Phan Cong Vinh
Nguyen Ha Huy Cuong
Emil Vassev (Eds.)

217

Context-Aware Systems and Applications, and Nature of Computation and Communication

LNICST

6th International Conference, ICCASA 2017
and 3rd International Conference, ICTCC 2017
Tam Ky, Vietnam, November 23–24, 2017
Proceedings

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 217

Editorial Board

Ozgur Akan
Middle East Technical University, Ankara, Turkey
Paolo Bellavista
University of Bologna, Bologna, Italy
Jiannong Cao
Hong Kong Polytechnic University, Hong Kong, Hong Kong
Geoffrey Coulson
Lancaster University, Lancaster, UK
Falko Dressler
University of Erlangen, Erlangen, Germany
Domenico Ferrari
Università Cattolica Piacenza, Piacenza, Italy
Mario Gerla
UCLA, Los Angeles, USA
Hisashi Kobayashi
Princeton University, Princeton, USA
Sergio Palazzo
University of Catania, Catania, Italy
Sartaj Sahni
University of Florida, Florida, USA
Xuemin Sherman Shen
University of Waterloo, Waterloo, Canada
Mircea Stan
University of Virginia, Charlottesville, USA
Jia Xiaohua
City University of Hong Kong, Kowloon, Hong Kong
Albert Y. Zomaya
University of Sydney, Sydney, Australia

More information about this series at http://www.springer.com/series/8197

Phan Cong Vinh · Nguyen Ha Huy Cuong
Emil Vassev (Eds.)

Context-Aware Systems and Applications, and Nature of Computation and Communication

6th International Conference, ICCASA 2017
and 3rd International Conference, ICTCC 2017
Tam Ky, Vietnam, November 23–24, 2017
Proceedings

Springer

Editors
Phan Cong Vinh
Nguyen Tat Thanh University
Ho Chi Minh City
Vietnam

Nguyen Ha Huy Cuong
Quang Nam University
Tam Ky City
Vietnam

Emil Vassev
University of Limerick
Limerick
Ireland

ISSN 1867-8211 ISSN 1867-822X (electronic)
Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering
ISBN 978-3-319-77817-4 ISBN 978-3-319-77818-1 (eBook)
https://doi.org/10.1007/978-3-319-77818-1

Library of Congress Control Number: 2018937363

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

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

Preface

ICCASA and ICTCC 2017 are international scientific conferences for research in the field of intelligent computing and communication and were held during November 23–24 2017 in Tam Ky City, Vietnam. The aim of the conferences is to provide an internationally respected forum for scientific research in the technologies and applications of intelligent computing and communication. These conferences provide an excellent opportunity for researchers to discuss modern approaches and techniques for intelligent computing systems and their applications. The proceedings of ICCASA and ICTCC 2017 are published by Springer in the *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* (LNICST) series (indexed by DBLP, EI, Google Scholar, Scopus, Thomson ISI).

For this sixth edition of ICCASA and third edition of ICTCC, repeating the success of previous years, the Program Committee received submissions by authors from nine countries and each paper was reviewed by at least three expert reviewers. We chose 22 papers after intensive discussions held among the Program Committee members. We appreciate the excellent reviews and lively discussions of the Program Committee members and external reviewers in the review process. This year we chose four prominent invited speakers, Prof. Phayung Meesad from King Mongkut's University of Technology North Bangkok in Thailand, Prof. Mohamed E. Fayad from San Jose State University in USA, Prof. Akhilesh K. Sharma from Manipal University in India, and Prof. Vijender K. Solanki from CMR Institute of Technology in India.

ICCASA and ICTCC 2017 were jointly organized by The European Alliance for Innovation (EAI), Quang Nam University (QNU), and Nguyen Tat Thanh University (NTTU). These conferences could not have been organized without the strong support from the staff members of the three organizations. We would especially like to thank Prof. Imrich Chlamtac (University of Trento and Create-NET), Daniel Miske (EAI), and Ivana Allen (EAI) for their great help in organizing the conferences. We also appreciate the gentle guidance and help from Prof. Nguyen Manh Hung, chairman and rector of NTTU, and Dr. Huynh Trong Duong, rector of QNU.

