and simulated annealing algorithms have reasonable standard deviation, although the genetic algorithm is better on this aspect. But, the greedy algorithm has much wider standard deviation.

Table 1. Best final objective function (*E*) values of 10 runs for different test case sizes.

Max. no. of events in test cases	**Genetic Algorithm**	**Simulated Annealing**	**Greedy Algorithm**
K=10	1.26	1.35	31.00
K=15	1.82	1.81	56.16
K=20	2.73	2.41	83.32

Table 2. Average final objective function (*E*) values of 10 runs for different test case sizes.

Max. no. of events in test cases	**Genetic Algorithm**	**Simulated Annealing**	**Greedy Algorithm**
K=10	1.31	1.77	35.08
K=15	1.87	2.17	61.57
K=20	2.92	2.78	88.66

Table 3. Standard deviation of final objective function (*E*) values of 10 runs for different test case sizes.

Max. no. of events in test cases	**Genetic Algorithm**	**Simulated Annealing**	**Greedy Algorithm**
K=10	0.03	0.19	2.01
K=15	0.03	0.23	3.28
K=20	0.17	0.14	3.73

5. CONCLUSIONS

We designed metaheuristic algorithms for testing web applications. We also modeled the dynamic features of Web 2.0 using state transition diagrams. A genetic algorithm was used to generate test cases. These test cases were generated as sequences of semantically interacting events. We also formulated a fitness function that is based on the criteria of providing high coverage of events, high diversity of events covered, and definite continuity of events. The experimental results show that the genetic algorithm is a promising approach for testing web applications and state-based software.

ACKNOWLEDGEMENTS

This work was partially supported by the Lebanese American University and its contributions were equal among the three authors.

REFERENCES

[1] T. O'Reilly (2013) “Design Patterns and Business Models for the Next Generation of Software,” Retrieved from <http://oreilly.com/web2/archive/what-is-web-20.html> on July 18.

[2] A. Marchetto, P. Tonella, and F. Ricca (2008), “State-based Testing of AJAX Web Applications,” In Proceedings of IEEE International Conference on Software Testing, Lillehammer, Norway, April.

[3] A. Andrews, J. Offutt, and R. Alexander (2005) “Testing Web Applications by Modelling with FSMs,” Software and System Modelling, vol. 4, no. 3, July.

[4] A. Tarhini., N. Mansour. H. Fouchal (2010) “Testing and Regression Testing for Web Services Based Applications,” International Journal of Computing and Information Technology, vol. 2, no. 2, pp. 195 – 217.

[5] G. A. Di Lucca, A. R. Fasolino, F. Faralli, and U. D. Carlini (2002) “Testing Web applications,” In Proceedings of the International Conference on Software Maintenance, Montreal, Canada, October. IEEE Computer Society.

[6] Web Application Testing Tools. Retrieved on July 18, 2013 from <http://logitest.sourceforge.net/logitest/index.html>.

[7] A. Petrenko, S. Boroday and R. Groz (2004) “Confirming Configurations in EFSM Testing,” IEEE Transactions on Software Engineering, vol. 30, pp. 29-42.

[8] M. Memon, M. Pollack and L. Soffa (2001) “Hierarchical GUI Test Case Generation using Automated Planning,” IEEE Transactions on Software Engineering, vol. 27, no. 2, pp. 144–155.

[9] C. Liu, D. Kung, P. Hsia, and C. Hsu (2000) “Structural Testing of Web Applications,” In Proceedings of the 11th IEEE International Symposium on Software Reliability Engineering, pp. 84–96, October.

[10] F. Ricca and P. Tonella (2001) “Analysis and Testing of Web Applications,” In Proceedings of the International Conference on Software Engineering, pp. 25–34, May.

[11] A. Marchetto, P. Tonella, and F. Ricca (2009) “Search-Based Testing of AJAX Web Applications,” In Proceedings of IEEE Search Based Software Engineering, May.

[12] M. Benedikt, J. Freire, and P. Godefroid (2013) “VeriWeb: Automatically Testing Dynamic Web Sites,” Retrieved from <http://www2002.org/CDROM/alternate/654/> on July 18.

[13] S. Elbaum, G. Rothermel, S. Karre, and M. Fisher (2005) “Leveraging User Session Data to Support Web Application Testing,” IEEE Transactions on Software Engineering, vol. 31, no. 3, pp. 187-202.

[14] B. Fejes (2013) “Test Web Applications with HttpUnit,” Retrieved on July 18, from <http://www.javaworld.com/javaworld/jw-04-2004/jw-0419-httpunit.html >.

[15] N. Mansour and M. Salame (2004) “Data Generation for Path Testing,” Software Quality Journal, vol. 12, pp. 121-136.

[16] N. Mansour, H. Zeitunlian, and A. Tarhini (2013) "Optimization Metaheuristic for Software Testing," In: Schütze O. et al. (eds) EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation II. Advances in Intelligent Systems and Computing, Vol. 175. Springer, Berlin, Heidelberg.

[17] D. Goldberg, Genetic algorithms in search, optimization and machine learning, Addison-Wesley, 1989.

[18] R.A. Rutenbar (1989) "Simulated annealing algorithms: an overview," IEEE Circuits and Devices Magazine, vol. 5, Issue 1.

THE BASES OF ASSOCIATION RULES OF HIGH CONFIDENCE

Oren Segal[1], Justin Cabot-Miller[2], Kira Adaricheva[2], J.B.Nation[3] and Anuar Sharafudinov[4]

[1]Department of Computer Science, Hofstra University, Hempstead, NY 11549
[2]Department of Mathematics, Hofstra University, Hempstead, NY 11549
[3]Department of Mathematics, University of Hawaii, Honolulu HI 11823
[4]AILabs, Astana, 010000 Kazakhstan

ABSTRACT

We develop a new approach for distributed computing of the association rules of high confidence in a binary table. It is derived from the D-basis algorithm [1], which is performed on multiple sub-tables of a table given by removing several rows at a time. The set of rules is then aggregated using the same approach as the D-basis is retrieved from a larger set of implications. This allows to obtain a basis of association rules of high confidence, which can be used for ranking all attributes of the table with respect to a given fixed attribute using the relevance parameter introduced in [2]. This paper focuses on the technical implementation of the new algorithm. Some testing results are performed on transaction data and medical data.

KEYWORDS

Association rules, implication, binary table, D-basis, parallel computing

1. INTRODUCTION

In data mining, the retrieval and sorting of association rules is a research problem of considerable interest. Association rules uncover the relationships between the attributes of a set of objects recorded in a binary table. This, for example, can be a transaction table, where the objects are sale transactions and the attributes are groups of products: position (r, c) in the table, in row r and column c, is marked by 1 if the transaction r includes a product from that group c, otherwise, it is marked by 0. Association rule $X \to b$ in the table means that the entire set of transactions shows the tendency that whenever a transaction includes products from all groups in set X (i.e., there are 1s in all columns from X), then some product in group b will appear as well. The confidence of such a rule is measured by a portion of all transactions that include b among all transactions that have products from X.

In the world of transaction data, a rule $X \to b$ with confidence of 0:1 might demonstrate that b sells together with all products in group X.

David C. Wyld et al. (Eds) : CSITY, DTMN, NWCOM, SIGPRO - 2018
pp. 39–51, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81104

There is an immense effort in the data mining community to develop reliable tools for the discovery of meaningful association rules. However, the hurdles encountered while developing such solutions are numerous. The benchmark algorithms, such as Apriori in Agrawal et al. [5], have time complexity that is exponential in regards to the size of the input table. Moreover, the number of association rules is staggering, and thus analyzing them requires further tools to obtain a short subset of rules that are significant. There are no strong mathematical results confirming a particular choice of such short subsets, and numerous approaches to the filtering process are described in various publications devoted to the topic. See, for example, Kryszkiewicz [15] and Balc´azar [7]. Recent approaches include constraint-based patterns, preference learning, instant mining and pattern space sampling, which are often interactive methods targeting user's implicit preferences, see [6], [18], [19].

One particular subset of association rules, the implications, or rules of full confidence, merit particular attention in data mining. They are also at the center of ongoing theoretical research. From a practical point of view, implications are the strongest rules available in a given table because they hold true for any row of the table.

Some types of data present cases where the implications uncovered in the table contain nonessential information because the support of those implications might be very low. The support of association rule $X \rightarrow b$ is the number of rows where ones appear in all columns from X ∪ b. For example, the transaction data might have only a few implications with small sets X, whose support could be a single-digit percentage of all transactions. Mining of transaction data tends to uncover rules of lower confidence but relatively large support, rather than those implications which hold everywhere.

The apriori algorithm and the concept of generating rules of lower confidence are relevant when considering medical data. The attributes of a table that represent genetic and clinical data of patients (rows) may have a tighter connection than relationships in transaction data, in which case the confidence of relevant association rules could be expected to be high and closer to 1. Implications would serve as the imperfect representation of the laws of nature in this data. At the same time, every data set may contain errors, missing entries and miscalculations. Additionally, some patients may have extraneous conditions affecting the value in the target column (e.g., co-morbidity with an untracked illness). Even if only one row contains such deviations, it may prevent us from discovering important implications that would otherwise hold in the table.

Because errors may exist in the data in small numbers, the type of association rules one would want to discover would be those rules whose confidence is sufficiently close to one. Where ”sufficiently close” can be decided on a case-by-case basis.

In this paper, we expand the approach developed in [2] of the extraction of implications and ranking of attributes with respect to a target attribute.

Our goal is not to uncover particular rules and rank them with respect to some measurement. Rather, we want to generate a basis Δ of association rules which satisfactorily describes dependencies among attributes. We could then use Δ to rank the importance of attributes with respect to target attribute, b. In medical data, b may describe high survival probability of a patient after particular treatment, when other attributes may record physical parameters in the patients. Having large Δ is not necessarily a bad feature; on the other hand, the optimally small set is desirable. Consider the case where:

$X \rightarrow b$, $X \cup$ c $\rightarrow b$ and $X \cup d \rightarrow b$ are in the basis and have high confidence. We may want to keep A $\rightarrow b$ in Δ and remove the other rules which may unnecessarily inflate the relevance of attributes *c* and *d* for *b*. This is because attributes *c*, *d* could be completely unrelated to *b,* however only appear because they are not explicitly *blocked* elsewhere.

The paper is organized as follows.We give a short description of the proposed algorithm in section 2. In section 3, we explain how the association rules are used to compute the relevance parameter of one attribute with respect to the other. In section 4, we discuss the technical implementation of the algorithm. We discuss the performance of the algorithm and comparison with existing implementations of Apriori algorithm in section 5. In the final section, section 6, we summarize the future work that we plan to do with our approach.

2. General Flow of The Algorithm

Herein we describe a several stage approach that allows us to compute, beyond just implications, the potentially most valuable association rules whose confidence is rather high, say, > 0.9

At the core of our approach is the connection between a binary table, its implications, and the closure operator defined on the set of columns and associated lattice of closed sets, known as the concept lattice or Galois lattice, see [11].

An algorithm in Adaricheva and Nation [1] works to extract the basis of implications of the table, and it is known as the *D-basis*, which was introduced in Adaricheva, Nation, and Rand [3]. The advantage of this basis is in the possibility to use algorithms for dualization of an associated hypergraph that are known to be sub-exponential in their complexity, see Fredman and Khachiyan [10]. The algorithm in [1] avoids generating the Galois lattice from the table and only uses the arrow relations, which can be computed in polynomial time, to produce a hypergraph for each requested attribute. In that way, the existing code for hypergraph dualization, such as in Murakami and Uno [16], can be borrowed for execution.

The main idea of the current algorithm follows from the observation that the association rules of high confidence may be computed by removing one or more rows (objects) from the table and computing implications (on attributes) of the shorter table. Upon the program's execution and output, one can record only new implications that were not present in the original data set.

A new rule may be derived from the shorter table given that one of the removed rows contradicts it. If new rules are present in this table and exhibit high support, or if numerous new rules are found with average support, then the row(s) temporarily removed to form this shorter table are called blockers due to their tendency to block the rules that were found with their removal.

We can choose various strategies to identify the set of blockers. Together with several straightforward statistics on new rules found on a shorter table, we are considering several heuristics and ranking systems. The goal is to identify a set S of rows/objects that are potential blockers.

In the next stage of the suggested procedure we choose $n \leq |S|$ which is a number of rows from S that will be deleted from original table to form a shorter table. With S and n fixed, we can run algorithm *k* times, where *k* is specified by a user, and limited by $C(|S|, n) = \frac{n!}{|S|!(|S|-n)!}$, which is a number of combinations to choose n rows from set *S*. If k < *C* (|*S*|, *n*), then the choice can be done randomly, otherwise, the rows can by systematically removed.

This process of removing sets of rows and re-running the program can be organized in parallel, and all the outputs combined and aggregated following the same procedure as used to aggregate the D-basis of implications in original D-basis algorithm.

The final set of rules is guaranteed to have the probability of at least $\frac{N-n}{N}$ that each rules holds in a table, where N is the total number of rows(objects) in the table, and n is the number of deleted rows in each run of the algorithm. If *s* is the support of some implication A → *b* in a shortened table with *N* – *n* rows, then on average the support of A in *n* deleted rows will be $s \cdot \frac{n}{N-n}$.

In worst case scenario, i.e. when b = 0 in all deleted rows, the support of A → *b* on *n* deleted rows will be 0. This will give a lower estimate of the confidence of association rule A → *b* in the full table as $\frac{s}{s+s\cdot\frac{n}{N}} = \frac{N}{N+n}$.

For example, given an original table of 90 rows, the confidence of a rule found as an implication in a sub-table of 80 rows, i.e., after deleting 10 rows, will be, on average, around $\frac{90}{100} = 0.9.$

3. RANKING THE ATTRIBUTES OF THE TABLE WITH RESPECT TO A GIVEN FIXED ATTRIBUTE

The algorithm described in the previous section, when we try to identify the blockers among the objects/rows of the data, can be interpreted as unsupervised learning about the data set. These blockers are then interpreted as outliers.

In this section we will take a look at supervised exploration of the data set, when one of the parameters is a target column/attribute, and we try to discover other attributes which might be essential for describing the behavior of the target parameter.

The algorithm in [1] allows us to retrieve only those implications *X* → *b* in the D-basis that have a fixed attribute *b* as a conclusion. This is called a *b-sector* of the basis. It is important to notice, for comparison with Apriori algorithm in section 5, that one does not need to obtain the full basis in order to get particular *b*-sector of the basis.

In our current approach, instead of the set of implications, we obtained the set Δ of association rules, where we traded the confidence for the higher support of the rules. We choose a number of shorter tables, as described in algorithm of section 2, and compute b-sectors of implications. A final part of the algorithm then performs a special trimming of the rules, called an aggregation, only leaving the *strongest* rules with respect to the *binary part* of Δ, which consists of rules of the form *a* → *d*. Thus, in order to obtain the *b*-sector of basis Δ, one needs only the binary part of Δ and combination of *b*-sectors of implications of shorter tables. The resulting *b*-sector of Δ after its aggregation will be denoted Δ(*b*). Note that one can generate multiple sets Δ(*b*), thus, following computations will depend on the particular instance of Δ(*b*).

Similarly, we could make the formation of $\Delta(\neg b)$, where the original attribute was replaced by its complement $\neg b$.

Having fixed $\Delta(b)$ and $\Delta(\neg b)$, we then used the approach described in [2] to rank the attributes by the relevance parameter.

For any attribute a, the relevance $rel_b(a)$ of *a* to *b* is computed based on frequency of *a* appearing in the antecedents of implications/association rules related to *b* in *b*-sectors $\Delta(b)$ and $\Delta(\neg b)$, From this definition one can see some relation between the relevance parameter and conviction, see for example [21].

The computation of this parameter takes into account the support of each individual implication in the basis where a appears. Since this time we have association rules of different confidence, we include the confidence into computation of the relevance as well. For rule α = (*X* → *b*), *con f (α)* = $\frac{sup(X \cup b)}{sup(X)}$.

We believe that, for each attribute $a \in A \setminus b$, the important parameter of relevance of this attribute to $b \in A$ is a parameter of *total support*, computed with respect to set of rules Δ:

$$tsup_b(a) = \Sigma\{\frac{|sup(X)|}{|X|} \cdot conf(X \to b) : a \in X, (X \to b) \in \Delta(b)\}.$$

Thus $tsup_b(a)$ shows the frequency of parameter a appearing together with some other attributes in implications *X* → *b* of set Δ(b). The contribution of each implication *X* → *b,* where $a \in X$, into the computation of total support of a is higher when the support of *X* is higher, i.e., column a is marked by 1 in more rows of the table, together with other attributes from *X*, but also when *X* has fewer other attributes besides a.

While the frequent appearance of a particular attribute *a* in implications *X* → *b* might indicate the relevance of a to b, the same attribute may appear in implications $X \to \neg b$.

Replacing $\Delta(b)$ by $\Delta(\neg b)$ in above formula, we can also compute the *total support* of $\neg b$, for each $a \in A \setminus b$:

$$tsup_{\neg b}(a) = \Sigma\{\frac{|sup(X)|}{|X|} \cdot conf(X \to \neg b) : a \in X, (X \to \neg b) \in \Delta(\neg b)\}.$$

Define now the relevance of parameter $a \in A \setminus b$ to parameter b, with respect to bases $\Delta(b)$ and $\Delta(\neg b)$:

$$rel_b(a) = \frac{tsup_b(a)}{tsup_{\neg b}(a) + 1}.$$

The highest relevance of a is achieved by a combination of high total support of a in rules $X \rightarrow b$ and low total support in rules $X \rightarrow \neg b$. This parameter provides the ranking of all parameters $a \in A\backslash b$.

As we indicated above, the computation of the relevance can be done not only with implications but with any set of association rules Δ. We believe that association rules of high confidence may provide a better set for computation of the relevance.

Observation with the data shows, and theoretical results confirm [4], that a rule $A \rightarrow b$ that fails in one or a few rows of table may appear through the set of implications A ∪ $d \rightarrow b$, with multiple attributes d, which may inflate $tsup_b(a)$ for element $a \in A$. When one or a few rows failing the rule $A \rightarrow b$ are deleted, then $A \rightarrow b$ will be discovered, and the process of the aggregation will eliminate all the rules A ∪ $d \rightarrow b$ from the final set of rules used for computation of the relevance.

4. Technical Implementation

Sequential algorithms are of little practical use when dealing with sufficiently computational complex problems and/or sufficiently large problems [8]. Since the beginning of the demise of Dennard's scaling in the mid 2000's [9], the focus of mainstream computing hardware has moved away from sequential acceleration into parallel acceleration using multiple cores of execution [14]. Since then, algorithms can no longer rely on regular incremental improvements in serial execution speed due to Dennard's scaling. Instead the focus of high performance programming shifted to parallel algorithms. In addition, due to the demands of the age of big data and the decline in custom computing hardware, many of the algorithms in use today must also scale beyond the confines of a single homogeneous computing environment into a distributed heterogeneous execution environment [8].

Finding association rules is a computationally complex problem [1] and our approach is meant to address this issue using a collection of parallelizable algorithms that are scalable in a distributed processing environment.

4.1 Algorithm Descriptions

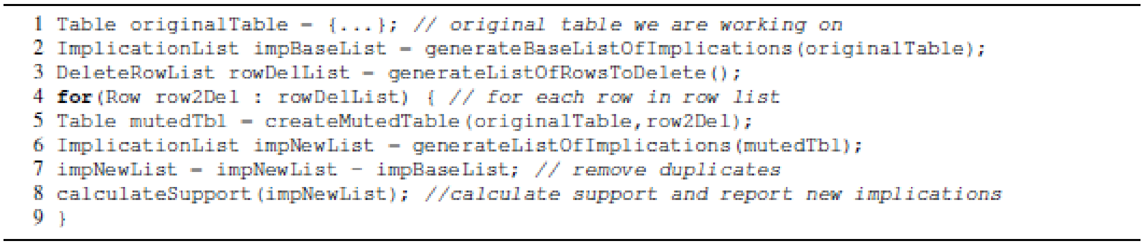
At the heart of our new approach for discovering new implications are two highly parallelizable algorithms:

1. One Row Delete Algorithm (ORD) - Remove rows from the original table one row at a time and check for new discoverable rules.