November 2017

Phan Cong Vinh
Nguyen Ha Huy Cuong
Emil Vassev

Organization

Steering Committee

Imrich Chlamtac (Chair)	CREATE-NET, Italy
Phan Cong Vinh	Nguyen Tat Thanh University, Vietnam
Thanos Vasilakos	Kuwait University

Honorary General Chairs

Huynh Trong Duong	Quang Nam University, Vietnam
Nguyen Manh Hung	Nguyen Tat Thanh University, Vietnam

General Chair

Phan Cong Vinh	Nguyen Tat Thanh University, Vietnam

Technical Program Committee Chairs

Le Tuan Anh	Thu Dau Mot University, Vietnam
Vangalur Alagar	Concordia University, Canada
Nguyen Ngoc Tu	Missouri University of Science and Technology, USA

Technical Program Committee Track Leader

Nguyen Kim Quoc	Nguyen Tat Thanh University, Vietnam

Publications Committee Chair

Phan Cong Vinh	Nguyen Tat Thanh University, Vietnam

Marketing and Publicity Committee Chair

Do Nguyen Anh Thu	Nguyen Tat Thanh University, Vietnam

Workshops Committee Chair

Emil Vassev	University of Limerick, Ireland

Patron Sponsorship and Exhibits Committee Chair

Nguyen Ho Minh Duc	Nguyen Tat Thanh University, Vietnam

Panels and Keynotes Committee Chair

Nguyen Ha Huy Cuong — Quang Nam University, Vietnam

Demos and Tutorials Committee Chair

Abdur Rakib — The University of Nottingham, Malaysia Campus, Malaysia

Posters Committee Chair

Vo Thi Hoa — Quang Nam University, Vietnam

Industry Forum Committee Chair

Waralak V. Siricharoen — Burapha University, Thailand

Special Sessions Committee Chair

Vangalur Alagar — Concordia University, Canada

Local Arrangements Committee Chair

Vu Thi Phuong Anh — Quang Nam University, Vietnam

Website Committee Chair

Thai Thi Thanh Thao — Nguyen Tat Thanh University, Vietnam

Technical Program Committee

Rasha Shaker Abdulwahab	College of Applied Sciences, Oman
Govardhan Aliseri	Jawaharlal Nehru Technological University Hyderabad, India
Le Hong Anh	Ha Noi University of Mining and Geology, Vietnam
Krishna Asawa	Jaypee Institute of Information Technology, India
Muhammad Athar Javed Sethi	University of Engineering and Technology (UET) Peshawar, Pakistan
Prasanalakshmi Balaji	Professional Group of Institutions, India
Shanmugam BalaMurugan	Kalaignar Karunanidhi Institute of Technology, India
Chintan Bhatt	Charotar University of Science and Technology, India

Aniruddha Bhattacharjya	Narasaraopeta Engineering College, India
Nguyen Phu Binh	Institute for High Performance Computing, A*STAR Singapore
Nguyen Thanh Binh	Ho Chi Minh City University of Technology - HCMVNU, Vietnam
Giacomo Cabri	University of Modena and Reggio Emilia, Italy
Jamus Collier	University of Bremen, Germany
Nguyen Ha Huy Cuong	Quang Nam University, Vietnam
Nguyen Hung Cuong	Hung Vuong University in Phu Tho Province, Vietnam
Issam Damaj	The American University of Kuwait, Kuwait
Shahed Mohammadi Dehnavi	Ragheb Isfahani Higher Education Institute, Iran
Charu Gandhi	Jaypee Institute of Information Technology, India
Kurt Geihs	University of Kassel, Germany
Hafiz Mahfooz Ul Haque	The University of Lahore, Pakistan
Huynh Xuan Hiep	Can Tho University, Vietnam
Huynh Trung Hieu	Ho Chi Minh City University of Industry, Vietnam
Phan Ngoc Hoang	Ba Ria-Vung Tau University, Vietnam
Zhu Huibiao	East China Normal University, China
Dao Huu Hung	FPT Japan Co. Ltd, Japan
Chia-Hung Hung	Missouri University of Science and Technology, USA
Nguyen Quoc Huy	Saigon University, Vietnam
Muhammad Fahad Khan	Fedral Urdu University of Arts, Science and Technology, Pakistan
Ashish Khare	University of Allahabad, India
Asad Masood Khattak	Zayed University, UAE
Ondrej Krejcar	University of Hradec Kralove, Czech Republic
Xuan Khoa Le	Ulster University
WenBin Li	Missouri University of Science and Technology, USA
Nguyen Phuoc Loc	Sunflower Soft Co., Vietnam
Manmeet Mahinderjit Singh	Universiti Sains Malaysia, Malaysia
Moeiz Miraoui	University of Quebec, Canada
Rohin Mittal	State University of New York at Buffalo, USA
Amol Patwardhan	Louisiana State University, USA
Gabrielle Peko	The University of Auckland, New Zealand
Nguyen Thanh Phuong	Polytechnic University of Bari, Italy
Nguyen Kim Quoc	Nguyen Tat Thanh University, Vietnam
Abdur Rakib	The University of the West of England, UK
Sreekanth Rallapalli	Botho University, Botswana
Ognjen Rudovic	Imperial College London, UK
S. Satyanarayana	KL University, India
Chernyi Sergei	Admiral Makarov State University of Maritime and Inland Shipping, Russia