2. Multiple Row Delete Algorithm (MRD) - Remove groups of rows from the original table, aggregate the new rules and discover new rules that emerge from the aggregation.

The following code listings are written in pseudo C++ code; some code is omitted for brevity.

Listing 1.1: One Row Delete Algorithm(ORD)



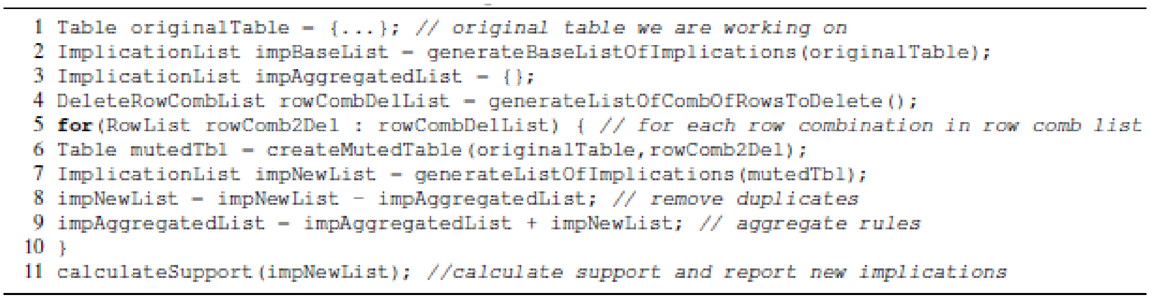
```
1 Table originalTable = {...}; // original table we are working on
2 ImplicationList impBaseList = generateBaseListOfImplications(originalTable);
3 DeleteRowList rowDelList = generateListOfRowsToDelete();
4 for(Row row2Del : rowDelList) { // for each row in row list
5 Table mutedTbl = createMutedTable(originalTable,row2Del);
6 ImplicationList impNewList = generateListOfImplications(mutedTbl);
7 impNewList = impNewList - impBaseList; // remove duplicates
8 calculateSupport(impNewList); //calculate support and report new implications
9 }
```

The ORD algorithm (seen in listing 1.1) breaks the problem of finding blocking rules in the original table into a list of *N* sub-problems where *N* is equal to the number of rows in the original table. In each of the sub problems we then need to discover new rules given a mutated original table with one of the *N* rows removed.

We start by generating the base list of implications without changes to the original table (line 2). Then we mutate the table (line 5) by removing different rows in each iteration of the loop and generate a new list of rules (line 6). In the next step we make sure the new rules are previously undiscovered by comparing them to the original rules and removing any repetitions (line 7). In the last step (line 8) we report the new rules. As can be seen in lines 5–8 of the algorithm, each iteration of the for loop can be executed in parallel, therefore allowing us to spawn up to N parallel executions of the DBasis algorithm.

Communication overhead and synchronization between parallel execution units are important factors in performance of parallel algorithms [8]. In the ORD algorithm only the original table and the list of rows to be removed need to be communicated across parallel execution unit boundaries. In addition, no synchronization is necessary since each parallel unit can have its own copy of the original table and the row/rows to remove.

Listing 1.2: Multiple Row Delete Algorithm(MRD)



```
1 Table originalTable = {...}; // original table we are working on
2 ImplicationList impBaseList = generateBaseListOfImplications(originalTable);
3 ImplicationList impAggregatedList = {};
4 DeleteRowCombList rowCombDelList = generateListOfCombOfRowsToDelete();
5 for(RowList rowComb2Del : rowCombDelList) { // for each row combination in row comb list
6 Table mutedTbl = createMutedTable(originalTable,rowComb2Del);
7 ImplicationList impNewList = generateListOfImplications(mutedTbl);
8 impNewList = impNewList - impAggregatedList; // remove duplicates
9 impAggregatedList = impAggregatedList + impNewList; // aggregate rules
10 }
11 calculateSupport(impNewList); //calculate support and report new implications
```

The MRD algorithm (seen in listing 1.2) shares much of the ideas and code with the ORD algorithm but differs in the following ways:

1. It works on groups of rows instead of one group at a time (line 6)

2. The new rules are aggregated inside the for loop (line 9)

3. we only calculate the support in the end of the loop after aggregating all the new rules found in each loop iteration (line 11)

The MRD algorithm allows us to execute lines 6–7 of each iteration of the for loop in parallel. Discovering new rules and accumulating the results of each iteration needs only to wait for individual iterations to complete in order to perform partial summing of the results (lines 8–9). The only part that needs to wait for all parallel iterations to complete is calculating the aggregated support (line 11).

5. TESTING

In addition to core programs discussed in previous section, this project develops a series of additional subroutines that perform secondary analysis of retrieved rules. We treat the output as an aggregated set, or basis of rules, that allows for statistical analysis.

First, we performed a series of tests with random matrices whose sizes mimicked real data at various densities (total number of ones in the matrix divided by the total number of entries) to investigate two things: the densities for which the algorithm may uncover a considerable number of rules, and to find the average relevance of rules in random matrices. Initial results indicate that the probability of obtaining high relevance of one attribute with respect to another remains very low for densities 0.3-0.4, and it increases when the density increases.

Testing shows that randomly generated binary tables may have pairs of attributes *a*, *b* such that the relevance parameter $rel_b(a)$ could be considerably higher than 1. Several thousand random binary tables of fixed size 20 x 32 and two different densities were analyzed, and their relevance characteristics are described in table 1. A slightly more thorough comparison is done in Fig:1, which shows the average relevance of random attributes. This data implies that a minimum relevance threshold might be recommended for practical data analysis given the possibility of unimportant attributes being ranked highly.

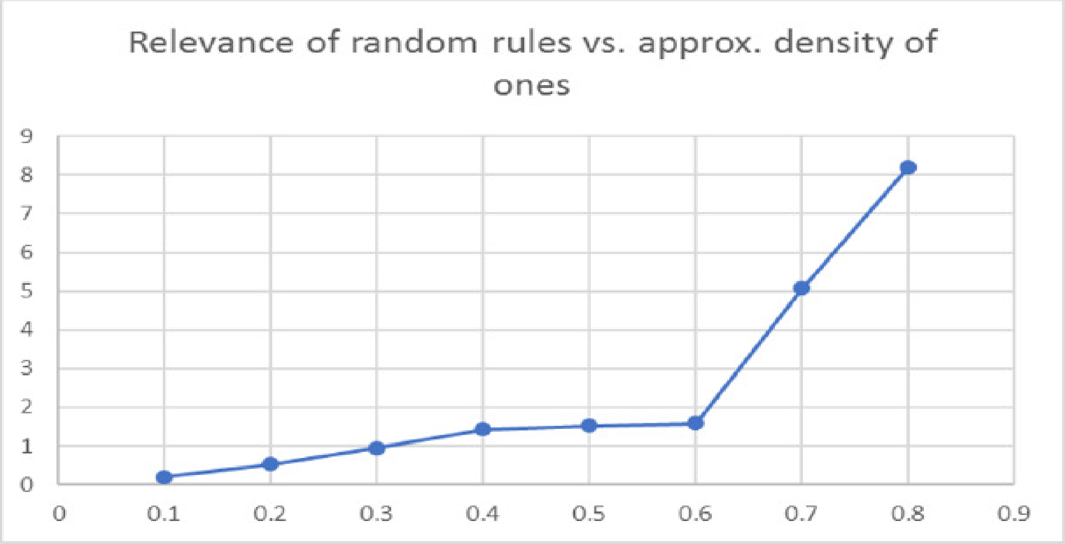
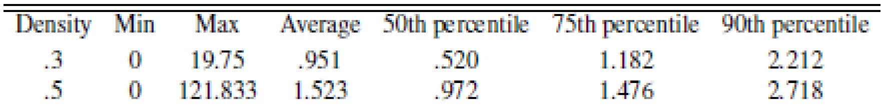


Fig. 1: A how random noise presents itself in the relevance of rules.

We also simulated data that carries a few essential rules and imposed various levels of the noise to observe whether ORD or MRD would recapture the rules which were blocked by simulated noise, defined as a certain probability p that b would become $\neg b$. In many cases the data was recaptured, although the supporting statistics were diminished. Relevance decreased inversely proportional to the amount of noise added, leveling off at around 40-50% noise. After this, many attributes' relevance statistics were indistinguishable from the rules generated by noise.

Table 1: Example of the relevance at two densities



Density	Min	Max	Average	50th percentile	75th percentile	90th percentile
.3	0	19.75	.951	.520	1.182	2.212
.5	0	121.833	1.523	.972	1.476	2.718

These studies will be presented in the full size publication when all the testing is summarized and analyzed. We are planning to use the method for analysis of medical or biological data where we are looking for the rules of confidence close to 1.

For the purposes of this short presentation we ran a comparison of our approach with existing implementations of Apriori algorithm.

We conducted our tests on Ubuntu 14.04.1(VM) running on Intel(R) Xeon(R) CPU E5-2660 v2 @ 2.20GHz with 4 cores and 16GiB of DDR3 RAM provisioned.

Two data sets were used: one transactional set and one medical.

The first data set was taken from the Frequent Itemset Mining Dataset Repository [25]. It is the retail market basket data from an anonymous Belgian retail store.We took the first 90 rows, converted them to a binary matrix format with size 90502 and (low) density 0.0162.

For our initial D-basis run, the time was 17.60 seconds, resulting in 422,273 implications of support ≥ 1.

At the time of writing, our implementation of the MRD algorithm is sequential. It takes 42.63 minutes to run the D-basis program on the original table and then a batch of 200 smaller sub-tables (with several of rows removed).

For the Apriori implementation, we initially used Microsoft SQL Server Business Intelligence Development Studio, 2008 (Data Mining Technique Microsoft Association Rules) [22]. It is worth noting that Apriori was designed specifically for mining association rules in retail data, which is normally produces the table of low density. The parameters for Microsoft were: minimal support = 3, confidence (probability)= 1.0, max number of rules = 9000, the ranking by 'importance' (= 'lift').

Most of the 9000 rules were weaker rules than those given by the D-basis algorithm. For example the top rule found by Microsoft, ranking by the lift, was 96, 48 $\rightarrow$ 95, which corresponds to equivalence 96 $\leftrightarrow$ 95 in the D-basis. On the other hand, some of the rules of support 5 and 8 from the D-basis did not appear on the 9000 list (majority of the rules in the 9000 set have support 1).

We then tested with the Apriori algorithm implemented in the R package 'arules' [12].We ran with the following parameters: minimum support = 3, max time = 10 (default), confidence (probability) = 0.8. There were 9 rules reported back from ’arules’, with only two rules having confidence < 1.

The test with R found similar set of rules as the batch of 200 runs in MRD algorithm with 10 random rows removed at a time. All the rules found in R’s Apriori function were accounted for in the output of our program. On the other hand, our batch revealed that some of rules found by our algorithm are shorter versions of those found by the R algorithm. For example, our algorithm did not report the rule 36, 48 → 38, because it reported rule 36 → 38, which is a shorter version of the rule.

Runtimes are immensely different, however, when one approaches non-sparse medical data such as the gene/survival data that was tested in [2] with the D-basis algorithm.

We re-tested again the data set treated in [2], with 291 ovarian cancer patients, split into 4 survival groups, together with expression levels of 40 genes identified as essential for this type of cancer by other methods. The goal of analysis is to find a small group of genes, say 5 to 7, that may predict the good or bad survival of a patient. The data was converted to the binary format, where the first 80 columns represented expression levels of 40 genes, while columns 81 through 84 marked the survival subgroups of patients. The ranking of 80 columns by the relevance parameter to any column b = 81, 82, 83, 84 would provide important information for medical specialists, and dependence of survival on identified subgroup of genes could be verified by other means such as Kaplan-Meier test [13].

The D-basis algorithm allows to compute only a sector of the basis, with all the rules $X \rightarrow b$ with particular consequent b. We set b = 81 and computed the b-sector of the D-basis in 135 seconds: either 200,000 implications (rules of confidence = 1) of minimum support 3, or just 91 of minimum support 5.

For the testing of MRD algorithm, we ran a batch of 25 runs of shorter tables removing randomly 20 rows at a time, requesting the rules X → 81 of minimum support 5.

Sequential time was 51.11 minutes, with roughly 125 seconds per run. The average confidence of the new rules was 0.93, and a total of 219 rules were retrieved.

Thus, together with 91 rules of confidence = 1, we found the set $\Delta(b)$ of 310 rules of high confidence. Note that no aggregation or relevance computation was performed, because the purpose of the test was the comparison with software that does the retrieval of the rules, but not the ranking of the attributes.

Note that the density of the medical data matrix is 0.34 (compared to just 0.0162 in transactional data), and the rules have a tendency to be long. Among the rules of confidence = 1 there were 35 rules $X \rightarrow$ b with $|X| = 6$ and 7 rules with $|X| = 7$. Among rules of less confidence, there were numerous rules with $|X| > 9$.

When requesting ’arules’ in R, the program runs with the user’s parameters. In order to reduce the time taken for the computation, there is a parameter *maxtime*, which limits the time used per frequent sets of attributes. This parameter stops the program when the time to produce some set of rules per subset exceeds expected times of computation.

For Apriori to analyze a more dense data set such as the medical data set, it will need to generate all frequent sets containing b = 81, 82, 83, 84, but the number of such frequent sets will be much larger than the eventual rules with consequent b. This highlights exponential time complexity of Apriori vs. the sub-exponential time it would take to analyze the same data set with the D-basis algorithm.

Requesting 'arules' on the equivalent data for the data set in [2] was unsuccessful due to the inordinate memory complexity and time complexity of the algorithm. We had successfully analyzed subsets of length 7 before the memory demands were too high. On the same machine that ran the *D*-basis program, Apriori in R ran with minimum support = 5 and maximum length of rules = 7 and took 43.67 seconds.

While faster than the D-basis program, the set of rules was restricted, and any attempt to test with larger parameters stopped execution due to large memory requirements. These rules were also limited in size so that only rules with $|X| \leq 6$ were generated.

6. CONCLUSION AND FUTURE WORK

In this paper we have discussed the development of an algorithm for the retrieval of association rules in a binary table i.e., a table consisting of ones and zeroes as another representation of the data. This is done via dualizing the hypergraph associated with the dataset, then reducing the task of rule generation to traversing this associated hypergraph via any sub-exponential time-complexity algorithm [1].

Several development proceedings were discussed, including the eventual parallelization of the program and a "top-down" method of retrieving rules which hold in all rows of the table except a few removed rows. Analyzing a slightly smaller sub-table allows to discover rules which have high confidence and may fail in a few rows due to noise in the data. The ability for this code to be parallelized and its low theoretical time complexity make it a powerful tool for data mining.

The relevance parameter for determining the importance of one attribute to an outcome was also tested to see how it might be used in analyzing real data. It was seen that even random rules could produce a notable relevance values, summarized in Fig. 1. Then, a synthetic rule was constructed and noise was added to the matrix in order to see what effect noise would have on the relevance of the parameters of the constructed rule, and whether or not the process of retrieving rules of high confidence via row deletion could recover the synthetic rule if it was blocked. The first part of these tests revealed that relevance stayed relatively stable up until approximately 30-40% noise, and that depending on the noise and how many rows were removed, the process could recover the synthetic rule albeit with less support and confidence.

Lastly, we tested several methods for "ranking" which rows should be deleted in order to recover rules lost to noise. One of tested heuristic is based upon analyzing the inverted table, however the reason for this heuristic's efficacy is not yet identified.

The development of this program is still underway, with real implementation of distributed computing expected later in 2018. We currently have several data sets in biology, medicine and meteorology which we plan to explore, working in collaboration with Biology Department, Geology, Environment and Sustainability Department of Hofstra University,as well as Donald and Barbara Zucker School of Hofstra-Northwell. We also plan to continue the collaboration with the Cancer Center of University of Hawai'i, and contribute to the exploration of data sets of various cancers, which combines several available methods [17].

ACKNOWLEDGEMENTS

We used for testing the data set on ovarian cancer available at [23], which was also used in [2] and managed in the lab of Dr. Gordon Okimoto, at Cancer Center of Univeristy Hawai'i. We thank the support of Computer Science Department of Hofstra University for providing the virtual machine environment for this project. The second author was partly supported by Honor's College of Hofstra University's Research Assistant Grant.

REFERENCES

[1] K. Adaricheva, and J.B. Nation, Discovery of the D-basis in binary tables based on hypergraph dualization, v.658 (2017), Theoretical Computer Science, Part B, 307–315.

[2] K. Adaricheva, J.B. Nation, G. Okimoto, V. Adarichev, A. Amanbekkyzy, S. Sarkar, A. Sailanbayev, N. Seidalin, and K. Alibek, Measuring the Implications of the D-basis in Analysis of Data in Biomedical Studies, Proceedings of ICFCA-15, Nerja, Spain; Springer, 2015, 39–57.

[3] K. Adaricheva, J.B. Nation and R. Rand, Ordered direct implicational basis of a finite closure system, Disc. Appl. Math. 161 (2013), 707–723.

[4] K.Adaricheva and T. Ninesling, Direct and binary-direct bases for one-set updates of a closure system, manuscript; presented in poster session of ICFCA-2018 http://icfca2017.irisa.fr/program/accepted-papers/

[5] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A.I. Verkamo, Fast discovery of association rules, Advances in Knowledge discovery and data mining, AAAI Press, Menlo Park, California (1996), 307–328.

[6] Guillaume Bosc, Marc Plantevit, Jean-Franois Boulicaut, Moustafa Bensafi, and Mehdi Kaytoue, h (odor): Interactive discovery of hypotheses on the structure-odor relationship in neuroscience, in ECML/PKDD 2016 (Demo), 2016.

[7] J.L. Balc´azar, Redundancy, deduction schemes, and minimum-size bases for association rules, Log. Meth. Comput. Sci.6 (2010), 2:3, 1–33.

[8] Gordon Bell, Jim Gray, and Alex Szalay. Petascale computational systems.Computer, 39(1):110–112, 2006.

[9] Raffaele Bolla, Roberto Bruschi, Franco Davoli, and Flavio Cucchietti. Energy efficiency in the future internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures. IEEE Communications Surveys & Tutorials, 13(2):223–244, 2011.

[10] M. Fredman and L. Khachiyan, On the complexity of dualization of monotone disjunctive normal forms, J. Algorithms 21 (1996), 618–628.

[11] B. Ganter, and R.Wille, Formal concept Analysis: Mathematical Foundations, Springer, 1999.

[12] M. Hahsler, C. Buchta, G. Bettina Gruen and K. Hornik, arules: Mining Association Rules and Frequent Itemsets, R package version 1.6-0. https://CRAN.R-project.org/package=arules 2018

[13] E.L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, J. Amer. Statist. Assn. 53 N282 (1958), 457–481.

[14] Jonathan Koomey, Stephen Berard, Marla Sanchez, and Henry Wong. Implications of historical trends in the electrical efficiency of computing. IEEE Annals of the History of Computing, 33(3):46–54, 2011.

[15] M. Kryszkiewicz, Concise representation of association rules, Proceedings of the ESF Exploratory Workshop on Pattern Detection and Discovery, Springer-Verlag, London, UK, 92–109.

[16] K. Murakami and T. Uno, Efficient algorithms for dualizing large scale hypergraphs, Disc. Appl. Math. 170 (2014), 83–94.