Manik Sharma	DAV University, India
François Siewe	De Montfort University, UK
Waralak V. Siricharoen	Burapha University, Thailand
Vijender Kumar Solanki	Institute of Technology and Science, Ghaziabad, India
Areerat Songsakulwattana	Rangsit University, Thailand
David Sundaram	The University of Auckland, New Zealand
Tran Huu Tam	University of Kassel, Germany
Prashant Vats	Career Point University, India
Chien-Chih Yu	National ChengChi University, Taiwan

Contents

Context-Aware Systems and Applications

Nature of Computation and Communication

Context-Aware Systems and Applications

A Resource-Aware Preference Model for Context-Aware System

Ijaz Uddin[1(✉)] and Abdur Rakib[2]

[1] School of Computer Science,
The University of Nottingham Malaysia Campus, Semenyih, Malaysia
khyx4iui@nottingham.edu.my
[2] Department of Computer Science and Creative Technologies,
The University of the West of England, Bristol, UK
Rakib.Abdur@uwe.ac.uk

Abstract. In mobile computing, context-awareness has recently emerged as an effective approach for building adaptive pervasive computing applications. Many of these applications exploit information about the context of use as well as incorporate personalisation mechanisms to achieve intended personalised system behaviour. Context-awareness and personalisation are important in the design of decision support and personal notification systems. However, personalisation of context-aware applications in resource-bounded devices are more challenging than that of the resource-rich desktop applications. In this paper, we enhance our previously developed approach to personalisation of resource-bounded context-aware applications using a derived context-based preference model.

Keywords: Context-aware · Preferences · Personalisation
Defeasible reasoning

1 Introduction

Context-awareness is one of the core features of ubiquitous computing. While the concept of context-awareness exists since early 1990s [1], it has gained fast popularity in the recent years due to the evolution of smartphones and the growth in the usage of Internet and sensor technology. Nowadays, almost all modern smartphones are equipped with visually rich and dynamic user interfaces, as well as a range of sensors including, accelerometers, GPS, Gyro, pulse and finger print sensor. The embedded sensors in the smartphones can be used to acquire contextual data from various context sources, e.g., users, environments or other devices. The low-level sensed contextual data can be translated into machine-readable data for higher level context inference using e.g., a suitable knowledge representation and reasoning technique. In the literature, the term *context* has been defined in various ways within the context-aware computing research, however, one of the most widely accepted definitions was provided by [2] as *context* is

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 3–13, 2018.
https://doi.org/10.1007/978-3-319-77818-1_1

any information that can be used to characterise the status of an entity. Common context types include the user-related context(e.g., profile, identity, activity, preference, location), physical or environment-related context (e.g., noise levels, temperature, wind speed, location, room number, time of day), and device-related context (e.g., resources, network connectivity, resolution). A system is said to be context-aware if it can adapt its behaviour to a given situation and provide relevant information and/or services to the users [1,2]. In the literature, various context modelling approaches and context-aware system development architectures have been proposed, however, ontology-based approach has been advocated as being the most promising one [3,4]. In our research, we model context-aware systems as ontology-driven multi-agent rule-based reasoning systems [5,6], where context is formally defined as *(subject, predicate, object)* triple that states a fact about the subject where—the subject is an entity in the environment, the object is a value or another entity, and the predicate is a relationship between the subject and object. That is, we model context as first order function free predicates, a context state corresponds to a belief state of an agent or content of its working memory, and firing of rules that infer new contexts may determine context changes and representing overall behaviour of the system [6]. In context-aware systems, user preferences play an important role in adapting their behaviour to satisfy the individual user in different contexts. The mechanism generally relies on implicit and/or explicit user, device, physical or environment-related context that manipulate working mechanism that control the way applications react to the context in use. For example, in our case only a subset of the rules of an agent's rule base could be active based on the given preferences. In this paper, we present and enhance our previously developed approach [7] to personalisation of context-aware applications using a derived context-based preference model. The main idea of our approach is that preferences are specified as derived or externally communicated/sensed context so that they can be easily controlled to personalise the system behaviour without modifying the internal settings or agent's program.

The rest of the paper is structured as follows. In Sect. 2, we briefly review closely related work. In Sect. 3, we discuss motivation for undertaking this study. In Sect. 4, we present the proposed context-aware preference model, which extends the existing framework [7] by incorporating derived-context based user preference. In Sect. 5, we discuss derived-context based user preference in more detail. In Sect. 6, we present a simple case study to illustrate the usefulness and effectiveness of the proposed approach, and conclude in Sect. 7.

2 Related Work

The use of preferences in context-aware systems for decision making and personalization has been a highly researched topic. For instance, incorporating preferences in context-aware applications, mainly in manipulating the context, storing, management and its use in the future has been a subject of interest to many researchers (see, e.g., [8–10]). Even the research in database technology has seen

the effect of personalised queries where the result of a query depends on the current context available [9]. However, these methods are used for developing resource-rich systems with large scale databases. Some more recent preference oriented works consider different approaches, e.g., in [11] authors use user profiling technique for storing contexts of different users. It matches all the rule instances with the facts stored in the working memory and the profile is loaded based on the current context. This approach perhaps requires extensive memory to run the system.