[17] J. B. Nation, G. Okimoto, T. Wenska, A. Achari, J. Maligro, T. Yoshioka, and E. Zitello, A Comparative analysis of MRNA expression for sixteen different cancers, preprint, http : ==www:math:hawaii:edu jb=lust 2017 615:pdf

[18] Arnaud Soulet and Bruno Crmilleux, Mining constraint-based patterns using automatic relaxation, Intell. Data Anal., 13(1):109–133, 2009.

[19] Arnaud Soulet, Chedy Rassi, Marc Plantevit, and Bruno Cremilleux, Mining dominant patterns in the sky In IEEE 11th Int. Conf on Data Mining (ICDM 2011), pages 655–664. IEEE, 2011.

[20] C. Spearman, The proof and measurement of association between two things, Amer. J. Psychol. 15 (1904), 72–101.

[21] D. Prajapati, S. Garg and N.C.Chanhan Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment, Future Computing and Informatics Journal 2 (2017), 19–30.

[22] Zhao Hui Tang, Jamie MacLennan, Data Mining with SQL server, Wiley 2005.

[23] The Cancer Genome Atlas Research Network, The Cancer Genome Atlas Pan-Cancer analysis project, Nature Genetics 45 (2013), 1113–1120.

[24] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

[25] Frequent Itemset Mining Dataset Repository, publicly available at http:// fimi. ua. ac.be /data/retail.dat

INTENTIONAL BLANK

A BFS-BASED SIMILAR CONFERENCE RETRIEVAL FRAMEWORK

Qing Guo[1, 2]

[1]Nanyang Technological University, Singapore
[2]SAP Innovation Center Network,Singapore

ABSTRACT

Literature review is part of scientific research. Online references management tools help researchers in finding relevant literature and documents. Finding relevant conferences is the key step to understand the research field. Researchers usually rely on the conference names to find out whether they are related. However, the conference name rarely reflects the diverse topics it covers. For instance, for the two conferences, "International Conference on Data Mining and Applications" and "Special Interest Group on Information Retrieval" which represent similar research topics and research areas, but the names fail to capture the similarity. One possible method to compute the similarity between all the papers in the two conferences but it's time-consuming. Instead of computing the similarity, this work builds a search engine based on Lucene and find similar conferences given a query conference based on the index. A BFS-based algorithm is proposed to address this problem and experiments on DBLP dataset shows the proposed approach can generate comparable results with the similarity-based approach.

KEYWORDS

Literature Review, Information Retrieval, Breadth-First-Search

1. INTRODUCTION

The ability of retrieving relevant information is of fundamental importance in the big data era [2]. Google, Yahoo, Bing or Baidu are popular search engines. With massive amount of data owing into the search engine, efficiency and scalability are key factors that influence the performance of the system and user experience. In this paper, searching similar conferences or journals are studied. Among the activities in the literature review, finding relevant conferences and journals is a critical step which helps the author gain an overview of the research field. A greedy approach is to get a ranking list of top-N conferences based on the similarity of the query conference and the others. In this way, the papers of the conferences need to be aggregated into one document to facilitate the similarity computation. Nevertheless, the large amount of papers makes this approach inefficient in both processing speed and storage. This work aims to address these two issues by leveraging index structure built by Lucene. Firstly, a IR system is developed for computer science publication based on the DBLP[1]dataset. An efficient algorithm based on BFS (Bread-First-Search) is developed to discover the top-N most similar conferences to query (publication venue and year, e.g., "SIGIR+1990").

[1] http://dblp.uni-trier.de/faq

David C. Wyld et al. (Eds) : CSITY, DTMN, NWCOM, SIGPRO - 2018
pp. 53–60, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81105

2. FRAMEWORK

2.1. Dataset

DBLP is a computer science bibliography website that contains more than 2.6 million of documents, published by more than 1.4 million authors stored in the form of XML. Considering the multiple dimensions for each record, it is impossible to use a DOM XML parser to infer the required attributes since the parse tree would become too big to fit the memory. SAX[2] (Simple API for XML) is utilized for parsing XML file. The information fields are extracted by the parser including authors, title, venue, year, type and paper id. In this work, the articles and the conference that compose about the 93.8% of the entire dataset are further used for implementation of the algorithm.

2.2. Lucene

Moreover, recent version of Lucene 5.0.0 supports some pre-processing/cleaning step like tokenization, stemming or lower-case normalization Lucene[3] is a project maintained by the Apache foundation which offers a complete set of APIs for building a search engine and is applied in many commercial systems. Lucene provides a complete set of tools to perform document indexing and search/ranking operations.

2.3. Indexing

To index the paper information, several operationsare conducted by Lucene:

- Extract paper information from the XML file.
- Lucene Analyzer normalizes the extracted text in different fields through tokenization, stemming, lower case and stopword removal. The tokens are saved in the document.
- The inverted index is constructed by the IndexWriter.

By indexing, users can input keywords in the various fields to trigger the search. For instance, the user can input a conference as a query (venue: SIGIR, Year: 1991) and the system will trigger the search based on this query.

3. SEARCH SIMILAR CONFERENCES

This task can be formulated as: given the paper titles of a certain conference as a query, e.g., SIGIR 2010, search 10 conferences that are most similar to the query conference. Document similarity is an extensively studied topic in text mining area. Various similarity calculation methods have been proposed including Jaccard Index, Levenshtein,N-Gram distance, etc. In this task, we propose to exploit N-Gram distance to capture the similarity between two conferences.

3.1. Similarity Computation

N-Gram model is widely applied in text mining. In the fields of computational linguistics and natural language processing, an n-gram is a contiguous sequence of n items froma given sequence of text or speech. N-Gram model has been used in a variety of NLP tasks, such as spelling

[2] http://www.saxproject.org/

[3] http://lucene.apache.org/

correction, word breaking and text summarization. Another application of N-Gram model is for extracting features for supervised Machine Learning models such as SVMs, Naive Bayes, etc. N-grams are basically a set of co-occurring words within a given window and when computing the n-grams, we typically move one word forward. For example, given a sentence “I like the information retrieval course”, when $n = 2$(also known as bigrams), then the n-grams can be generated as: “I like”, “like the”, “the information”, “information retrieval”, "retrieval course". In this task, we use N-Gram distance [1] to compute the similarity between two conferences ($sim(c_1, c_2)$). Nevertheless, the details about this technique would not covered in this report. To implement N-Gram distance, we use N-Gram Distance class in Lucene library.

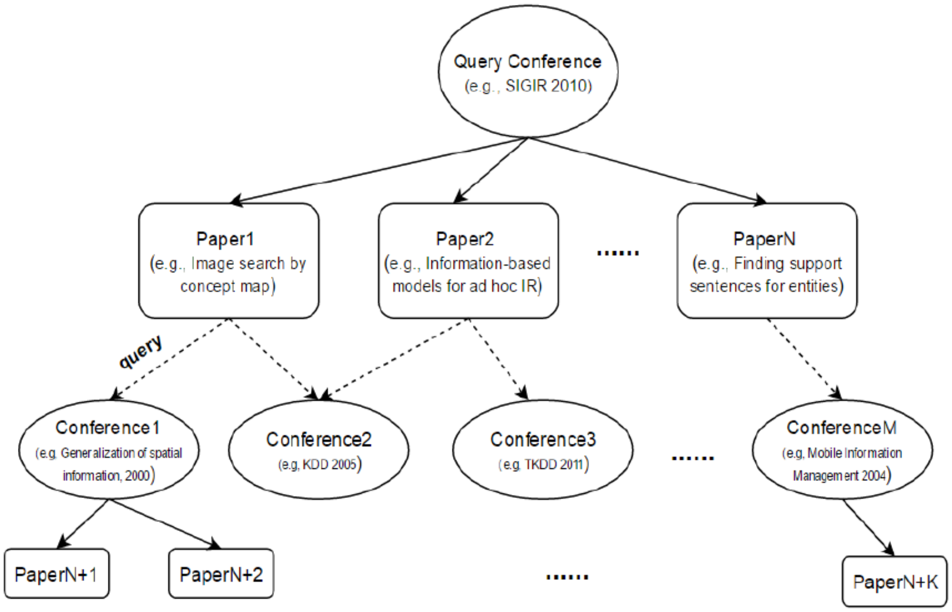

Query Conference
(e.g., SIGIR 2010)
Paper1
(e.g., Image search by concept map)
Paper2
(e.g., Information-based models for ad hoc IR)
......
PaperN
(e.g., Finding support sentences for entities)
query
Conference1
(e.g. Generalization of spatial information, 2000)
Conference2
(e.g, KDD 2005)
Conference3
(e.g. TKDD 2011)
......
ConferenceM
(e.g, Mobile Information Management 2004)
PaperN+1
PaperN+2
......
PaperN+K

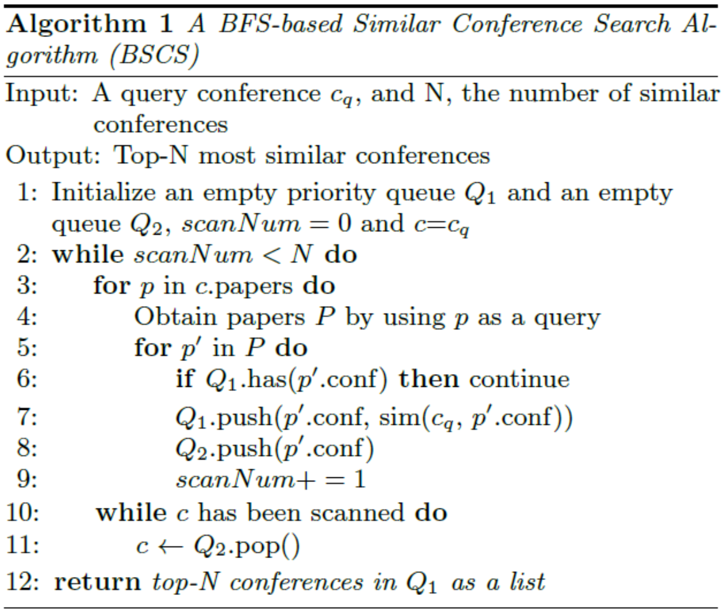
Figure 1. Tree structure of conferences and papers

Table 1. Time required for the tasks.

Task	Average time
Retrieve all paper titles of a conference	$2350ms$
Compute similarity of two conferences	$50ms$
Obtain top 10 similar conferences of a given conference	$(250 + 50) \times \frac{21000}{1000} = 6300s$

3.2. A BFS-based Similar Conferences Search Algorithm (BSCS)

An intuitive idea to search top N similar conferences is greedy search. We attempt to analyse the feasibility of greedy search by simply calculating the following statistics in Table 1. Random queries are selected to estimate the total time. For the first task, we randomly pick 20 conferences to retrieve all the paper titles of each one using the built system and each retrieval averagely takes about $250ms$. Next, 100 random conference pairs are generated, e.g., <SIGIR 1990, Pictorial Information Systems 1988>, then we apply *N Gram Distance* to them and finally the average time is $50ms$. In total, there are nearly 21000 conferences, thus for a single conference, it may take more than$1.5h$ (not include ranking) to find top 10 similar conferences, making it impossible as an efficient application. In this section, we propose a highly efficient method to address this task, named as a BFS-based similar conference search algorithm shown in Algorithm 1 and Figure 1. The searching component is triggered by the query conference of which all the paper titles, $c.papers$, are returned. Then the paper titles are taken as keywords to find similar papers, P, based on the index built by Lucene. We assume if two papers are similar, their conferences are also relevant. Hence, the conferences of these similar papers are further retrieved as $p'.conf$. The similarity of retrieved conferences and query conferences are measured by N-Gram distance which is maintained by a priority queue, Q_1. A variable $scanNum$ is set to update the number of similar conferences. If this number exceed N, then the searching stops and return the top-N elements of Q_1.



Algorithm 1 *A BFS-based Similar Conference Search Algorithm (BSCS)*

```
Input: A query conference $c_q$, and N, the number of similar
       conferences
Output: Top-N most similar conferences
Initialize an empty priority queue $Q_1$ and an empty
queue $Q_2$, $scanNum = 0$ and $c = c_q$
while $scanNum < N$ do
    for $p$ in $c$.papers do
        Obtain papers $P$ by using $p$ as a query
        for $p'$ in $P$ do
            if $Q_1$.has($p'$.conf) then continue
            $Q_1$.push($p'$.conf, sim($c_q$, $p'$.conf))
            $Q_2$.push($p'$.conf)
            $scanNum += 1$
    while $c$ has been scanned do
        $c \leftarrow Q_2$.pop()
return top-N conferences in $Q_1$ as a list
```

Figure 2. A BFS-based Similar Conference Search Algorithm (BSCS)

3.3. Performance Analysis

3.3.1. Efficiency Analysis

Note that the main motivation of our proposed algorithm is retrieving top 10 similar conferences at an acceptable speed. Though there are other approaches for this purpose we also considered, such as clustering documents by either TF-IDF features or topics. In this direction, given a query conference and its cluster label, we could only take the conferences belonging to the same cluster as candidates. However, this approach would be extremely time-consuming and require high-performing machines to provide high computation power (it has taken us more than 2 days for an unfinished clustering task), forcing us to give up this method.

Without compromising the quality of results, we take an economic strategy that make the full advantage of already built index by Lucene. Having tested 100 queries, the average time is about 45 seconds (ranging from 8*s* to 75*s*) which could be acceptable for users due to the large-scale data to process. Meanwhile, an additional advantage of our algorithm is that it does not rely any intermediate results from other algorithms (e.g., topic model or clustering), saving the system huge amount of storage and reducing the complexity of this job.

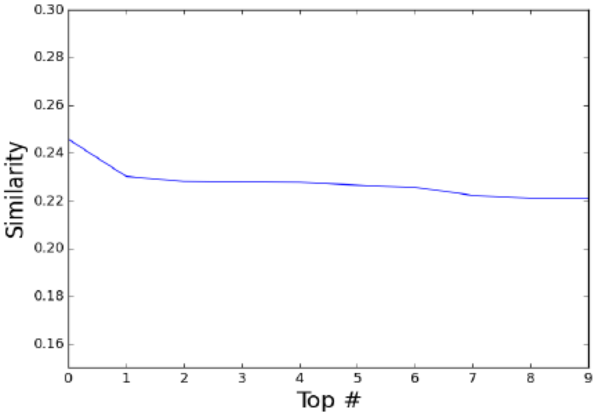


Figure 3. Similarity between Top 10 results and "SIGIR 1990"

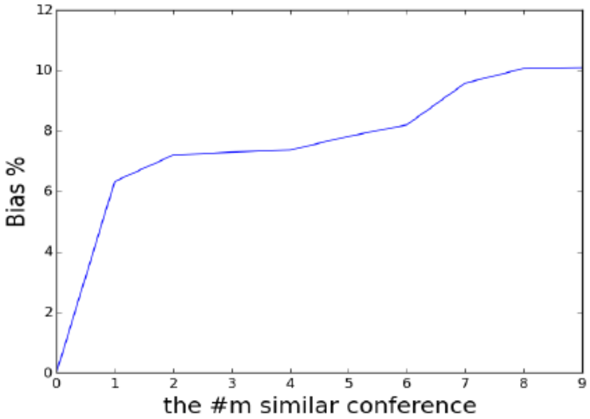


Figure 4. "SIGIR 1990": Difference between R_m and "SIGIR 1991"

3.3.2. Effectiveness Analysis

The effectiveness of BSCS will be evaluated from two perspectives: quantitative and qualitative evaluation.

- **Quantitative evaluation.**

Several query conferences from various research areas are tested and almost all the similarity lies in the range from 0.20 to 0.25. "SIGIR 1990" is chosen as an example presented in Figure 2. However, the evaluation only relying on similarity value is unfair to demonstrate the effectiveness of BSCS since similarity itself depends on the data quality. Thus, we conduct comparison analysis that can ease this issue. Since it would take much time (>1.5h) to retrieve the top 10 similar conferences given a conference by greedy search, we adopt an assumption that two conferences ($C_{v,i}$ and $C_{v,i+1}$) published in the same venue (v) and in consecutive years (i and $i + 1$) are similar, for example,"SIGIR 2010"and "SIGIR 2011" are considered to be similar. For a particular conference $C_{v,i}$, the bias between its mth similar conference (Rm) and $C_{v,i+1}$ is calculated as:

$$bias(R_m, C_{v,i+1}) = sim(C_{v,i}, R_m) = \frac{sim(C_{v,i}, R_m) - sim(C_{v,i}, C_{v,i+1})}{sim(C_{v,i}, C_{v,i+1})}$$

We choose “SIGIR 1990” and “Telecommunication Systems 2003” from different research areas as examples shownin Figure 3-4. In total, 20 conferences are chosen to be evaluated as shown in Figure 5. From the plotting, the difference between the results by BSCS and $C_{v,i+1}$ almost all lie within 10% which proofs that BSCS algorithm is effective.

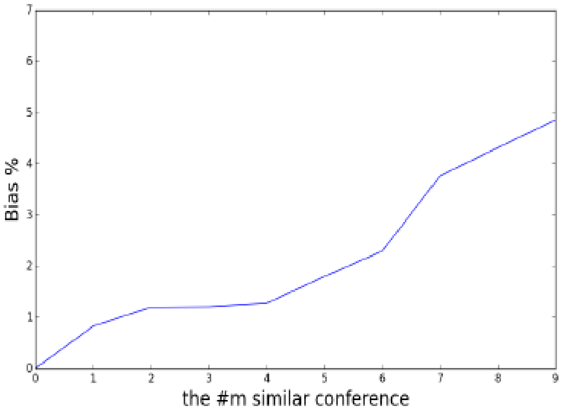


Figure 5. “Telecommunication Systems 2003”: Difference between Rm and “Telecommunication Systems 2004”

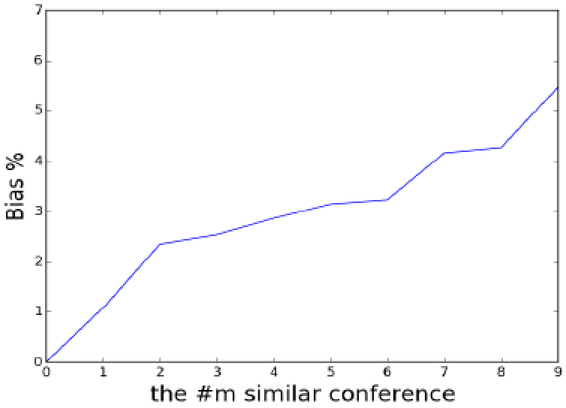
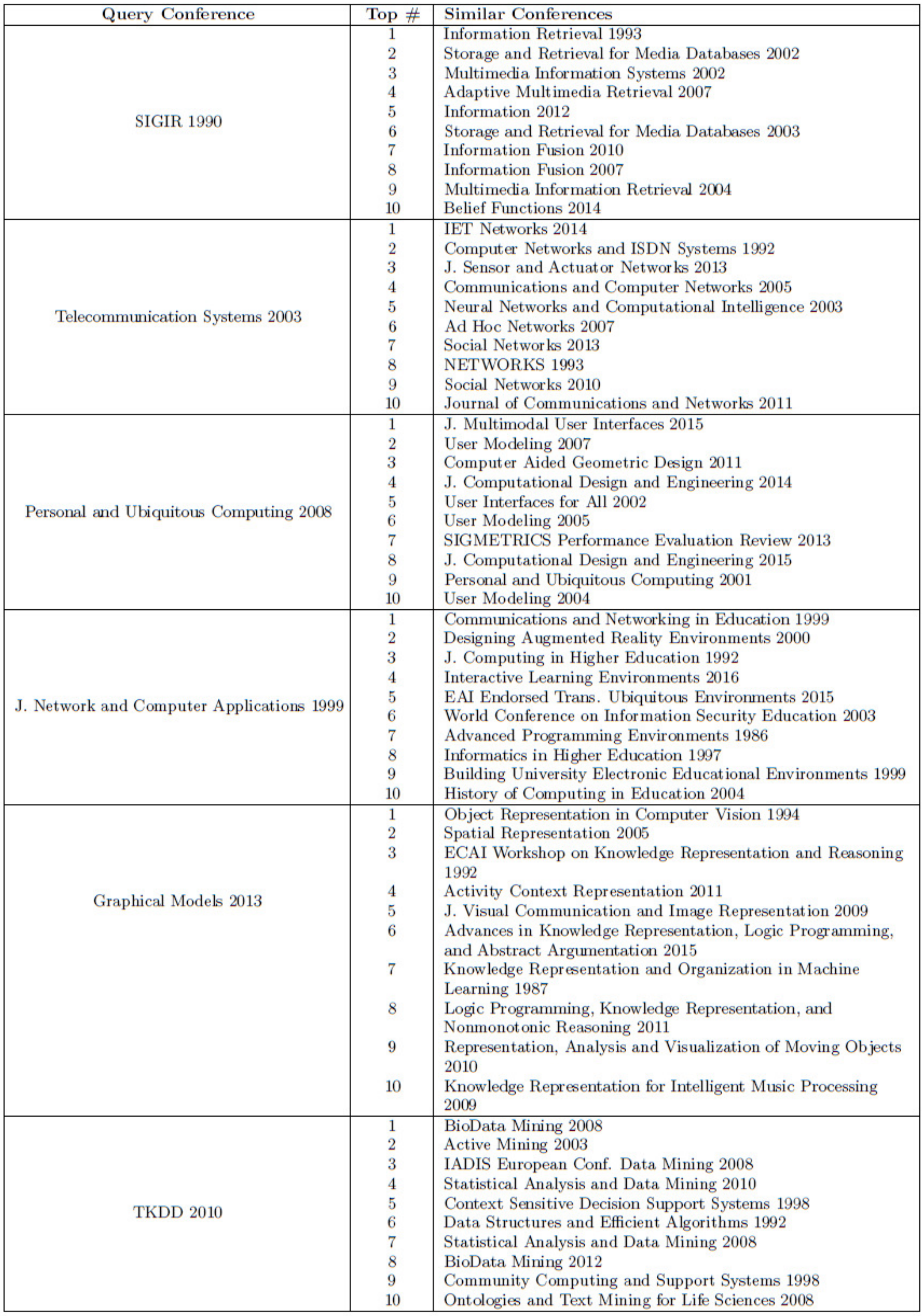


Figure 6. The Evaluation of 20 Conferences

- **Qualitative evaluation.**

Due to the space limitation, the results of 6 query conferences are shown in Table 2. From the conference names, we can observe that BSCS indeed can extract similar conferences. Taking “TKDD 2010” as an example, the conference names contains words like “computing”, “mining”, “learning”. More significantly relevant phrases such as “data mining”, “text mining” and “statistical analysis”, etc. appear in the results, highlighting the relevance between “TKDD 2010” and the results.



Query Conference	Top #	Similar Conferences
SIGIR 1990	1	Information Retrieval 1993
	2	Storage and Retrieval for Media Databases 2002
	3	Multimedia Information Systems 2002
	4	Adaptive Multimedia Retrieval 2007
	5	Information 2012
	6	Storage and Retrieval for Media Databases 2003
	7	Information Fusion 2010
	8	Information Fusion 2007
	9	Multimedia Information Retrieval 2004
	10	Belief Functions 2014
Telecommunication Systems 2003	1	IET Networks 2014
	2	Computer Networks and ISDN Systems 1992
	3	J. Sensor and Actuator Networks 2013
	4	Communications and Computer Networks 2005
	5	Neural Networks and Computational Intelligence 2003
	6	Ad Hoc Networks 2007
	7	Social Networks 2013
	8	NETWORKS 1993
	9	Social Networks 2010
	10	Journal of Communications and Networks 2011
Personal and Ubiquitous Computing 2008	1	J. Multimodal User Interfaces 2015
	2	User Modeling 2007
	3	Computer Aided Geometric Design 2011
	4	J. Computational Design and Engineering 2014
	5	User Interfaces for All 2002
	6	User Modeling 2005
	7	SIGMETRICS Performance Evaluation Review 2013
	8	J. Computational Design and Engineering 2015
	9	Personal and Ubiquitous Computing 2001
	10	User Modeling 2004
J. Network and Computer Applications 1999	1	Communications and Networking in Education 1999
	2	Designing Augmented Reality Environments 2000
	3	J. Computing in Higher Education 1992
	4	Interactive Learning Environments 2016
	5	EAI Endorsed Trans. Ubiquitous Environments 2015
	6	World Conference on Information Security Education 2003
	7	Advanced Programming Environments 1986
	8	Informatics in Higher Education 1997
	9	Building University Electronic Educational Environments 1999
	10	History of Computing in Education 2004
Graphical Models 2013	1	Object Representation in Computer Vision 1994
	2	Spatial Representation 2005
	3	ECAI Workshop on Knowledge Representation and Reasoning 1992
	4	Activity Context Representation 2011
	5	J. Visual Communication and Image Representation 2009
	6	Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation 2015
	7	Knowledge Representation and Organization in Machine Learning 1987
	8	Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning 2011
	9	Representation, Analysis and Visualization of Moving Objects 2010
	10	Knowledge Representation for Intelligent Music Processing 2009
TKDD 2010	1	BioData Mining 2008
	2	Active Mining 2003
	3	IADIS European Conf. Data Mining 2008
	4	Statistical Analysis and Data Mining 2010
	5	Context Sensitive Decision Support Systems 1998
	6	Data Structures and Efficient Algorithms 1992
	7	Statistical Analysis and Data Mining 2008
	8	BioData Mining 2012
	9	Community Computing and Support Systems 1998
	10	Ontologies and Text Mining for Life Sciences 2008

Figure 6. Results of 6 conferences

4. CONCLUSIONS

This paper proposes an algorithm that can help researchers find similar conferences. The proposed BFS-based framework is based on Lucence. Our proposed algorithm can efficiently retrieve relevant conferences compared with traditional methods. For evaluation, we employ DBLP dataset to measure the performance both quantitively and qualitatively.

REFERENCES

[1] G. Giannakopoulos, V. Karkaletsis, G. Vouros, and P. Stamatopoulos. Summarization system evaluationrevisited: N-gram graphs. ACM Trans. Speech Lang.Process., 5(3):5:1{5:39, Oct. 2008.

[2] S. Lawrence and C. L. Giles. Accessibility of information on the web. Intelligence, 11(1):32{39, Apr. 2000.

AUTHORS

I am currently a 4th year PhD student majored in Information System in Nanyang Technological University and SAP Innovation Center Network, Singapore. My research interests are recommender system, information retrieval and spatial-temporal data analysis.

NETO-APP: A NETWORK ORCHESTRATION APPLICATION FOR CENTRALIZED NETWORK MANAGEMENT IN SMALL BUSINESS NETWORKS

Dewang Gedia and Levi Perigo

Interdisciplinary Telecom Program, University of Colorado Boulder
530 UCB, Boulder, Colorado, USA 80309

ABSTRACT

Software-defined networking (SDN) is reshaping the networking paradigm. Previous research shows that SDN has advantages over traditional networks because it separates the control and data plane, leading to greater flexibility through network automation and programmability. Small business networks require flexibility, like service provider networks, to scale, deploy, and self-heal network infrastructure that comprises of cloud operating systems, virtual machines, containers, vendor networking equipment, and virtual network functions (VNFs); however, as SDN evolves in industry, there has been limited research to develop an SDN architecture to fulfil the requirements of small business networks. This research proposes a network architecture that can abstract, orchestrate, and scale configurations based on small business network requirements. Our results show that the proposed architecture provides enhanced network management and operations when combined with the network orchestration application (NetO-App) developed in this research. The NetO-App orchestrates network policies, automates configuration changes, and manages internal and external communication between the campus networking infrastructure.

KEYWORDS

Ansible, Automation, Flask, Network Management System, Network Programmability, NetO-App, OpenStack, OpenContrail, OpenFlow, Orchestration, Python, SDN.

1. INTRODUCTION

Software-defined networking (SDN) pronounced its presence in networking, but the attempts to create SDN architectures that can address the needs of small business networks have been limited [13]. According to Cisco Systems, small business networks require highly secure and reliable data networks that meet rigorous requirements such as remote workers, accessing customer data from any place and time, and cost-effective support of new applications [31]. Small business network infrastructures that employ network elements such as cloud operating systems, virtual machines (VMs), containers, traditional vendor networking equipment, and virtual network functions (VNFs); constantly need efficient control and configuration management mechanisms to dynamically cater to changing workloads. Such an environment is subjected to software and hardware restrictions, repetitive deployments and configurations, and dynamic business requirements. Small business enterprises, which are typically defined as possessing less than 500 employees in the United States and less than 250 employees in Europe, need to adopt an

David C. Wyld et al. (Eds) : CSITY, DTMN, NWCOM, SIGPRO - 2018
pp. 61–72, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81106

infrastructure that is efficient to configure and manage, inexpensive to deploy and operate, highly scalable, easy to operate, and secured from internal and external threats [32,17].

While SDN can be difficult to define, the Open Networking Foundation (ONF) defines an SDN architecture as a networking model that is directly and programmatically configured, decouples the network control functions from the forwarding functions, logically centralizes the control, and is open standards-based and vendor-neutral [3]. In a small business networking environment, the infrastructure incorporates both SDN and traditional devices and must use an architecture that can flexibly manage both traditional and SDN domains [13].

Traditional network engineering relies on device configuration via the command line interface (CLI) and does not scale to meet the complexity of multi-vendor SDN/traditional networks in small businesses. Programmability of traditional devices is cumbersome because they lack open, programmable interfaces, which prohibits developers from programming the network in the most efficient method [21,22]. Furthermore, integrating SDN and traditional networks is difficult due to the disparities between how they function: traditional networks operate with the help of MAC address tables and routing tables, whereas SDN with OpenFlow uses flow entries in flow tables. These disparities need a different methodology to integrate as a system, and research indicates that only a limited number of tools can handle these problems efficiently [23].

Network automation reduces the manual effort required for completing routine tasks and decreases the amount of human error caused by traditional, manual CLI configurations. Starting with scripting and progressing to intelligent network control and efficient translation and deployment of network plans and policies, network automation is a key tool to facilitate traditional network management and operations. While using information from configuration files and deploying routine configurations onto multiple network devices is a step towards automation, this approach can be made more dynamic by creating a graphical user interface (GUI) that automates configuration from minimal user input, simplifies the process, abstracts the network infrastructure from the programmer, because it does not require the programmer to know vendor-specific CLI commands, and reduces the number of misconfigurations [25].

Network programmability coupled with network automation can address SDN and traditional network limitations and can also provide a better platform for centralized configuration management of the cloud infrastructure in small business networks. Cloud computing is a rapidly growing paradigm for consuming data center resources in the form of Services: Platform-as-a-Service (PaaS), Infrastructure-as-a-Service (IaaS), and Software-as-a-Service (SaaS) [15]. Private cloud offers small business networks control over the infrastructure, choice of hardware/software tools, and offers control over the desired network security; thus, minimizing the scope of network vulnerability from external attacks. The benefits network virtualization and cloud computing offer when combined with SDN and network automation provide a framework that is suitable for small businesses.

The remainder of the paper is organized as follows: Section 2 provides a review of the existing body of knowledge, state of the art applications, and how our scheme extends it. Sections 3 and 4 describe the methodology and results of our experiment respectively. Section 5 concludes our research and addresses scope for future enhancements.

2. RELATED WORK

SDN has changed management of network infrastructure by decoupling the network control plane and the data plane [17]. With the help of ONF, there has been wide scale industry adoption of the OpenFlow protocol as the standard southbound interface (SBI) to communicate with pure and hybrid OpenFlow SDN switches. Although there have been attempts to create network architectures that are easily manageable, scalable, fault-tolerant, and inexpensive, there has been limited results that meet all these requirements for small businesses [13].

Small business network environments are constrained by limited software, hardware, and network capabilities. They also are hindered by repetitive tasks, limited administrative skillset, and the capability to dynamically adapt to constantly changing user workloads. Furthermore, the confined financial budgets dedicated to small business network environments prove to be a monumental restriction. In such an agile and restrictive environment, it becomes essential for the small business network infrastructure to meet or exceed these minimum requirements to have state-of-art network facilities with limited resources. The NetO-App developed from this study addresses these limitations by using free and open source software, a user-friendly web interface, and provides an advantage of having an automated and orchestrated infrastructure that reduces operating cost and increased ROI [27]. NetO-App efficiently addresses these requirements by using OpenStack Kolla as an orchestrator that deploys Docker containers that are lightweight, quickly scalable, require less storage space thus, catering to small business network requirements.

With the advent of cloud computing, more Internet applications such as DNS, DHCP, and web servers are deployed in the cloud. To manage these applications, a greater level of automation and orchestration is required. SDN helps build a level of abstraction and orchestration for VM management where hypervisors leverage the real-time network information before migration to minimize network-wide communication costs of resulting traffic dynamics [1]. A large-scale SDN capable infrastructure, the OF@TEIN playground, was initially targeted to build and operate OpenFlow enabled networks, but shifted its efforts to establish an open and shared consortium for new potential collaborators with the intention to build and operate a federated multi-site SDN-Cloud-leveraged infrastructure using the ONOS SDN controller, OpenStack cloud, and Quagga router [2]. Using Quagga to facilitate the transition from a traditional network to an SDN has provided a platform for exchanging border gateway protocol (BGP) routing information [3]. Although such an architecture provides inter-platform networking capabilities, it still lacks centralized application for configuration and management of a multi-platform infrastructure that our proposed model provides.

One of the motivations for developing SDN was to overcome challenges faced by data centers. An architectural framework provided by ONF is Central Office Re-architected as Data Center (CORD). This platform helps service providers deliver a cloud-native, open, and programmable platform to enable services to end-users [4]. This architecture primarily enables residential, mobile, and enterprise subscribers to appropriately route traffic using defined network policies residing on the XOS (CORD controller) node. While CORD tackles policy based routing through XOS, the framework, lacks necessary components for deploying and scaling VMs/containers, and incorporating multi-vendor traditional network hardware present in small business network environments. The proposed NetO-App specifically addresses such business network requirements and appropriately automates the VM/container deployment.

Open Network Automation Platform (ONAP) aims to provide a comprehensive platform for the real-time deployment and policy-driven orchestration of network functions for cloud providers and operators to automate new services [20]. Additionally, it offers the capability to monitor the service behavior based on the specified design and provides healing capabilities by scaling the resources to adjust any demand variations. Although the ONAP platform can deliver service design, creation, and lifecycle management in an OpenStack VM environment, it lacks capabilities to host and monitor the VNFs in a container platform which increases the expense of this platform due to the exponential storage space required for VMs over containers. Furthermore, the amount of dedicated hardware required to operate ONAP is difficult for small business network environments. NetO-App addresses these limitations by leveraging Ansible to proactively monitor containers and VMs hosted in network, and can be deployed on a single server. Another disadvantage of ONAP is that it has a complex architecture and needs a thorough understanding of every module to tailor desired services needed for a small business network. NetO-App provides a simple architecture that is easy to control through a user-friendly web portal to dynamically create VMs/containers.

The COSIGN project highlighted the integration of SDN controllers and the OpenStack orchestrator for optimizing the selection of resources in a virtual data center [5]. A fabric topology using SDN helps overcome the bandwidth utilization and network scalability challenges posed by fat tree topology [6]. However, the approaches still fail to deliver self-healing incident-response (pre-defined) capabilities in an orchestrated cloud environment that NetO-App addresses.

As shown in [7], integrating OpenStack with SDN provides benefits when managing complex and virtualized applications. With a better GUI, it is easy to manage SDN topologies which were demonstrated by integrating OpenStack and the Ryu SDN controller [7]. While transitioning from traditional networks toward centralized SDN, SDN placement planning can help achieve better controllability in the early 70% of the deployment [8]. Using the OpenDaylight (ODL) SDN controller for achieving network programmability (flow control and network isolation) in an OpenStack environment through the provided Neutron plugin can help provide centralized management in cloud operating systems as well [9]. However, the Neutron service fails to provide the necessary encapsulation for the traffic flowing between different tenants. Our proposed model overcomes this limitation by employing the OpenContrail SDN controller.

As new SDN design architectures emerge, a framework is required to manage and coordinate different implementations. The concept of a network hypervisor was introduced in [10] which provides a platform to use the existing low-level application programmable interfaces (API) provided by different SDN implementations in an autonomous system and convert it to high-level APIs. This can ease the task of creating an SDN; however, it fails to leverage capabilities to remotely configure and manage infrastructure through a centralized application. It was found in [12] that when integrated into an OpenStack environment, ODL has inferior performance in terms of delay and throughput when compared to Floodlight and Ryu, but the ODL controller showed higher resiliency.

In this paper, the primary research question we answered was “**Can an application provide centralized network management to configure and manage multi-platform, software-defined, and traditional networking environments for small business networks?**" This research question was strategically divided into subproblems that addressed an individual technological research aspect to collectively answer the primary research question.

A. Can we achieve orchestrated control to configure and manage multi-platform network elements using an application?

B. Can we achieve a scalable and resilient control infrastructure for remote network management and configuration?

C. Can BGP be used to facilitate interconnectivity between VM/containers, SDN, and traditional networks?

The contribution of this paper is to design and implement a network architecture and application that can be used together to orchestrate multi-platform environments, such as virtualized cloud, SDN, and traditional networks in small business networks.

3. RESEARCH OVERVIEW

3.1. Environment

The network architecture developed for this research was comprised of an overlay and underlay network (Fig. 1). The overlay network consists of a multi-node OpenStack setup which is operating on five x86 servers, VMs with Docker containers for specific applications, OpenContrail and Floodlight SDN controllers, and the network orchestration application (NetO-App) developed from this study. The underlay network is composed of x86 servers running Ubuntu, OpenFlow-capable switches including Arista, Cisco, Dell, HP, OvS, OpenSwitch, and Pica8, which establish an OpenFlow v1.3 channel to the Floodlight container for the SDN, and traditional networking equipment including ADTRAN, Arista, Cisco, and Juniper.

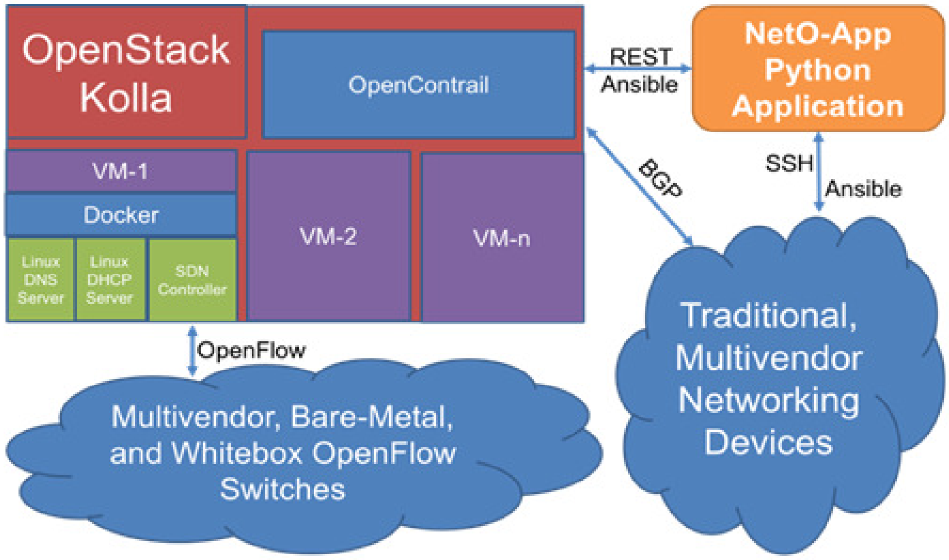

OpenStack Kolla
OpenContrail
VM-1
Docker
Linux DNS Server
Linux DHCP Server
SDN Controller
VM-2
VM-n
REST
Ansible
NetO-App Python Application
BGP
SSH
Ansible
OpenFlow
Multivendor, Bare-Metal, and Whitebox OpenFlow Switches
Traditional, Multivendor Networking Devices

Figure 1: Network Architecture Design

3.2. Hypervisor and Containers

For this study, OpenStack was selected as the private-cloud OS because it provides a virtualized, cloud infrastructure that leverages abstraction, orchestration, and automation capabilities for the network infrastructure. In the network architecture created as a result of this research, we implement a specific project forked from the master OpenStack project, OpenStack Kolla. OpenStack Kolla provides Representational State Transfer (REST) APIs, greater programmability, increased resource management, network virtualization, visibility and real-time monitoring, as well as multi-tenancy support. OpenStack Kolla deploys the OpenStack services in containers; thus, reducing the underlying server storage and achieves rapid boot time of the services with auto-scaling functionality [15]. The auto-scaling functionality of OpenStack Kolla implements a service called Senlin which facilitates automatic VM redundancy because it can dynamically distribute dedicated compute resources due to failure, user-defined thresholds, and utilization.