Similarly, context-aware recommendation applications are also part of user preferences, where an application is recommended to the user based on his past patterns. In [12], the authors have proposed a model for personalising recommendations and improving user experience by analysing the context in use. They have used ranking algorithms for context based items. The system integrates the social media to explore the user preferences and based on those preferences it personalises the user experiences.

As digital healthcare often designed to exploit recent advances in computing technology, traditional healthcare information systems make use of context-aware technologies to improve the quality of healthcare services. In [13], the authors proposed a context-aware system framework for automated assistance and independent living of senior citizens. It mainly focuses on the personalisation and adaption of preferences. Besides other tasks, a local context manager is used in order to process the data from low-level to high-level. The decision making module is the IDSS or intelligent decision support system, which is a cloud based service. This IDSS has in itself large number of reasoners such as *Lifestyle Reasoners and Management*, which works on different data types. The reasoner can store long-term data that have certain patterns or routines, which defines the lifestyle of some users. Thus it can detect changes and indicate changed behaviour of users in terms of their health status. In [14], the authors propose using defeasible logic rules to describe system behaviour and for modelling context-dependant preferences. Their work is closely related to our work presented in this paper. However, in our work we use defeasible reasoning to model and describe behaviour of the context-aware agents.

3 Motivation

The motivation for undertaking this study is that, the usage of social networks and cloud computing has dominated the context-aware platform by providing more resource-rich techniques on server/cloud. It is practically possible to scale a high end system with the use of resource-rich cloud computing. However, there is certainly attention required when systems are developed considering tiny resource-bounded devices. To add more, if a system is intended for elder care or patient care then the chances are that a patient might not have his social networking account or may not be using it actively. Development of a system which is independent of other services can be beneficial for rapid implementation of elder care or remote system where resources are limited. Further to this, our

previously developed externally received context-based preference mechanism [7] works on different indicators provided by the user to generate a preference set. However, there are some contexts which can not be obtained from external or embedded sensors, and a user might be interested in those contexts in order to generate the preference sets. For example, a context *Patient(Alan)*, the status of a person of being a patient can not be obtained from a sensor, instead it has to be derived using some rules. Based on the status of a person being a patient, the system can generate a preference set accordingly. Similarly, derived context based approach could be useful for generating a preference set when the context that was actually expected from an external source cannot be obtained perhaps due to a sensor malfunction. For example, if the contextual information of user's presence in his office cannot be received from the GPS, an agent may derive it using a set of rules and information obtained from a occupancy sensor. One such example can be found in the work by [15], which mainly deals with the survivor tracking at the current stage but can be evolved further to be used in elder care or patient care system. In light of the above literature, we propose a preference model suitable for implementing context-aware systems that run on resource-bounded devices. Furthermore, the preferences in our model are filtered through two different layers, one is generalised preference that deals with a particular context, e.g., preference required at office or home [7], second is when a conflict occurs between the rules of the preference set [14]. By incorporating these two different preference layers, we propose an approach aimed at providing preferences to the users with minimal usage of system resources and independent of any other services.

4 Resource-Aware Preference Model

The logical framework and its extension to accommodate preferences presented in [6,7] serve as the basis of the whole framework. In this paper, we extend our previous work [7] to incorporate preferences using a derived context-based preference model, while maintaining the resource utilisation factor intact [16]. Note that our approach to preferences is based on two levels. First level works on the basis of communicated/sensed or derived context, while second level assigns priorities to different rules to give preference to one rule over another to resolve conflicts. In [7], the preferences were based on the user provided or externally communicated/sensed contexts. However, the implicitly derived contexts were not considered to make changes to the preference sets. Here, we consider the derived contexts to be dealt as input in case if they are indicated to be the contexts of interest by the user. The structure of inference engine and internal set-up remain the same. However, some changes are made within the preference manger layer of the system architecture and to the point when new contexts are derived.

4.1 Context-Aware System Architecture

As mentioned before, we design context-aware systems as multi-agent rule-based reasoning agents. In general, there are several different ways agents in a multi-agent system can be programmed. In our case, programming agent behaviour

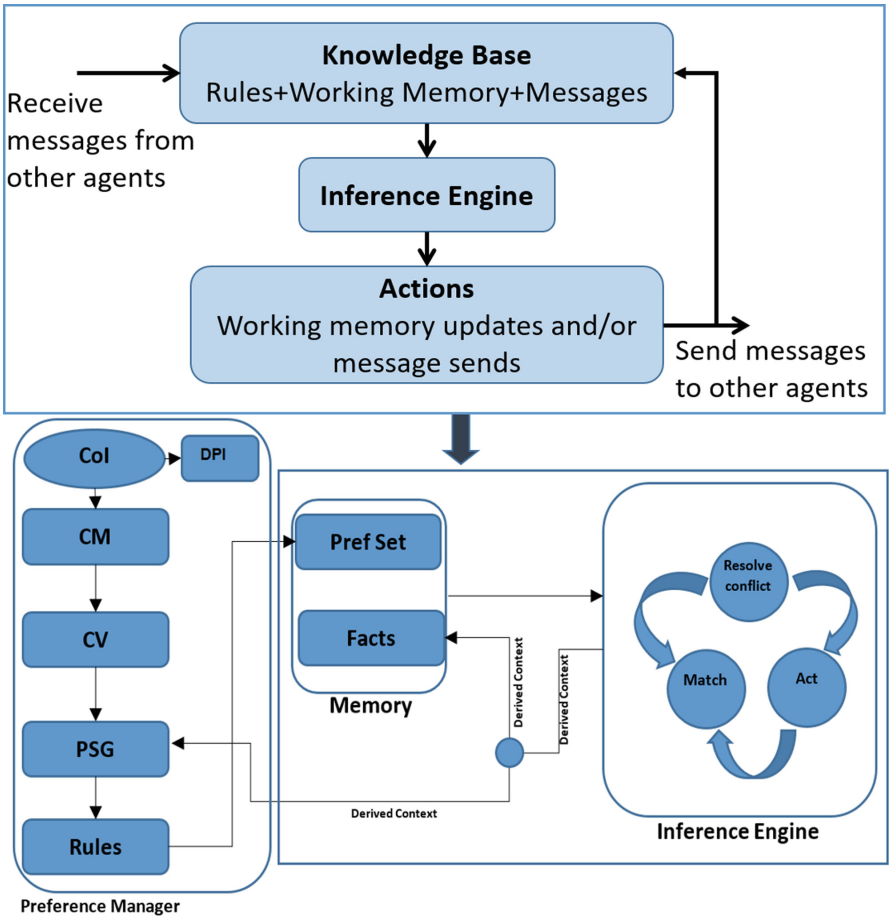


Fig. 1. System architecture and preference generation overview

using a declarative rule language consists in building a layered architecture using the Horn clause rules at the upper layer and Android Java is used in the lower layer to handle agent communication. The knowledge base is the upper layer of the architecture, which contains annotated ontology-driven rules (translated from OWL2 RL ontology augmented with SWRL rules). The upper part of the Fig. 1 represents the layered architecture of our system. A formal specification of the rule syntax is given in the following section.

4.2 Rule Structure

A typical rule format of our framework can be found in [7], while some changes are made when derived-context based preference is intended. The typical structure of a rule looks like: $m : P_1, P_2, \ldots, P_n \rightarrow P_0 \ : F : CS$ where $n \geq 0$.

The $CS(= \{-||P||\_P||tag\})$ is mainly used for the preference set generation. The different CS indicators are used by the framework to determine the nature of preferences required by the user. In case when we do not wish to attach a rule to any of the preference set then we can simply use it as a general rule that can be indicated by the "−" sign. That is, any rule with a "−" sign will be considered as a common rule and will be added to any preference set. The predicate P can be a context/fact, e.g., *hasLocation(Alan, UNMC)*. The predicate $\_P$ indicates that the rule attached to this format is only selected when P is derived by the inference mechanism. Thus, it is a potential context to be used as a preference only if an agent derives it by the inference mechanism. For example, *_hasLocation(Alan, UNMC)* is a potential context to be used as a preference, however, the preference set will be generated based on this preference if the context *hasLocation(Alan, UNMC)* is inferred by the agent confirming that the user is indeed located at *UNMC* (The University of Nottingham Malaysia Campus), and hence he expects preferred services available at *UNMC*. The tag indicator is used for general preferences and can be used to gather different rules into one group identified by the literal or tag given. For example, a rule with a tag of "L" may refer to the context related to location, hence all the rules with tag "L" are considered to be the members of the corresponding tag.

4.3 Preference Manager Layer

To incorporate the preferences, preference manager layer plays its role in managing the modules it carries, and to give a user the feel of personalization and also allows the inference engine to work with minimum overload. The general idea of the preferences provided is to extract a subset of rules from the whole rule base based on the user preferences. The preference manager layer is composed of Preference Set Generator (PSG), Context Monitor (CM), Context Set (CS), Context of Interest (COI), Context verifier (CV), and Derived Preference Indicator (DPI). The lower part of the Fig. 1 depicts the preference manager module and relationship between these components. The detailed description of the CS, CM and PSG can be found in [7]. Due to space limitations, we only briefly describe the newly added components.