Containers have proved to be advantageous over their VM counterpart because of their portability, highly responsive lifecycle management, orchestration, agility, and elasticity [11]. In this research, we have selected Docker containers to be used within VMs in the OpenStack Kolla environment. The services we deployed in Docker were the DHCP server, DNS server, and the SDN Floodlight controller.

3.3. SDN Controllers

SDN implementations are evolving, but there has been limited research on providing seamless interconnectivity between varied platforms [26]. There is a need to adopt a routing mechanism for inter-connectivity between VMs/containers and both SDN and traditional networking devices. To achieve this, our network architecture implements an OpenContrail SDN controller to provide networking service for the virtualized OpenStack Kolla environment, the Floodlight controller for the OpenFlow SDN, and the vendor routers for the traditional network. This is beneficial in this research because OpenContrail has an intuitive Python REST API for automation and utilizes BGP to connect both SDN and traditional networks; thus, serving as an optimal solution for bringing inter-platform connectivity between the OpenStack Kolla environment and both the traditional and SDN devices.

The Floodlight SDN controller was implemented in this research to control the SDN infrastructure via OpenFlow. Floodlight was selected as the SDN controller because it is well-tested and has well-defined APIs. The APIs of Floodlight provide a documented menu which allows researchers to create network control and management applications with relative ease [12]. This was critical in this research because of the need for abstraction and orchestration between platforms.

3.4. The NetO-App

The NetO-App developed from this research to manage and configure multi-platform environments in small business networks was built using the Python programming language. Python was selected because it has an object-oriented design that provides high-level, built-in data types, user-friendly data structures, and support libraries [14]. The NetO-App is comprised of two primary modules: the abstraction module and the implementation module (Fig. 2).

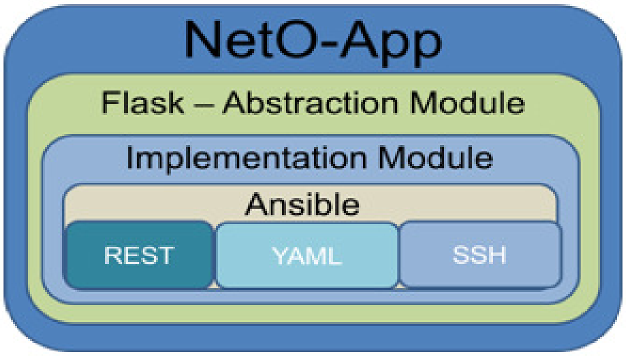


Figure 2: The Python NetO-App model

3.5. The abstraction Python module

The abstraction module provides a front-end interface for configuration, management, and network automation using the Python Flask web framework. This module provides a user-friendly GUI that abstracts the multi-platform network infrastructure to the user and provides the user with flexibility to select functions to orchestrate the entire network architecture.

3.6. The implementation Python module

The implementation module is comprised of the following sub-modules: Ansible, YAML, REST, and SSH. Each of the sub-modules is used to configure and manage various parts of the proposed network infrastructure environment.

To interact with network nodes for configuration, the implementation module of NetO-App uses the Python-based infrastructure automation framework Ansible. Ansible is a configuration and automation tool [18] that is prevalent in the industry for managing and configuring network devices and servers [19]. It was designed to make remote configurations quicker using SSH [18]. Ansible is agentless, so it does not require an agent to be installed on the client; instead, it uses SSH to push changes to the remote server or host defined in Playbooks. Playbooks describe hosts and tasks and are defined in YAML Ain't Markup Language (YAML) format [19]. Apart from SSH, Ansible can also communicate using APIs; thus, extending the number of network elements that can be configured using Ansible. For this research, Ansible is a tool that can be used to configure the cloud, SDN, and traditional networks.

4. RESULT AND ANALYSIS

In the experiments, we answered the sub-problems defined in section II to ultimately answer the primary research question.

4.1. Research sub-problem: Can we achieve orchestrated control to configure and manage multi-platform network elements using an application?

This research sub-problem guided the creation of NetO-App. The NetO-App combined with the proposed network infrastructure from this study provides a solution for centralized network management of cloud, SDN, and traditional networks that can dynamically push network configurations to the overlay and underlay network devices using SSH, Ansible, and REST. To make the solution user-friendly, we have developed a Flask front-end that abstracts the specific commands from the user. The user can make changes via the GUI, and then the back-end implementation module of NetO-App will execute the appropriate platform configuration scripts such as Ansible, REST, or SSH. This provides convenience to the user, without having to memorize multiple vendor specific commands, or understand a programming language. This abstraction layer is important to the research because small business network environments do not have the dedicated, skilled resources that other institutions have; thus, an intuitive web front-end is paramount to the success of the small business network model. The NetO-App communicates via REST and Ansible/SSH to dynamically configure and manage VMs within the OpenStack Kolla environment and uses Ansible/SSH to update or change configurations as described within Ansible Playbooks. When the user clicks on desired functions in the GUI, an appropriate Python script is executed, which invokes the Paramiko module inside of Ansible to SSH into either the VM, software-defined device, traditional device, or all. In the event of making changes to an existing configuration on any platform, the application checks for the present configuration and only pushes configurations that need to be updated; thus, providing efficiency and consistency across relevant nodes.

The NetO-App takes the user input based on mandatory parameters - hostname of node (on which VMs/containers would reside), tenant name, number of VMs/containers to deploy (preconfigured with desired packages), and the type of VMs/containers (Ryu controller/ONOS controller/ODL controller/Mininet/OVS). Optional parameters include – validation checks (verify VMs/containers are configured per the user-defined requirements), and fresh install (revert the

VMs/containers back to the clean state). Once the user has defined the parameters, NetO-App specifically uses Ansible/SSH to instantiate and configure VMs/containers per the requirement. The NetO-App is beneficial in small business networks because it makes it easier for the network user to orchestrate, manage, and configure a multi-platform environment consisting of cloud, SDN, and traditional networks from a centralized point without having to understand the underlying vendor-specific CLIs, programming language, or cloud operating systems. Furthermore, the NetO-App is free and open source software which appeals to small business networks' budgetary constraints.

4.2. Research sub-problem: Can we achieve a scalable and resilient control infrastructure for remote network management and configuration?

In our proposed model, we deployed multiple services, such as Floodlight SDN controllers within Docker containers on the OpenStack Kolla environment to provide a scalable and resilient control infrastructure to manage and configure the SDN. We could provide resiliency in the event of the failure of a VM because the Senlin service that is enabled within the OpenStack Kolla environment constantly monitors VMs for utilization. For example, when the threshold compute value of the containerized SDN controller is reached, the Senlin service spawns the secondary controller container and disables the primary controller, providing high availability of the control plane and optimizes resource efficiency. Furthermore, the OpenStack Kolla environment is hosted on multiple physical servers, which provides dynamic resilience and physical redundancy for the virtualized control platform.

To understand the amount of downtime achieved through this automated OpenStack scaling approach, we conducted a test that compared a manual configuration to an automated configuration. It was found that manually it took 57 seconds to create a secondary SDN controller whereas, Senlin could detect and configure a secondary SDN controller (container) in 1.2 seconds, and do it automatically without manual intervention. We incrementally added the configuration time based on the number of devices to understand the average number of configured containers in a stipulated time. Fig. 3 demonstrates that it took approximately 15 seconds to configure 10 controllers. This particularly demonstrates that the current OpenStack approach deployed in this research can reduce the downtime to seconds instead of minutes, and is fully automated requiring no manual intervention to self-heal.

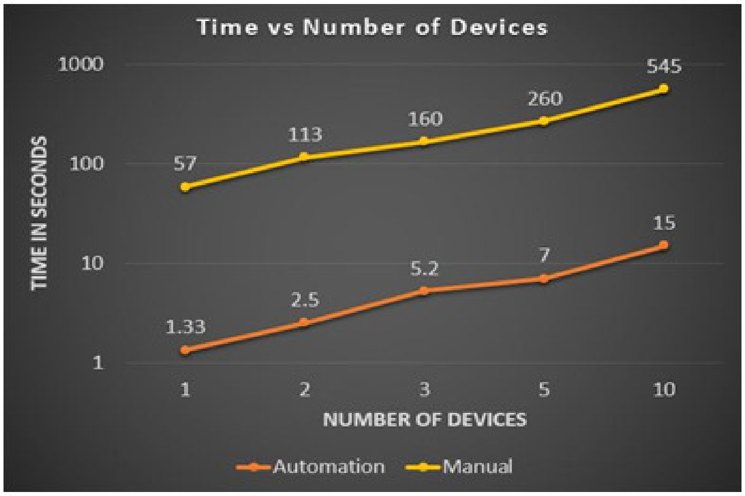


Figure 3: Time taken to configure SDN controller manually versus OpenStack automated approach

Therefore, to achieve a scalable and resilient control infrastructure for network management and configuration the research from this study answered the sub-problem by designing an architecture consisting of a virtualized cloud environment using OpenStack, Docker containers for required, lightweight service functionality, and the NetO-App which monitors and manages this environment to dynamically resolve physical and virtual failures and configuration quickly.

4.3. Research sub-problem: Can BGP be used to facilitate interconnectivity between VM/containers, SDN, and traditional networks?

To provide the virtual and physical L2 and L3 external connectivity, we used OpenContrail as a networking solution that works with the neutron service of OpenStack to provide cloud networking capabilities [24]. The OpenContrail architecture consists of two main components: vRouter and Controller. The OpenContrail controller uses Extensible Messaging and Presence Protocol (XMPP) to communicate with the vRouters and BGP/NETCONF to communicate with the traditional networking devices. As shown in Fig. 4, the control node present within the OpenContrail controller is responsible for processing routing information and applying them to the forwarding table of the vRouter service that handles networking for the OpenStack Kolla environment. To exchange routing information with the traditional vendor routers, the control node uses BGP. Thus, the use of BGP by the control node helps provide connectivity between VMs/containers and SDN devices with the traditional BGP speaking vendor routers.

OpenContrail provides REST APIs that are used by NetO-App to dynamically orchestrate the configuration of the VNFs inside of OpenStack. Therefore, the research answers this sub-problem by using OpenContrail as the VNF manager within OpenStack to communicate between the virtualized SDN and the traditional networking environment via BGP. Additionally, OpenContrail allows NetO-App to centralize control of this platform through the built-in OpenContrail REST API.

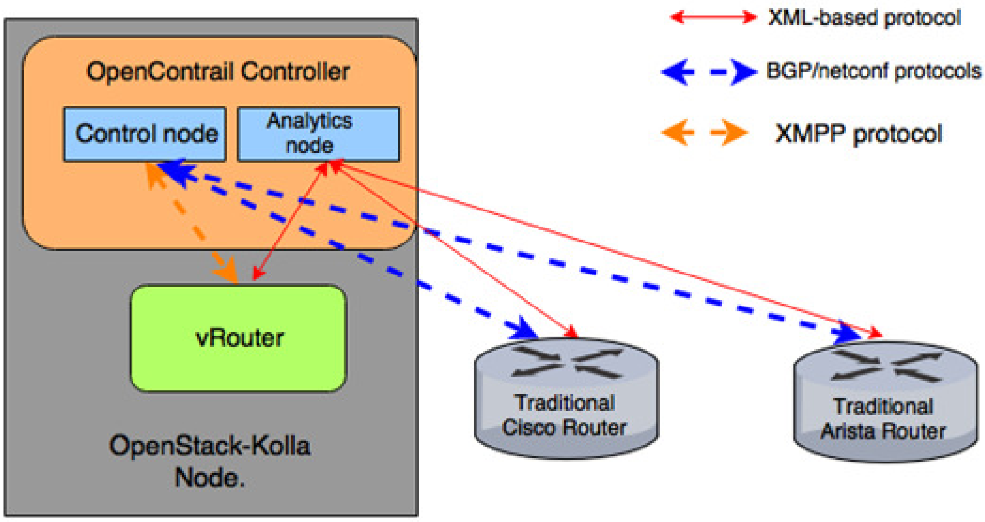


Figure 4: OpenStack-OpenContrail Networking Diagram

5. CONCLUSIONS

In this paper, we propose a network architecture and centralized network management application that can configure and manage multi-platform, software-defined, and traditional networking environments for small business networks. The small business network environment is subjected to limited software/hardware resources, repetitive deployments and configuration changes, dynamic adaptability to changing business requirements on a limited budget. It is essential for a small business organization to adopt an automated and orchestrated infrastructure that is efficient to configure and manage, inexpensive to deploy and operate, highly scalable, and provides ease of operation. We achieved this by designing an environment consisting of SDN, traditional networks, and cloud architectures. By utilizing OpenStack Kolla, we created a scalable and resilient control infrastructure for remote network management, configuration, scalability, and resiliency. OpenContrail facilitated interconnectivity between VMs, SDN, and traditional networks, through BGP and the NetO-App developed from this research implemented a user-friendly Flask front-end to abstract the underlying technologies from the user by utilizing Ansible, SSH, and REST to orchestrate the multi-platform cloud, SDN, and traditional network. Small business institutions can deploy the network architecture and NetO-App designed from this research to create a low-budget optimized network that is platform independent, centrally managed, resilient, scalable, easy to use, and inexpensive to implement.

Currently, this research focused on small business networks, but a future scope could enhance this by targeting other sectors with similar requirements, such as academia. Furthermore, NetO-App has minimal self-healing capabilities. NetO-App is able to execute predefined tasks upon a failure of a respective service. The future scope of this research can be improved by including a module for greater self-healing capabilities using TensorFlow (machine learning) for various failure scenarios [30]. This can be achieved by performing big data analysis and dissecting traffic parameters and performing actions to correct errors through TensorFlow. Specifically, Grafana [28] and Platform for Network Data Analytics (PNDA) [29] provide a platform for carrying out such analysis which serves as a future scope of the research.

REFERENCES

[1] R. Cziva et al, “SDN- based Virtual Machine Management for Cloud Data Centers” in 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet). IEEE, 2014, pp. 388 - 394.

[2] A. Risdianto et al, “Leveraging Open-Source Software for Federated Multisite SDN_Cloud Playground” in NetSoft Conference and Workshops, 2016 IEEE. IEEE, 2016, pp. 423 - 427.

[3] Software Defined Networking Definition [Online]. Available: https://www.opennetworking.org/sdn-definition/. [Accessed: Oct. 15, 2017].

[4] ONF [Online]. Available: https://www.opennetworking.org/projects/cord/

[5] S. Spadaro et al, “Orchestrated SDN based VDC Provisioning over Multi-Technology Optical Data Center Networks” in 2017 19th International Conference on Transparent Optical Networks. IEEE, 2017, pp. 1 - 4.

[6] L. Chen et al, “An SDN-Based Fabric For Flexible Data-Center Networks” in 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing. IEEE, 2015, pp. 121 - 126.

[7] S. Chen, and R. Hwang, “A scalable Integrated SDN and OpenStack Management System” in 2016 IEEE International Conference on Computer and Information Technology (CIT). IEEE, 2016, pp. 532 - 537

[8] W. Wang, W. He, and J. Su, “Boosting the Benefits Of Hybrid SDN” in 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2017, pp. 2165 - 2170.

[9] L. Wang et al, “Combining Neutron and OpenDaylight for Management and Networking” in 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). IEEE, 2017, pp. 457 - 462.

[10] S. Huang, and J. Griffioen, “Network Hypervisors: Managing the Emerging SDN Chaos” in 2013 22nd International Conference on Computer Communication and Networks (ICCCN). IEEE, 2013, pp. 1 - 7.

[11] H. Shimonishi, Y. Shinohara, and Y. Chiba, “Vitalizing data-center networks using OpenFlow” in 2013 IEEE Photonics Society Summer Topical Meeting Series. IEEE, 2013, pp. 250 - 251.

[12] O. Tkachova, M. Salim, and A. Yahya, “An Analysis of SDN-OpenStack Integration” in 2015 International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T). IEEE, 2015, pp. 60 -62.

[13] T. Huang, et. al., “A survey on Large-Scale Software Defined Networking (SDN) Testbeds: Approaches and Challenges” in IEEE Communications Survey & Tutorials. vol. 19 issue: 2, 2016, pp. 891 - 917.

[14] A. Rangola, “Applications of Python in Real World”[Online]. Available: https://www.invensis.net/blog/it/applications-of-python-in-real-world/?utm_source=invensis-blog&utm_campaign=blog-post&utm_medium=content-link&utm_term=benefits-of-python-over-other-programming-languages. [Accessed: Oct. 10, 2017].

[15] OpenStack Kolla [Online]. Available: https://wiki.openstack.org/wiki/Kolla

[16] OpenContrail Networking [Online]. Available: http://www.opencontrail.org/why-contrail-is-using-bgpmpls/

[17] T. Bakhshi, “State of the Art and Recent Research Advances in Software Defined Networking” in Wireless Communications and Mobile Computing. vol. 2017, 2017.

[18] P. Venezia, “Puppet vs. Ceph vs. Salt vs. Ansible” [Online]. Available: https://www.networkworld.com/article/2172097/virtualization/puppet-vs--chef-vs--ansible-vs--salt.html [Accessed: 15th Dec. 2016].

[19] Ansible [Online]. Available: [Accessed: 30th Jan., 2017]. https://www.ansible.com/blog/2016-community-year-in-review

[20] ONAP [Online]. Available: https://www.onap.org.

[21] L. Lwakatare, P. Kuvaja and M. Oivo, “An Exploratory study of DevOps extending the dimensions of DevOps with practices”, in The Eleventh International Conference on Software Engineering Advances.

[22] S. Sezer et al., "Are we ready for SDN? Implementation challenges for software - defined networks," in IEEE Communications Magazine, vol. 51, no. 7, pp. 36 - 43, July 2013.
doi: 10.1109/MCOM.2013.6553676.

[23] O. Salman, I. H. Elhajj, A. Kayssi and A. Chehab, "SDN controllers: A comparative study," in 2016 18th Mediterranean Electrotechnical Conference (MELECON), Lemesos, 2016, pp. 1 - 6.

[24] OpenContrail SDN [Online]. Available: http://www.opencontrail.org/the-importance-of-abstraction-the-concept-of-sdn-as-a-compiler/.

[25] E. Borjesson, and R. Feldt, “Automated System Testing using Visual GUI Testing Tools: A Comparative Study in Industry” in 2012 IEEE Fifth International Conference on Software Testing, Verification and Validation. IEEE, 2012, pp. 350 - 359.

[26] L. He, et. al., “Design and Implementation of SDN/IP hybrid space Information Network Prototype” in 2016 IEEE/CIC International Conference on Communications in China. IEEE, 2016, pp. 1- 6.

[27] Cisco Automation Benefits [Online]. Available: https://www.cisco.com/c/dam/en/us/products/collateral/cloud-systems-management/network-services-orchestrator/white-paper-c11-738289.pdf.

[28] Grafana [Online]. Available: https://grafana.com

[29] PNDA [Online]. Available: http://pnda.io

[30] TensorFlow [Online]. Available: https://www.tensorflow.org

[31] Small Business Network Basics, Cisco Systems. Available [online]: https://www.cisco.com/c/en/us/solutions/small-business/resource-center/connect-employees-offices/primer-networking.html.

[32] V. Ngugi, and C. Yoshida, "Digital Media Platform To Connect Small and Medium Enterprises In Nairobi" in 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). IEEE, 2016.

AUTHORS

Dewang Gedia, University of Colorado Boulder

Dewang Gedia is a Ph.D. student at the University of Colorado Boulder having primary research focus in Network Functions Virtualization and Software Defined Networks domain. He achieved his Master's degree from Interdisciplinary Telecom Program (ITP) at the University of Colorado Boulder in 2017.

Dr. Levi Perigo, University of Colorado Boulder

Dr. Perigo is a Scholar in Residence in Interdisciplinary Telecom Program (ITP) at the University of Colorado Boulder where he focuses on next generation networks, SDN/NFV, and network automation.

TRACKS MATCHING BASED ON AN ACOUSTIC SPECTRUM SIGNATURE DETECTION AND A MULTI-FRAME FUSION ALGORITHMS

Dahai Cheng[1], Huigang Xu[1] and Ruiliang Gong[2], Huan Huang[2]

[1]School of Electric and Automatic Engineering, Changshu Institute of Technology, Changshu, Jiangsu, China. 215500
[2]Changshu Ruite Electric Co. Ltd 2nd Changshu, Jiangsu, China. 215500

ABSTRACT

In this paper, an acoustic spectrum signature tracks matching algorithm based on the Manhattan distance and the Euclidean distance of signature vectors, and a multi-frame fusion algorithm are proposed for reliable real time detection and matching of boat generated acoustic signal spectrum signatures. The experiments results have shown that the proposed tracks matching algorithm has the ability to discriminate the tracks from different ships and the ability of matching of the tracks from the same ship; and the spectrum signature detection algorithm has captured the critical features of ship generated acoustic signals. In the process of signal spectrum signature detection, the observation of time and frequency space is structured by dividing input digitalized acoustic signal into multiple frames and each frame is transformed into the frequency domain by FFT. Then, a normalization of signal spectrum vector is carried out to make the detection process more robust. After that, an adaptive median Constant False Alarm Rate (AMCFAR) algorithm is used for the detection and extraction of boat generated spectrum signature, in which an extreme low constant false alarm rate is kept with relative high detection rate. Finally, the frame detections are accumulated to build up the track spectrum signatures.

KEYWORDS

Tracks Matching, Multi-Frame Fusion, Time-Frequency Observation Space, Spectrum Signature Detection

1. INTRODUCTION

Tracks matching plays an important role at detection and tracking of marine vehicles, including some underwater targets, such as submarine etc. The discrimination of targets is based on the comparison of the given spectra with the reference spectra available as endmembers in a spectral library. The comparison is done using the similarity as a criterion [1-4]. Stochastic measures such as spectral information divergence consider the spectral band to band variability as a result of uncertainty incurred by randomness. The spectrum can be modelled as a probability distribution so that the spectral properties can be further described by statistical moments of any order [1].

David C. Wyld et al. (Eds) : CSITY, DTMN, NWCOM, SIGPRO - 2018
pp. 73–92, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81107

The hybrid approaches of spectral angle mapper and spectral information divergence is found to increase the discriminatory power as against the individual measures [4].

Detection and extraction of underwater acoustic signal play an extremely important role in marine vehicle tracking and is one of the key technologies in underwater source detection, location, tracking, recognition and as well as acoustic communication [5–11]. Energy detection, feature detection and matched filter detection are commonly used range from high detection performance to low computation complexity, which can usually work well to some extent. Among these researches of the underwater passive detection methods, energy detection is mostly discussed and used in practical for its merits of lowest computational cost and easy to realize, while shows a poor performance under low SNR marine environment. However, over the past several decades, the marine environment has been more complexed in both natural and anthropogenic influences that ambient noise revealed. And as a consequence, the existing underwater passive detection methods are facing serious challenges, for which advanced detection scheme with better performance is still worthy investigating, especially under low signal-to-noise ratio (SNR) region [12-14].

This paper is focused on the research of theoretic algorithm development and experiments of tracks matching, and automatic detection ship spectrum signature based on boat generated acoustic signals from hydrophone ([15-18]). In this paper, an observation space is created by dividing input acoustic signals into multiple frames, with each frame sampled and transformed into the frequency domain. Then, an Adaptive Median Constant False Alarm Rate algorithm [19] is used for automatic target detection of boat-generated acoustic signals in each frame to provide a low constant false alarm rate with relatively high detection rate. Finally, the track signal spectrum signature is built up by a multi-frame detection vector fusion algorithm, in which the signal-to-noise ratio (SNR) in the observation period has been increased in the detection phase. The proposed algorithms have been tested on real boat generated acoustic signals obtained from a hydrophone.

2. Multi-Frame Spectrum Signature Detection in Time Frequency Domain

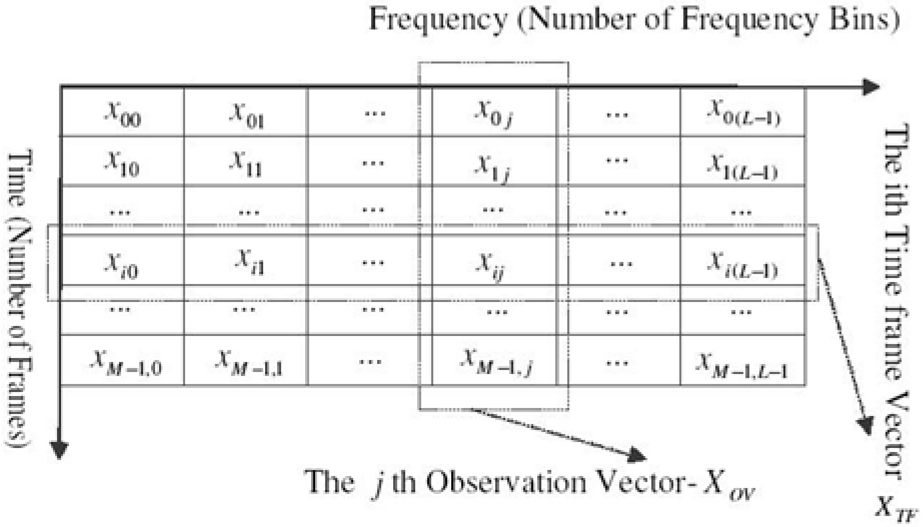
In this section, the observation is created first. Then, an adaptive signal spectrum components detection algorithm is used for the signature detection.

2.1. Multi-Frame Spectrum Observation Space

The observation space for multi-frame spectrum detection and fusion is created by dividing input time domain signal into multiple frames, and each frame in transformed into frequency domain by using FFT, which is shown in Fig. 1, in which x-direction represent number of frequency bins, and the y-direction represent time or number of frames.

2.2. Adaptive Spectrum Signature Detection Algorithm

An adaptive CFAR spectrum signature detection algorithm used in this paper are structured based on the idea of median CFAR thresholding, and the normalization operation based Neyman-Pearson (NP) criterion ([19-22]), which is widely adopted for signal detection application in either radar or sonar systems.


Frequency (Number of Frequency Bins)
Time (Number of Frames)
The ith Time frame Vector
X_{TF}
The j th Observation Vector- X_{OV}

x_{00}	x_{01}	...	x_{0j}	...	$x_{0(L-1)}$
x_{10}	x_{11}	...	x_{1j}	...	$x_{1(L-1)}$
...	...	...	...	...	...
x_{i0}	x_{i1}	...	x_{ij}	...	$x_{i(L-1)}$
...	...	...	...	...	...
$x_{M-1,0}$	$x_{M-1,1}$	...	$x_{M-1,j}$	...	$x_{M-1,L-1}$

Figure 1. Observation space for the acoustic signal spectrum

The whole multi-frame acoustic signal processing and detection algorithms employed in the paper is shown in Fig. 2. In which, the acoustic signals from hydrophone are sampled and converted into digital signals based on the Nyquist Sampling Theorem (Criterion) ([24-27]). In our experiments, the signals are originally sampled at 44.1kHz, then resampled to 2048 Hz, so the maximum frequency range in our scope is 1024 Hz. After that, the digital signal is divided into multiple frames with T=0.5 seconds for each frame, and the data processing period in digital format is N = 1024.

The digital signals are then transformed into the frequency domain by using FFT (Fast Fourier Transform). Since the data processing period is 1024, we chose the same length, i.e. 1024 points as FFT length.

Normally DC is the strongest component in the boat generated acoustic signals, which does not carry any useful information, but will affect the later processing, so it is necessary to remove DC component first.

We assume that the boat generated acoustic signals are stable random process during the observation period, such as 15 to 20 seconds.

In order to deal with various input signal strength and make the automatic target spectrum signature detection more robust, the normalization is performed in the frequency domain for the multi-frame spectrum fused vector by its magnitude. At this stage, each element in the multi-frame fusion vector is divided by the magnitude of the vector (geometric length). After normalization, a magnitude scaling factor of 40 dB (100 times) is used to give the signal a more practical range.

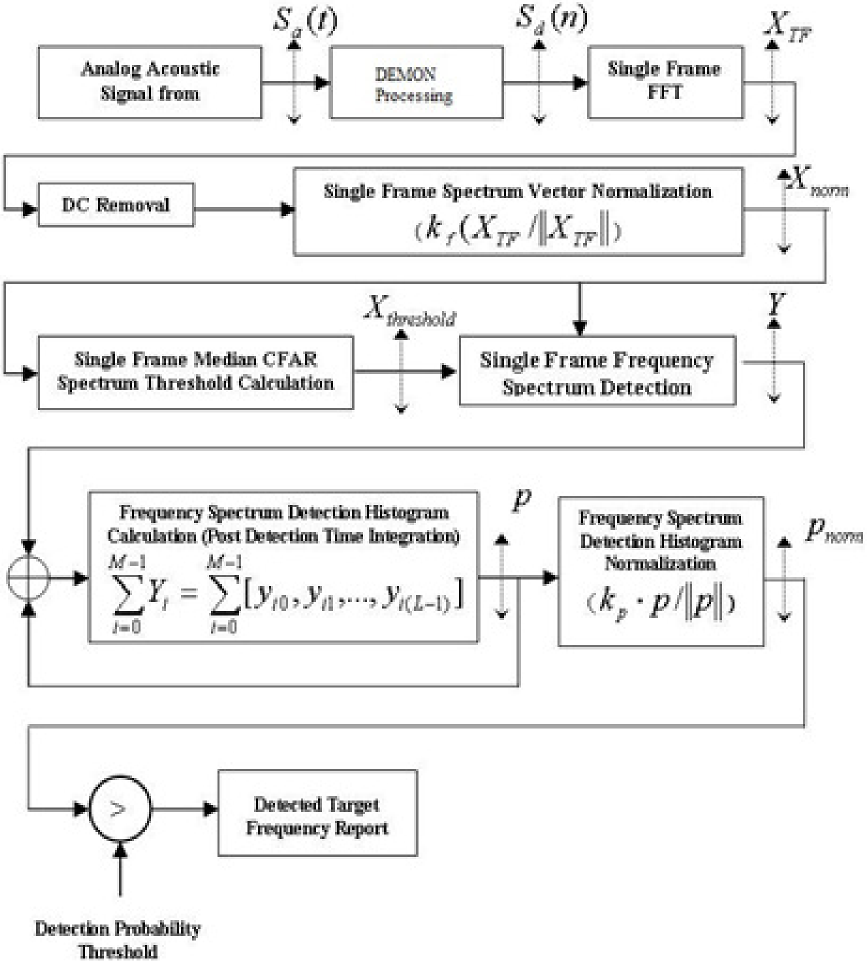

$S_a(t)$
$S_d(n)$
X_{TF}
Analog Acoustic Signal from
DEMON Processing
Single Frame FFT
DC Removal
Single Frame Spectrum Vector Normalization
$(k_f(X_{TF}/\|X_{TF}\|)$
X_{norm}
$X_{threshold}$
Y
Single Frame Median CFAR Spectrum Threshold Calculation
Single Frame Frequency Spectrum Detection
Frequency Spectrum Detection Histogram Calculation (Post Detection Time Integration)
$\sum_{i=0}^{M-1} Y_i = \sum_{i=0}^{M-1} [y_{i0}, y_{i1}, \ldots, y_{i(L-1)}]$
p
Frequency Spectrum Detection Histogram Normalization
$(k_p \cdot p/\|p\|)$
p_{norm}
>
Detected Target Frequency Report
Detection Probability Threshold

Figure 2. Adaptive Spectrum Signature Detection Multi-Frame Fusion Algorithms

The multi-frame spectrum fusion vector normalization is described as follows

$$X_{TF}^{norm} = k_f \cdot X_{TF} \Big/ \|X_{TF}\| = k_f \cdot \frac{\left[x_0, x_1, \cdots, x_{(L-1)}\right]^T}{\sqrt{x_0^2 + x_1^2 + \cdots + x_{(L-1)}^2}} \qquad (1)$$

in which L is the number of frequency bins in the fusion vector and kf is the scaling factor.

In traditional sonar systems, signal detection is realized by a constant thresholding. Instead, in our approach, a floating threshold vector for the multi-frame fusion vector is calculated based on an Adaptive Median CFAR algorithm, in which the threshold for each frequency bin is adapted by the median value over a sliding window. This is performed in two steps: in the first step, the median threshold of the normalized spectrum vector is subtracted from the original normalized spectrum vector. In the second step, the difference calculated in the first step is compared with a constant CFAR threshold (Δ), and if the difference is big enough to cross the constant threshold, the bin is reported as the target frequency, otherwise, it is reported as noise component. The parameter is called sensitivity, as the bigger the Δ is, the less sensitive our detection system is on weak signals. Since the CFAR threshold in each bin is adapted to its neighbourhood background noise, it will keep our automatic spectrum signature detection system at a very low and constant false alarm rate. In the following, the AMCFAR algorithm is described in details.

The Median Constant False Alarm Rate (Median CFAR) threshold vector,

$$\begin{aligned} X_{threshold} &= [x_{threshold,0}, x_{threshold,1}, \cdots, x_{threshold,L-1}]^T \\ &= [x_{threshold,i}]^T \end{aligned} \tag{2}$$

is calculated by feeding into a Median filter (whose properties and size will be discussed in the next section) as follows:

$$x_{threshold,i} = Median\{x_{norm,i-k}, \cdots, x_{norm,i}, \cdots, x_{norm,i+k}\} \tag{3}$$

The Median filter size is $(2k+1)$, with $k=1,2,3$. In order to deal with the boundary case, both the input signal spectrum vector and the threshold vector are treated as wrapped period signals. The multi-frame fusion vector detection is based on the comparison of the difference between vectors Y_{norm} and $Y_{norm}^{threshold}$ against a threshold, with the output being a binary vector:

$$Y_i = [y_{i0}, y_{i1}, \cdots, y_{i(L-1)}]^T \tag{4}$$

with Y_j ,where j=0, 1, …, L-1 given by:

$$y_{ij} = \begin{cases} 1, \left(x_{norm,ij} - x_{threshold,ij}\right) \geq \Delta \\ 0, otherwise \end{cases} \tag{5}$$

The y_{ij} values are either equal to 1 or 0, which are the detected frequencies based on the multi-frame fusion detection. As boat-generated signals typically last at least 20 to 30 minutes, it is reasonable to treat the signal stable random process in the observation period of 15 to 20 seconds, and it has been proven that the proposed algorithm can significantly improve the detection rate.

In order to increase the signal-to-noise ratio in frequency domain, a multi-frame detection vector fusion is used in the paper as shown in Equation 6.

$$\begin{aligned} p &= [p_0, p_1, \dots, p_{(L-1)}] \\ &= \sum_{i=0}^{M-1} Y_i = \sum_{i=0}^{M-1} [y_{i0}, y_{i1}, \dots, y_{i(L-1)}] \end{aligned} \tag{6}$$

Where M is the number of frames in the observation period, and L is the number of frequency bins in each frame.

3. SIGNATURE TRACKS MATCHING

The discrimination of targets is based on the comparison of the detected signature spectrum with the reference spectra stored in our spectral database, and the comparison is carried out by using the similarity or distance between spectrum signature vectors. The tracks matching algorithm is shown in Fig. 3, in which we can see that the final decision of whether the two signature tracks are from different ships or from the same ship is based on blocks matching and final fusion, and each block consisted of 15 frames, and each frame is a 1024 point spectrum vector.

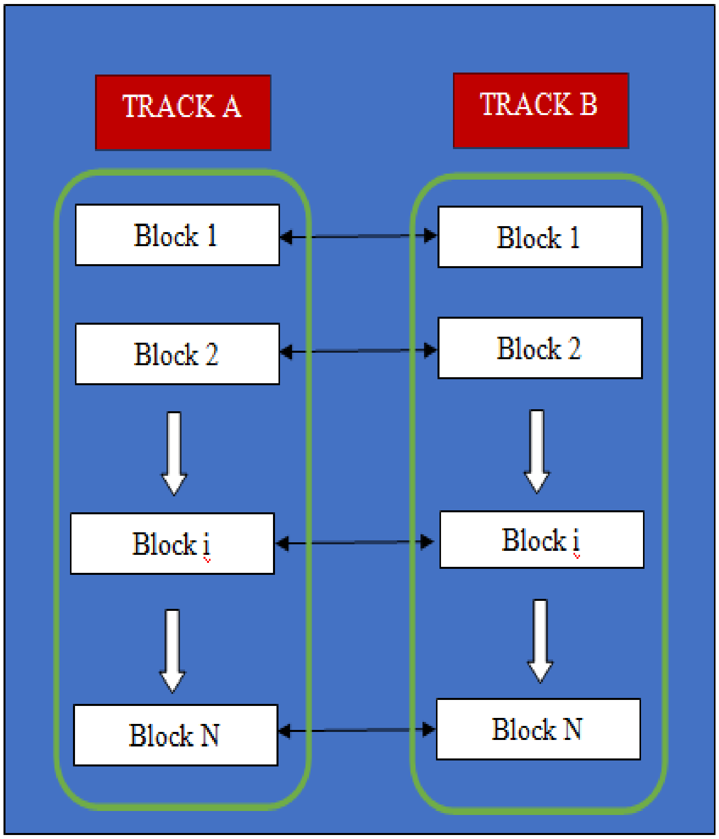


Figure 3. Tracks matching algorithm

The blocks matching is based on the similarity or distance measurements between two spectrum signature vectors, which include Manhattan and Euclidean distance distances in our proposed tracking algorithms.

Manhattan distance: Manhattan distance is also called taxicab distance. A taxicab geometry is a form of geometry in which the usual distance function or metric of Euclidean geometry is replaced by a new metric in which the distance between two points is the sum of the absolute differences of their Cartesian coordinates. The taxicab metric is also known as rectilinear distance, L1 distance, L1 distance or ℓ^1 norm (see L^p space), snake distance, city block distance, Manhattan distance or Manhattan length, with corresponding variations in the name of the geometry. The latter names allude to the grid layout of most streets on the island of Manhattan, which causes the shortest path a car could take between two intersections in the borough to have length equal to the intersections' distance in taxicab geometry.