- **Context Verifier (CV)** component is responsible for validating the contexts received from the sensors/agents and matches them with the user provided COI. If the COI matches with the sensed/received contexts then it can allow the PSG to generate the preference set. A straight forward example is location. If a user has COI *hasLocation(Alan, UNMC)* and the GPS sends the location as *hasLocation(Alan, Home)*, then it will drop the COI, as the location does not match with the COI. Hence the preference can not be added.
- **Derived Preference Indicator (DPI)** (or _COI) is responsible for generating a list of potential preferences from the COI. It matches a potential context with derived context in case a preference is enabled. If it finds a derived context that is being considered as a potential preferred context then DPI will send that context to the PSG. Unlike sensed/communicated context, derive context does not require validation and DPI directly sends it to the PSG.

To further elaborate the concept, let us suppose that an agent has a set of rules to model the behaviour of a person. Now a person can become patient if he is sick, which is a possibility. So, a system designer may add $\_Patient(Alan)$ as a derived preference. Which means that those rules related to the $\_Patient(Alan)$ will be added to the preference set once Alan gets sick.

5 Derived-Context Based User Preference

Since we have different indicators for the rules, it is necessary to determine the level of preferences required by the user. This mechanism is handled by the preference level monitor (PLM).

5.1 Preference Level Monitor (PLM)

Preference levels give user a choice of where the preferences are desired and up to which level the preferences are desired. The PLM can accommodate both the simple preference along with the facts/context value based preferences. As discussed in Sect. 4.2, the user can opt for any of the four different preference indicators. The Algorithm 1 goes through different checks to perform the better preferences and make the appropriate list of preferences. The algorithm presented in [17] has been revised to accommodate the derived context preference mechanism, changes are reflected in lines 16–22. One thing is to mention here is that the PLM Algorithm will make a separate list of derivable preference indicators, which will not be used by the CV, instead it will be passed once the contexts are derived. This is because, in advance, the CV will match the COI with the externally received contexts.

Since a system designer is aware of the different rules used to design the system and their possible outcome, it is fairly easy for him to use the preferences accordingly. In basic terminologies suppose we have a health care domain, where the system allows a user to monitor his blood pressure. The blood pressure can be categorised as High, Low and Normal levels besides declaring the user as a *Patient*. So, while keeping in mind that the possibility of a user to become a *Patient*, the *Patient* can be made as a derivable preference. Unless the user is derived as a *Patient*, the rules belong to the patient category will not be added to the corresponding preference set. In the next section, we explain the overall idea considering a simple case study.

6 A Simple Case Study

We consider a system consisting of a number of agents, including a person agent (Agent 1 represented by a smartphone) who is a user and may change his location detected by the GPS embedded into his smartphone. The user is also known to have his Blood pressure issues which is monitored by the BP device (Agent 2) and has heart rate monitor enabled (Agent 3). The user casually visits hospital

Input: **COI**: Current Context of Interest,_**COI**:Derivable COI,**R:** Rules, $\mathbf{F}_e$**:** Facts from external agents or sensors,$\mathbf{F}_d$**:** Facts derived, **CS:** Context Set, **Regex:** regular expression
Output: Preference Set based on COI

START
if *Regex(COI)==[a-zA-Z]* **then**
  Fetching Simple preference
  for *r→[R]* **do**
    if $\exists x \in$ ***COI*** *such that* $x \in$ ***CS[r]*** **then**
      Add r to Preference Set
    end
  end
end
else if *Regex(COI)==[a-zA-Z]+([a-zA-Z0-9]+)* ***OR*** *[a-zA-Z]+([a-zA-Z0-9]+,[a-zA-Z0-9])* **then**
  Fact-based preference of the form A(b) or B(b,c)
  for *r→[R]* **do**
    if $\exists x \in$ ***COI*** *such that* $x \in$ ***CS[r]*** *AND* $x \in \boldsymbol{F}_e$ **then**
      Add r to Preference Set
    end
  end
end
else if *Regex(_COI)==[a-zA-Z]+([a-zA-Z0-9]+)* ***OR*** *[a-zA-Z]+([a-zA-Z0-9]+,[a-zA-Z0-9])* **then**
  Derived-Context based preference of the form A(b) or B(b,c)
  for *r→[R]* **do**
    if $\exists x \in$ ***_COI*** *such that* $x \in$ ***CS[r]*** *AND* $x \in \boldsymbol{F}_d$ **then**
      Add r to Preference Set
    end
  end
end
else if ***CS[r]== "-"*** **then**
  Add r to general rule
end
END

Algorithm 1. PLM working algorithm

for the check up, and person agent can interact with Out Patient handling agent (Agent 4, located at Hospital). The user also has some preferences for his office which is located as UNMC. The office has an occupancy sensor (Agent 5), which can detect if the user is in the office or not.