Formal definition: The taxicab distance, d(Sig_1, Sig_2), between two vectors Sig_1 and Sig_2 in an N dimensional real vector space with fixed Cartesian coordinate system, is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes. More formally,

$$d(Sig_1, Sig_2) = \left|Sig_{1,1} - Sig_{2,1}\right| + \left|Sig_{1,2} - Sig_{2,2}\right| + \cdots + \left|Sig_{1,N} - Sig_{2,N}\right| = \sum_{i=1}^{N}\left|Sig_{1,i} - Sig_{2,i}\right| \tag{7}$$

where Sig_1 and Sig_2 are vectors Sig_1 = (Sig_{11}, Sig_{12},..., Sig_{1N}), and Sig_2 = (Sig_{21}, Sig_{22},..., Sig_{2N}).

For example, in the plane, the taxicab distance between Sig_1 and Sig_2 is $\sum_{i=1}^{N}\left|Sig_{1,i} - Sig_{2,i}\right|$.

Euclidean distance: In similarity measurement, we also use the Euclidean distance or Euclidean metric, which is the "ordinary" straight-line distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. The associated norm is called the Euclidean norm, or a generalized term for the Euclidean norm is the L^2 norm or L^2 distance.

The **Euclidean distance** between signature points Sig1 and Sig2 is the length of the line segment connecting them .In Cartesian coordinates, if the two spectrum signatures Sig_1 = (Sig_{11}, Sig_{12},..., Sig_{1N}) and Sig_2 = (Sig_{21}, Sig_{22},..., Sig_{2N}) are two points in Euclidean *N*-space, then the distance (d) from Sig_1 to Sig_2, or from Sig_1 to Sig_2 is given by the Pythagorean formula

$$d(Sig_1, Sig_2) = \sqrt{(Sig_{1,1} - Sig_{2,1})^2 + (Sig_{1,2} - Sig_{2,2})^2 + \cdots + (Sig_{1,N} - Sig_{2,N})^2} = \sqrt{\sum_{i=1}^{N}(Sig_{1,i} - Sig_{2,i})^2} \tag{8}$$

Stochastic measures such as spectral information divergence consider the spectral band to band variability as a result of uncertainty incurred by randomness. The spectrum can be modelled as a probability distribution so that the spectral properties can be further described by statistical moments of any order. The hybrid approaches of spectral angle mapper and spectral information divergence is found to increase the discriminatory power as against the individual measures.

4. Related Works and Discussion

In this paper, a tracks matching algorithm based on Manhattan distance and Euclidean distance, a multi-frame detection fusion and an adaptive signal spectrum detection algorithms are proposed, in which the tracks from different ships can be easily discriminated by both Manhattan distance and Euclidean distance; the spectrum signature detection algorithm captured the critical features, with a low false alarm and relative high detection rates in the frequency domain; and the signal-to-noise ratio in each frequency bin is increased by a multi-frame fusion algorithm. The input acoustic signal is sampled and divided into multiple frames. Then each frame is transformed into the frequency domain by using FFT. After that, the target generated frequency spectrum will be detected based on accumulated and normalized spectrum vector by the proposed adaptive algorithm, in which the basic idea of the detection algorithm is that of using, for each frequency bin, different, adaptive CFAR (Constant False Alarm Rate) thresholds [8] rather than a single, constant threshold (which is often the case in acoustic systems), which is described below. Then, a multi-frame fusion on is carried out by accumulating single frame detection result vectors in the observation period, for example, in 15 seconds, equivalent of 30 frames in our experiments. Finally, the tracks matching is carried out based on the given Manhattan distance and Euclidean distance measurements.

The basic idea of CFAR thresholding is that every single threshold of each frequency bin is computed based on the surrounding background noise. The higher the background noise, the higher the threshold is set. Moreover, our algorithm uses a median filter window centered at each frequency bin to adapt the threshold value. To the best of our knowledge, while this idea is often used in radar system to obtain lower false alarm rate with relatively higher target detection rate, it is applied here for the first time to sonar-generated acoustic signals.

Since the Median Filter is good at removing high frequency spot noise, it is a very effective way to calculate the threshold vector, which is independent of specific signals. As such, the Adaptive Median CFAR algorithm proves superior to other common approaches such as constant thresholds or average-based thresholds. The major advantage of the Median filter is in its ability to remove interferences such as strong signal or noise spikes without affecting the sharpness of edges (retaining sharp edges after filtering). Conversely, with an Averaging Low Pass Filter, which is equivalent to the Average CFAR algorithm, sharp edges will be blurred after filtering. Moreover, every bin in the averaging window will affect the threshold value, especially when the signal or a noise spike is strong. Evidence of the superiority of the median filter with respect to average filters for signal detection can also be found in [15].

The size of Median Filter window is an odd number, which can be $3,5,7,\cdots,(2k+1)$. From our experiments, a window size of 5 has been proved to be the most appropriate.

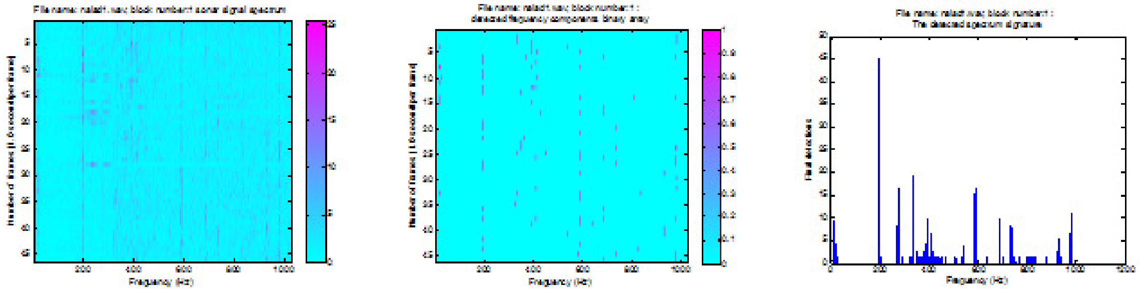
The combination of multi-frame spectrum fusion and adaptive CFAR detection makes the whole detection system extremely robust and reliable. As boat-generated signals have relatively long duration, we have used a multi-frame spectrum fusion, i.e. time-integration over multiple frames which significantly improves the SNR. Although this step introduces a delay in early detection of incoming boats in the order of 15 seconds, this is completely negligible with respect to the typical travelling speeds of monitored boats. The Adaptive Median Filter Constant False Alarm Rate (AMCFAR) algorithm is used to detect boat signature with relatively high detection rate while maintaining a low and constant false alarm rate.

5. Experimental Results of Tracks Matching and Adaptive Signature Detection Algorithm Based on The Multi-Frame Fusion

The test signals are provided by Soncom PTY LTD from “C-Buoy/Off-Buoy Processor Sea Trials’ at Low Islets, Australia (16.3833° S, 145.5667° E) on 17 June 2002. The proposed tracks matching, multi-frame spectrum fusion and adaptive signature detection algorithms have been successfully tested on several real ship generated signals in the following, which include the signals called “Naiad1” and “SF6” for reference.

5.1. SPECTRUM SIGNATURE DETECTION TEST

The proposed algorithms have been coded in Matlab, and the signature detection results have been shown in Figs. 4 and 5. Fig. 4 (a) shows us that the "Naiad1" boat signal has a pretty wide frequency band with a quite weak strength, which spreads between about 40 to 1k Hz, with the main frequency component at about 200 Hz and the strength of these frequency components between 0.5 to 4 dB. Fig. 4 (b) shows us the binary detections based on the Median CFAR threshold, in which the floating CFAR threshold vector has adapted to its original normalized multiple spectrum fusion vector to avoid any false detections. Fig. 4 (c) shows us the detected spectrum signature of "Naiad1" in the observation period with the normalized multiple frame spectrum vector, in which we can see that the spectrum signature has been reliably detected without any false detections.



(a) Naiad1 Boat Multi-Frame Signal Spectrum.

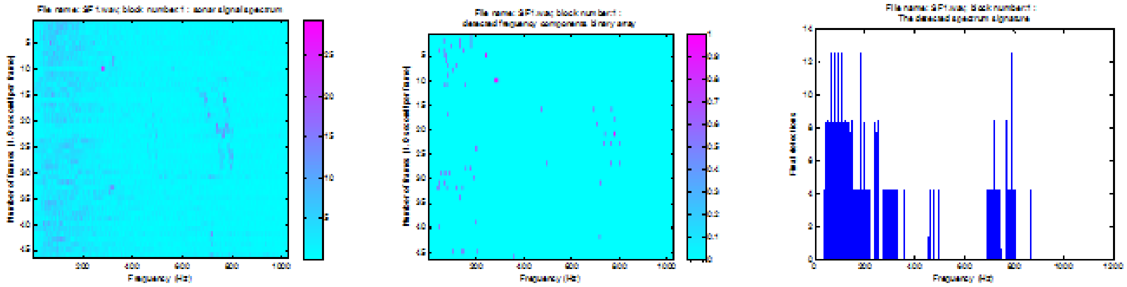
(b) Detected target frequencies by using the proposed Median CFAR algorithm with window size of 5

(c) The detected spectrum signature

Figure 4 Experimental results of "Naiad1" with the adaptive median CFAR algorithm with the sliding window size of 5, and proposed multiple frame fusion algorithms (block size: 45; initial frame number: 40).

5.2. "SF1" BOAT SIGNAL TEST

The proposed Adaptive Median CFAR detection and Multi-Frame spectrum fusion Algorithms have also been tested on "SF1'" boat signal, and the test results are shown in Fig. 5. Fig. 5 (a) shows us that the "SF1" boat signal has also a pretty wide frequency band, which spreads between about 50 to 380 Hz, and 700 to 900 Hz, with the main frequency component at about 200 Hz 750 Hz, and the strength of these frequency components between 5 to 25 dB. Fig. 5 (b) shows us the binary detections with the Median CFAR threshold, in which the floating CFAR threshold vector has adapted to its original normalized multiple spectrum. accumulated spectrum fusion vector, in which the SNR has been significantly improved, and it also shows us the normalized multiple spectrum fusion vector (blue) vs its fusion vector to avoid any false detections. Fig. 4 (c) shows us the detected spectrum signature of "SF6" in the observation period (red) with the normalized multiple frame spectrum fusion vector (green), in which we can see that the spectrum signature has been reliably detected without any false detections. The overall procedure is computationally light, thus allowing us cost-effective real-time implementation even on systems with limited computational power and size constraints such as on-board embedded computers.



(a) SF1 Boat Multi-Frame Signal Spectrum

(b) Detected target frequencies by using the proposed Median CFAR algorithm with window size of 5

(c) The detected spectrum signature.

Figure 5 Experimental results of "SF1" of the proposed multiple frame spectrum fusion and adaptive median CFAR algorithm with sliding window size of 5. (block size: 45; initial frame number 10)

5.3 TRACKS MATCHING BASED ON MANHATTAN DISTANCE- "NAIAID1" VERSUS "SF1" SIGNALS

The proposed tracks matching algorithm has been tested on signal "Naiad1" and "SF1". Each track has 9 blocks, and each block is consisted of 15 frames, and the spectrum signature of each block is calculated based on the single fame detection and fusion in the same block. The spectrum signatures of the two tracks are shown in Table 1 (a), (b) and (c).

Table 1(a). Detected block spectrum signatures (block size: 15 frames; number of blocks: 9).

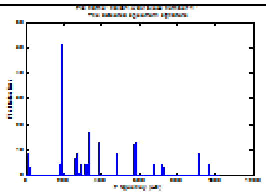
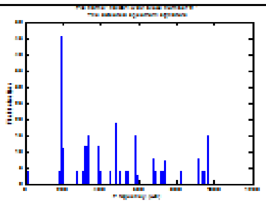
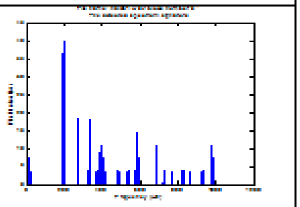
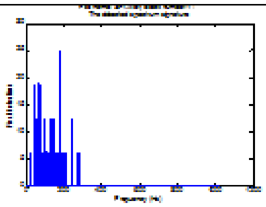
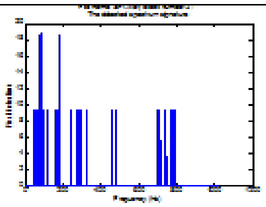
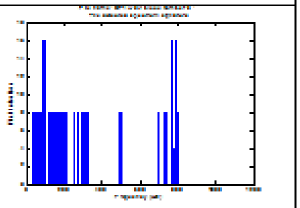
Block number:	1	2	3
Track A: Spectrum signatures			
Track B: Spectrum signatures			

Table 1(b). Detected block spectrum signatures (block size: 15 frames; number of blocks: 9).

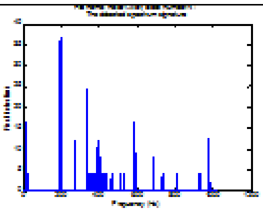
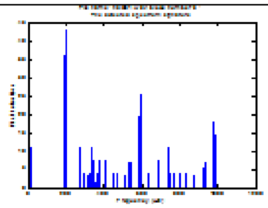
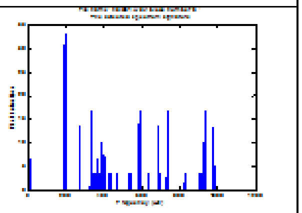
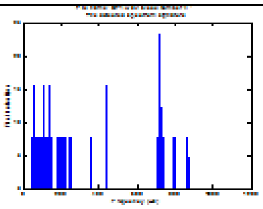
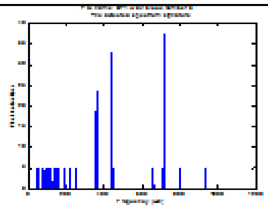
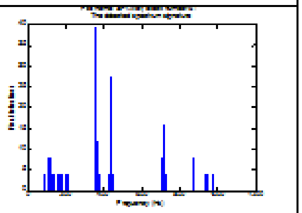
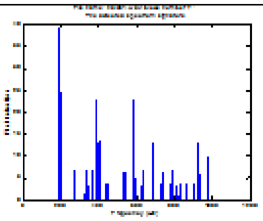
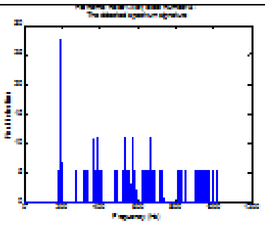
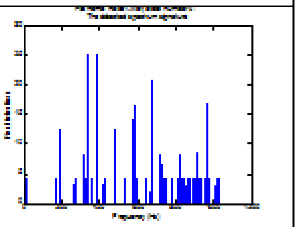
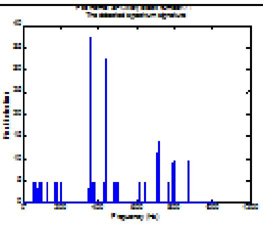
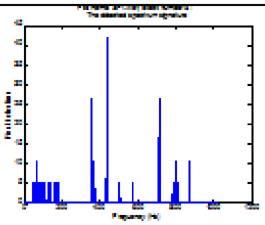
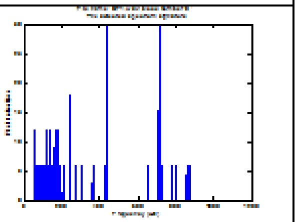
Block number	4	5	6
Track A: Spectrum signatures			
Track B: Spectrum signatures			

Table 1(c). Detected block spectrum signatures (block size: 15 frames; number of blocks: 9).

Block number	7	8	9
Track A: Spectrum signatures			
Track B: Spectrum signatures			

In Table 1, we can see that the spectrum signatures in two tracks of different ships are quite different, and the similarities of the spectrum signatures from the same ship are much more than that from the different ships, but there are also some variations in the same track.

The calculated Manhattan distances between the two tracks in Table 1 are shown in Table 2 and Fig. 6.

Table 2. The Manhattan distances between Track 1 (Naiad1) and Track 2 (SF1)

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Block 9
Naiad1/ SF1	1547.18	1685.19	1829.72	1549.87	1367.08	1585.52	1575.00	1964.62	1912.40

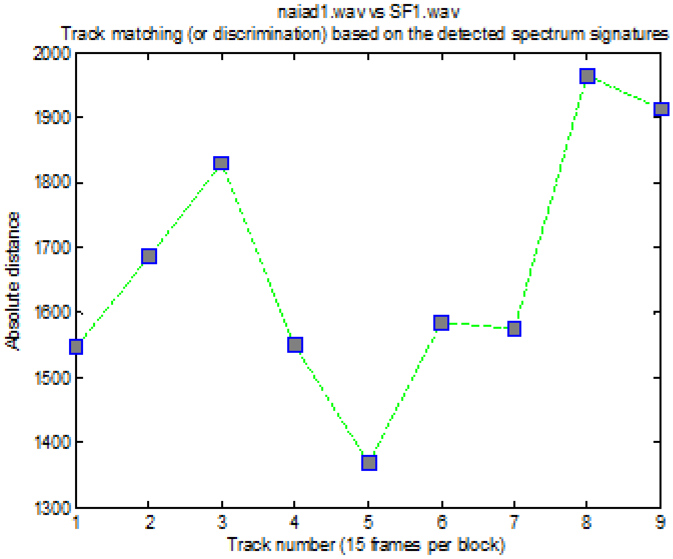


Figure 6. The Manhattan distances between Track 1 (Naiad1) and Track 2 (SF1)

Table 2 and Figure 6 show us that the Manhattan distances between Track 1 (Naiad1) and Track 2 (SF1) are around 1600.

5.4. TRACKS MATCHING BASED ON EUCLIDEAN DISTANCE- “NAIAID1” VERSUS “SF6” SIGNALS

The calculated Euclidean distances between the two tracks in Table 1 are shown in Table 3 and Fig. 7.

Table 3. The Euclidean distances between Track 1 (Naiad1) and Track 2 (SF1)

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Block 9
Naiad1/ SF6	139.11	141.42	140.58	140.51	140.45	139.93	139.90	138.52	140.71

Table 3 and Figure 7 show us that the Manhattan distances between Track 1 (Naiad1) and Track 2 (SF1) are around 140.

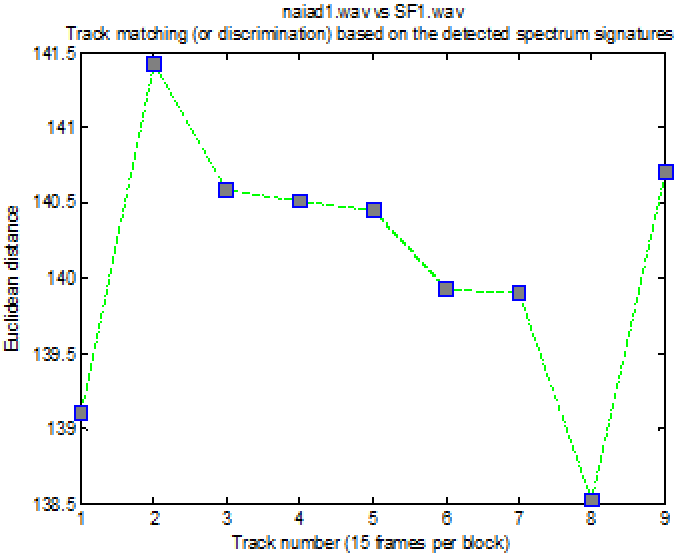


Figure 7. The Euclidean distances between Track 1 (Naiad1) and Track 2 (SF1)

5.5. TRACKS MATCHING BASED ON MANHATTAN DISTANCE- "NAIAID1" VERSUS "NAIAID1" SIGNALS

The matching of two tracks that are actually the same target ship is tested by using the data from the same track but shifted by about one block.

The calculated Manhattan distances between the two tracks (Naiad1 and Naiad1 by shifting around 1 block) are shown in Table 4 and Fig. 9.