6.1 Context-Based Preferences

As mentioned above, the user is not static and he may change his location time to time. When he arrives at hospital, his location is detected and processed to derive a new context being a patient. We will use this derived context to make

Table 1. Some example rules of Agent 1

Id	m	Rule	Identifier
R1	3	Patient(?p), hasBloodPressure(?p, Low) ⟶ hasSituation (?p, Emergency)	_Patient(Alan)
R2	3	Patient(?p), hasBloodPressure(?p, High) ⟶ hasSituation (?p, Emergency)	_Patient(Alan)
R3	2	Tell(2, 1, hasBloodPressure(?p, High)) ⟶ hasBloodPressure(?p, High)	_Patient(Alan)
R4	2	Tell(2, 1, hasBloodPressure(?p, Low)) ⟶ hasBloodPressure(?p, Low)	_Patient(Alan)
R5	1	Patient(?p), hasHeartRate(?p, Normal) ⟶ ∼ hasSituation(?p, Emergency)	_Patient(Alan)
R6	2	Tell(3, 1, hasHeartRate(?p, Normal)) ⟶ hasHeartRate(?p, Normal)	_Patient(Alan)
R7	1	Person(?p), GPS(?loc) ⟶ hasLocation(?p, ?loc)	-
R8	2	hasLocation(?p, Hospital), PatientID(101), hasPID(?p,101) ⟶ Patient(?p)	-
R9	2	Patient(?p), hasReason(?p, ?r), MedicalReason(?r) ⟶ isOutPatient(?p,?r)	_Patient(Alan)
R10	2	isOutPatient(?p, ?r) ⟶ Tell(1, 4, isOutPatient(?p, ?r))	_Patient(Alan)
R11	2	Tell(5, 1, hasOccupancy(?p, Yes)) ⟶ hasOccupancy(?p, Yes)	GPS(UNMC)
R12	2	hasOccupancy(?p,Yes) ⟶ Tell(1, 6, hasAircon(?p, On))	GPS(UNMC)

Table 2. Preference set transition

System status	COI	_COI	Facts in WM
Initial information	GPS(UNMC)	Patient(Alan)	PatientID(101), hasPatientID(Alan, 101)
Iterations of the system case scenario, where a user moves to different locations at different times with preferences enabled are GPS(UNMC) and Patient(Alan)			
User location	Derived facts	Preference indicator found in WM	Corresponding subset of rules
GPS(Home)	–	No	R7, R8
GPS(UNMC)	–	GPS(UNMC)	R7, R8, R11, R12
GPS(Hospital)	hasLocation(Alan, Hospital) Patient(Alan)	Patient(Alan)	R1, R2, R3, R4, R5, R6, R7, R8, R9, R10

a preference set for him at the hospital, which will illustrate how the sensed/externally received context-based preference as well as derived-context based preference work together to minimise the load on the agent's inference engine by reducing the number of rules while achieving the desired results. The rules in Table 1 are some example rules that are used to design Agent 1. The initial facts provided to the system are *PatiendID(101)* and *hasPatientID(Alan,101)*. The location is detected by the GPS sensor and also added to the agent's working memory as a fact. Once the COI is defined, the system checks and separates the COI from _COI. The _COI is put aside for the later use once the system starts working. As a result, the Table 2 shows us set of rules that are in the preference set for a given set of user provided preferences. In Table 2, we show the transition of facts, Context of Interest (COI) and how the rules are grouped. We assume that the initial location of the user is his Home. Later on, the user visits the

smart hospital and accordingly his location is detected which in turns deduce that the user is a Patient. Accordingly, the derived-context is used as a preferred context that helps generating a new set of rules by replacing the existing rules to be used in the agent's inference engine.

6.2 Rule-Based Preferences

It is always possible that a conflict occurs between the rules, and to resolve it we assign priorities (column m in Table 1) to the rules. The rule priorities give one rule preference over another rule. In this case study, we deliberately made a scenario where according to the facts we can have two different rules generating contradictory outcome as *hasSituation(Alan, Emergency)* and $\sim$*hasSituation(Alan, Emergency)*. Which if not handled can derive unwanted conclusion. Therefore, we assigned the priorities to rules, as a part of defeasible reasoning, and in the scenario described below, the rules R1 and R2 are assigned priority 3, while R5 has priority 1. Since R1 and R2 having higher priority than that of R5, the preference will be given to R1 and R2 over R5. Thus, avoiding any unwanted outcome. A more detailed discussion on defeasible reasoning can be found in [6].

7 Conclusion and Future Work

In this paper, we present derived-context based user preference as a personalisation mechanism into context-aware applications. The proposed approach supports preferences that could be easily controlled to personalise the system behaviour without modifying the internal settings or agent's program. We also present a revised algorithm to identify relevant user preferences. The research on context-aware user preferences, specifically on decision support system still in its early stages, many challenges remain in this area. In the future, we would like to explore the integration of social network based preferences into the system and analyse its effectiveness from different aspects, especially from the resource usage point of view.