Table 4. The Manhattan distances between Track 1 (Naiad1) and Track 2 (Naiad1)

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Block 9
Naiad1/ Naiad1	251.19	395.77	239.82	392.91	282.74	312.61	303.35	769.51	547.03

Table 4 and Figure 8 show us that the Manhattan distances between Track 1 (Naiad1) and Track 2 (Naiad1) with 1 block shift) are around 300.

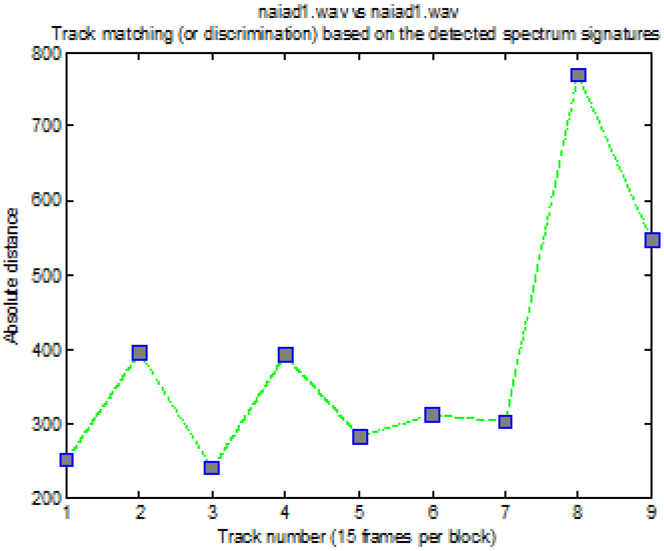


Figure 8. The Manhattan distances between Track 1 (Naiad1) and Track 2 (Naiad1) (with 1 block shift)

5.6. TRACKS MATCHING BASED ON EUCLIDEAN DISTANCE- "NAIAID1" VERSUS "NAIAID1" SIGNALS

The calculated Euclidean distances between the two tracks (Naiad1 and Naiad1 by shifting around 1 block) are shown in Table 4 and Fig. 9.

Table 5. The Euclidean distances between Track 1 (Naiad1) and Track 2 (Naiad1)

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Block 9
Naia d1/ Naia d1	34.86	48.96	33.02	46.88	35.65	39.72	38.04	81.39	51.71

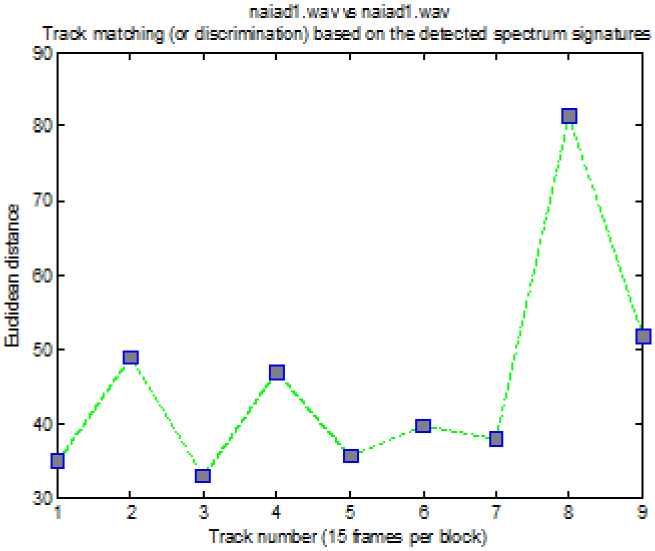


Figure 9. The Euclidean distances between Track 1 (Naiad1) and Track 2 (Naiad1)(with 1 block shift)

Table 5 and Figure 9 show us that the Euclidean distances between Track 1 (Naiad1) and Track 2 (Naiad1) with 1 block shift) are around 40.

5.7 TRACKS MATCHING (DISCRIMINATION) BASED ON MANHATTAN AND EUCLIDEAN DISTANCES

The power of discrimination (difference) or matching (similarity) of proposed algorithms are tested by plotting the Manhattan and Euclidean distances of tracks from different target tracks or from the same target tracks on the same space, drawing a line to separate them.

TRACKS MATCHING (DISCRIMINATION) BASED ON MANHATTAN DISTANCE:

The Manhattan distances between Track 1 (Naiad1) and Track 2 (Naiad1) (with 1 block shift) is shown in Figure 10 in red lines, and the Manhattan distances between Track 1 (Naiad1) and Track 2 (SF1) are shown in green lines.

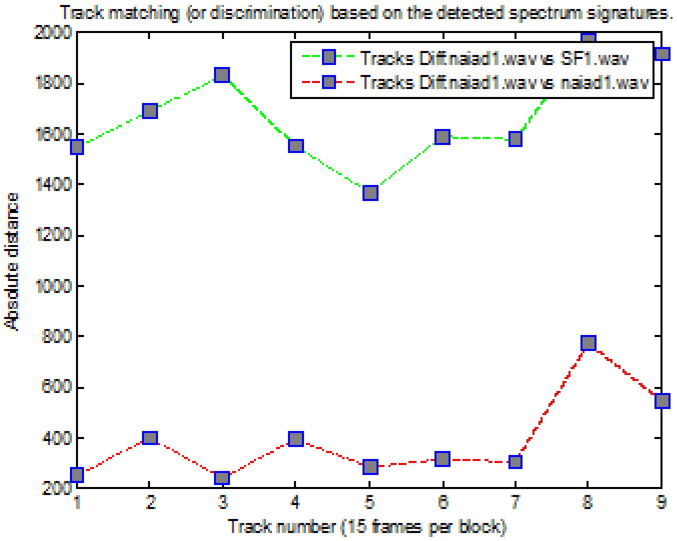


Figure 10. The Manhattan distances between Track 1 (Naiad1) and Track 2 (Naiad1) (with 1 block shift), and distances between Track 1 (Naiad1) and Track 2 (SF1).

Figure 10 shows us that the tracks distances of different target ships "Naiad1" versus "SF1" are around 1600, but the tracks distances of the same target ship "Naiad1" versus "Naiad1" are around 300. So, by drawing a line around 1000, it is very easy to separate the tracks of the same ship from the tracks from different ships.

TRACKS MATCHING (DISCRIMINATION) BASED ON EUCLIDEAN DISTANCE:

The Euclidean distances between Track 1 (Naiad1) and Track 2 (Naiad1) (with 1 block shift) is shown in Figure 11 in red lines, and the Euclidean distances between Track 1 (Naiad1) and Track 2 (SF1) are shown in green lines

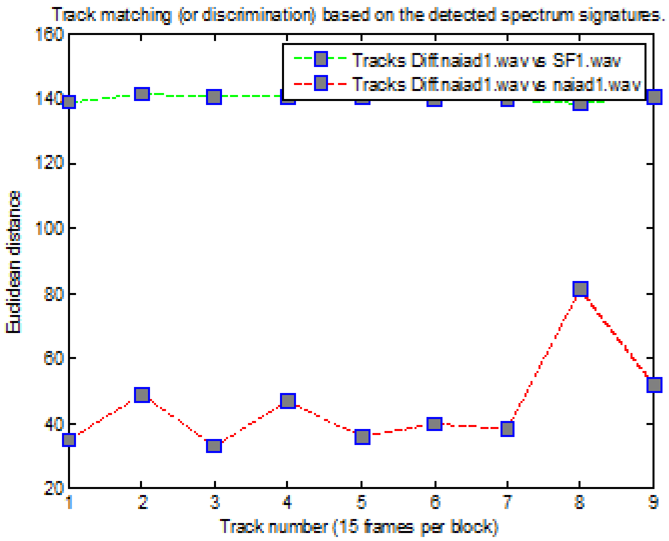


Figure 11. The Euclidean distances between Track 1 (Naiad1) and Track 2 (Naiad1) (with 1 block shift) , and distances between Track 1 (Naiad1) and Track 2 (SF1).

Figure 11 shows us that the tracks distances of different target ships "Naiad1" versus "SF1" are around 140, but the tracks distances of the same target ship "Naiad1" versus "Naiad1" are around 40. So, by drawing a line around 90, it is very easy to separate the tracks of the same ship from the tracks from different ships.

6. Conclusion and Future Work

In this paper, an acoustic spectrum signature tracks matching algorithm based on the Manhattan distance and the Euclidean distance of signature vectors, and a multi-frame fusion algorithm are proposed for reliable real time detection and matching of boat generated acoustic signal spectrum signatures. In which, an adaptive median constant false alarm rate algorithm has been used for effective spectrum signature detection of boat-generated acoustic signals, in which a low constant false alarm rate is kept with relatively high detection rate. The proposed algorithms have been tested on many real acoustic signals recorded from hydrophone at a site on the Australian coastline, two of them are shown in the paper. The statistical analysis and experimental results showed that the proposed algorithms have increased the SNR significantly in the observation period, and have kept a very low false alarm rate and relatively high detection rate for the whole detection system.

The following conclusions can also be drawn:

1) The experiments results have shown that the proposed tracks matching algorithm has the ability to discriminate the tracks from different ships and the ability of matching of the tracks from the same ship; and the spectrum signature detection algorithm has captured the critical features of ship generated acoustic signals.

2) The proposed multi-frame fusion algorithm has increased the SNR significantly, and made the whole detection more robust.

3) The proposed Adaptive Median CFAR algorithm is used to detect target frequency signature from a multiple frame fusion spectrum vector, keeping our automatic target detection system at low and constant false alarm rate. This algorithm has been proven that it is especially good for detecting LOFAR target frequency components.

4) A magnitude normalization (in the frequency domain) is used to keep our automatic detector more robust to noise and spurious frequencies.

5) With the default sensitivity value, most target frequency components are correctly detected. Further decreasing the sensitivity value makes the false detection rate (alarm rate) lower, but at the same time less target frequency components will be detected.

6) The boat-generated frequency spectrum signature can be detected with high accuracy. In the experiment reported in this paper, the detected boat-generated frequencies of ‘Naiad1’ and ‘SF1’ are very close to the “ground truth”.

Future work:

1) Recognition - The detected spectrum signatures can be used for ship recognition and tracking in the future, in which the study of similarity measures between ship spectrum signatures, and the neural network can also be possibly applied for the recognition of detected ships, based on a database of collected spectrum signatures.

2) Tracking-The proposed real time processing and detecting based sonars can be connected into a worldwide undersea network, in which the ships around each sonar can be detected, tracked, and displayed in a control center.

3) Arrays processing - Sonar array processing based on the proposed algorithm as a building block can also be used to increase the SNR of input signal, or detect the direction of incoming ships, [29-35]

ACKNOWLEDGEMENTS

The research is supported by 2017 Changshu Scientific and Technology Development Plan Project (CQ201701), and the test data was provided by Sonarcom PTY LTD in Australia.

REFERENCES

[1] Chang Cheini (2000), “An Information Theoretic Approach to Spectral Variability, Similarity, and Discrimination for Hyperspectral Image Analysis”, “IEEE Transactions on Information theory, Vol 46, No. 5, 2000, 19271932.

[2] Farifteh, J., Van Der Meer, F., Carranza, E.J.M. , 2006b, Similarity measures for spectral discrimination of saline soils, In Press: International Journal of Remote Sensing.

[3] Van Der Meer, F., 2006, The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery, International Journal of Applied Earth Observation and Geoinformation, 8, pp. 317.

[4] Du H, C.I. Chang, H. Ren, Cc. Chang, J. O. Jensen, And F. M. D¡¯Amico., 2004, New hyperspectral discrimination measure for spectral characterization, Optical Engineering, 43, no 8, pp. 1777¨C1786.

[5] J. X. Qiu, L. Zheng, Y. C. Wang, "Research on ship-radiated noise beat tune", Technical Acoustics, vol. 33, no. 4, pp. 322-325, 2014.

[6] F. S. Zhang, D. Feng, "Fusion and extraction of modulation feature from ship radiated-noise based on wavelet packet and ZFFT", Electronics World, vol. 14, no. 2, pp. 105-106, 2014.

[7] X. W. Luo, S. L. Fang, "Feature extraction from non-stationary amplitude modulated broad-band signal using the Hilbert-Huang transform", Signal Processing, vol. 27, no. 6, pp. 950-955, 2011.

[8] Xueyao Li ; Fuping Zhu; Harbin Eng. and Harbin Univ., Application of the zero-crossing rate, LOFAR spectrum and wavelet to the feature extraction of passive sonar signals, Proceedings of the 3rd World Congress on Intelligent Control and Automation, 2000., vol.4, pp. 2461-2463 (2000).

[9] G. Q. Wu, "Ship radiated-noise recognition (I) the overall framework analysis and extraction of line-spectrum", Acta Acoustica, vol. 23, no. 5, pp. 394-400, 1998.

[10] Y. S. Cheng, X. Gao, H. Liu, "A method for ship propeller blade-number recognition based on template matching", TechnicaI Acoustic, vol. 29, no. 2, pp. 228-231, 2010.

[11] Peng H.C., Long F., and Ding C., Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8, pp. 1226-1238 (2005).

[12] Roholing, H., “Radar CFAR Thresholding in Clutter and Multiple Target Situations,” IEEE Transactions on Aerospace and Electronic System, Vol. AES-19, No. 4, July 1983, pp. 608-621.

[13] S. Wang, J. X. Qiu, S. J. Wang, "Enhancement of ship radiated noise DEMON spectrum SNR based on correlation properties theory of principles of system_dynamics", Ship Science and Technology, vol. 35, no. 8, pp. 24-27, 2013.

[14] Y. S. Cheng, Y. C. Wang, "DEMON analysis of underwater target radiation noise based on modern signal processing", Technical Acoustics, vol. 25, no. 1, pp. 71-74, 2006.

[15] Chan, Y. T., “Underwater Acoustic Data Processing”, NATO ASI Series, Kluwer Academic Publishers.

[16] R. O. Nielsen, “Sonar Signal Analysis”. Boston, MA, Artech House, 1991, pp. 123-128.

[17] Merrill I, Skolnik, “Introduction to Radar Systems”. McGraw-Hill Book Company, 1980.

[18] Eric Dahai Cheng, Massimo Piccardi and Tony Jan, “Stochastic Boat-Generated Acoustic Target Signal Detection in Time-Frequency Domain”, IEEE International Symposium on Signal Processing and Information Technology (ISSPIT’04), Rome, Italy, December, 2004.

[19] Stergios Stergiopoulos, “Advanced Signal Processing Handbook”, CRC Press.

[20] Burdic, W. S., 1984, “Underwater Acoustic System Analysis”, Prentice-Hall, Englewood Clissf, NJ.

[21] Waite A. D., Sonar for practical Engineers, John Wiley, 3rd Ed. (2003).

[22] H. L.Van Trees, “Detection, Estimation and Modulation Theory”. Part I. New York, Wily, 1968, pp. 68-85.

[23] Soares Filho, W.; Manoel de Seixas, J.; Pereira Caloba, L., Principal component analysis for classifying passive sonar signals. International Symposium on Circuits and Systems, Sydney, Australia, vol. 2, pp. 592-595 May (2001).

[24] Soares Filho W., Seixas J. M. and de Moura N. N., Preprocessing passive sonar signals for neural classification. IET Radar, Sonar & Navigation, vol. 6, pp. 1-14 (2011).

[25] 1. R. J. Urick, Principles of Underwater Sound (3rd Edition), Washington, USA: McGraw-Hill, 1983.

[26] R. O. Nielsen, Sonar Signal Processing, MA: Artech House Inc., 1991.

[27] J. C. D. Martino, J. P. Haton and A. Laporte, "Lofargram line tracking by multistage decision," in Speech, and Signal Processing, USA, 1993. Lee, S.hyun. & Kim Mi Na, (2008) “This is my paper”, ABC Transactions on ECE, Vol. 10, No. 5, pp120-122.

[28] Johnson, D. H.; Dudgeon, D. E. (1993). Array Signal Processing. Prentice Hall.

[29] Van Trees, H. L. (2002). Optimum Array Processing. New York: Wiley.

[30] Krim, H.; Viberg, M. (July 1996). "Two Decades of Array Signal Processing Research" (PDF). IEEE Signal Processing Magazine: 67–94. Retrieved 8 December 2010.

[31] S. Haykin and K.J.R. Liu (Editors), "Handbook on Array Processing and Sensor Networks", Adaptive and Learning Systems for Signal Processing, Communications, and Control Series, 2010.

[32] E. Tuncer and B. Friedlander (Editors), "Classical and Modern Direction-of-Arrival Estimation", Academic Press, 2010.

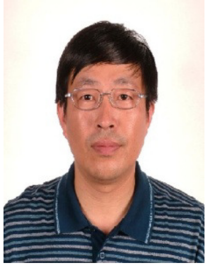
[33] Prof. J.W.R. Griffiths, Adaptive array processing, IEEPROC, Vol. 130,1983.

[34] N. Petrochilos,G. Galati, E. Piracci, Array processing of SSR signals in the multilateration context, a decade survey.

AUTHORS

Dahai Cheng (BE 1982, ME 1986, and PhD 1999) was born in 1961, in Xian, China; who is currently a professor at Changshu Institute of Technology. His research interests are radar and sonar signal processing, biological signal and image processing, computer vision, video surveillance and pattern recognition etc. He has about 40 academic papers published in recognized international journals and international conferences. He had been working on several ARC projects in NICTA, UTS, UNSW, and several R&D projects in Sonacom and MDA etc. He was also a specially invited professor at Xian University of Posts & Telecommunications.

E-mail: ericdahaicheng@qq.com



Huigang Xu was born in September 1969; who graduated from the Department of Applied Physics in East China Normal University (Bachelor degree), and obtained his PhD from Nanjing University of Science and Technology in control science and engineering in 2007, and currently he is a professor at Changshu Institute of Technology. He was a young leader of science and technology group in “333 high-level personnel training project”, and he was also selected as an “outstanding young teacher” in Jiangsu province, and a leader of key discipline project in the university. Professor Xu is a director of electrical and automation specialized committee in China Automation Society; and a director of electronic control system and device specialized committee, in China Electrotechnical Society; and also a director of Jiangsu automation society. Professor Xu is a master degree postgraduate supervisor of Suzhou University and China University of Mining and Technology, and is mainly engaged in the research work in signal detection and control, industrial automation, and has completed of 2 National Natural Science projects, and 2 provincial prospective research projects, and received a Jiangsu Provincial Science and Technology Progress Award, and a National Machinery Industry Science and Technology Progress Award.

E-mail: xuhuigang@cslg.edu.cn

Ruiliang Gong was born in November 1965, and currently the CEO of Changshu Ruite Electric Co. Ltd; finished his University Diploma from Changshu Institute of Technology in 1987. He joined Changshu Research Institute of Electronics in 1987, where he was engaged in project research and development, and jointly developed the "coal mine environmental monitoring system" with China University of Mining and Technology, which has passed the user assessment and identification. In 1989, He served as director of Changshu Research Institute of Electronics, and the director of the electronic instrument test workshop, and in charge of a R&D of "XXXH2 security monitoring alarm system", and won the "Jiangsu Province Science Technology Progress Award in 1992. Mr Ruiliang Gong was also in charge of the projects of “DH-UPS Navigation UPS” and “GLX marine auxiliary boiler control device” etc.

E-mail: gongrl@cs-ruite.com

Huan Huang was born in August, 1981; who is currently the chief engineer in Changshu Ruite Electric Co. Ltd. Mr Huang graduated from Department of Electrical Engineering, Tongji University (Bachelor) in 2004, and he obtained his master degree in automatic control from Tongji University in 2007. Mr Huang joined Changshu Ruite Electric Co. Ltd in 2008, first as the electronic engineer, then automation department manager, and deputy chief engineer. He has presided over the development of a set of electric field control devices, obtained five invention patents, and won the first prize of Suzhou civil military integration project, and currently involved in drafting a national standard for ships.

E-mail: 75189915@qq.com

AUTHOR INDEX