References

1. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: Proceedings of the First Workshop on Mobile Computing Systems and Applications, pp. 85–90. IEEE Computer Society, Washington (1994)
2. Dey, A.K.: Understanding and using context. Pers. Ubiquit. Comput. **5**(1), 4–7 (2001)
3. Baldauf, M., Dustdar, S., Rosenberg, F.: A survey on context-aware systems. Int. J. Ad Hoc Ubiquit. Comput. Arch. **2**(4), 263–277 (2007)
4. Perera, C., Zaslavsky, A.B., Christen, P., Georgakopoulos, D.: Context aware computing for the internet of things: a survey. IEEE Commun. Surv. Tutorials **16**(1), 414–454 (2014)

5. Rakib, A., Haque, H.M.U., Faruqui, R.U.: A temporal description logic for resource-bounded rule-based context-aware agents. In: Vinh, P.C., Alagar, V., Vassev, E., Khare, A. (eds.) ICCASA 2013. LNICST, vol. 128, pp. 3–14. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05939-6_1
6. Rakib, A., Haque, H.M.U.: A logic for context-aware non-monotonic reasoning agents. In: Gelbukh, A., Espinoza, F.C., Galicia-Haro, S.N. (eds.) MICAI 2014. LNCS (LNAI), vol. 8856, pp. 453–471. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13647-9_41
7. Uddin, I., Rakib, A.: A preference-based application framework for resource-bounded context-aware agents. In: Kim, K.J., Joukov, N. (eds.) ICMWT 2017. LNEE, vol. 425, pp. 187–196. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-5281-1_20
8. Lai, J., et al.: Bluespace: personalizing workspace through awareness and adaptability. Int. J. Hum. Comput. Stud. **57**(5), 415–428 (2002)
9. Stefanidis, K., Pitoura, E., Vassiliadis, P.: Modeling and storing context-aware preferences. In: Manolopoulos, Y., Pokorný, J., Sellis, T.K. (eds.) ADBIS 2006. LNCS, vol. 4152, pp. 124–140. Springer, Heidelberg (2006). https://doi.org/10.1007/11827252_12
10. Hong, J., Suh, E.H., Kim, J., Kim, S.: Context-aware system for proactive personalized service based on context history. Exp. Syst. Appl. **36**(4), 7448–7457 (2009)
11. Hoque, M.R., Kabir, M.H., Seo, H., Yang, S.-H.: PARE: profile-applied reasoning engine for context-aware system. Int. J. Distrib. Sens. Netw. **12**(7) (2016)
12. Alhamid, M.F., Rawashdeh, M., Dong, H., Hossain, M.A., Saddik, A.E.: Exploring latent preferences for context-aware personalized recommendation systems. IEEE Trans. Hum. Mach. Syst. **46**(4), 615–623 (2016)
13. Kyriazakos, S., et al.: eWALL: an intelligent caring home environment offering personalized context-aware applications based on advanced sensing. Wirel. Pers. Commun. **87**(3), 1093–1111 (2016)
14. Fong, J., Lam, H.-P., Robinson, R., Indulska, J.: Defeasible preferences for intelligible pervasive applications to enhance eldercare. In: 2012 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 572–577. IEEE (2012)
15. Thanakodi, S., Nazar, N.S.M., Tzen, B.S.P., Roslan, M.M.M.: Survivor tracking system based on heart beats. In: Kim, K.J., Joukov, N. (eds.) ICMWT 2017. LNEE, vol. 425, pp. 550–557. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-5281-1_61
16. Uddin, I., Rakib, A., Haque, H.M.U.: A framework for implementing formally verified resource-bounded smart space systems. Mobile Networks and Applications **22**(2), 289–304 (2017)
17. Uddin, I., Rakib, A., Haque, H.M.U., Vinh, P.C.: Modeling and reasoning about preference-based context-aware agents over heterogeneous knowledge sources. Mob. Netw. Appl. (2017). https://doi.org/10.1007/s11036-017-0899-5

A Context Adaptive Framework for IT Governance, Risk, Compliance and Security

Shree Govindji, Gabrielle Peko(✉), and David Sundaram

Department of Information Systems and Operations Management,
University of Auckland, Auckland 1142, New Zealand
bgovl53@aucklanduni.ac.nz,
{g.peko,d.sundaram}@auckland.ac.nz

Abstract. The technological solutions offered today evolve at a rapid pace, as this happens, risk management and security practices are becoming more relevant and in fact, now a necessity for most growing organisation. Governance, Risk management and compliance (GRC) are established and well-adhered functions in a business which have individually always been very important in business management. As individual topics, the application of all concepts have been fundamental for businesses in order to manage risks. However, over the years, the term GRC was developed and applied to describe the integration between the various areas due to the reason that a monolithic approach between the functions was no longer feasible in successful management of business risk. However IT GRC has been dealt with an isolated manner from IT Security. In this paper we explore IT GRC and Security and propose an integrated context adaptive framework that addresses the problems of monolithic approaches.

Keywords: Governance · Risk management · Compliance
Information technology · Security · Context adaptive

1 Introduction

According to De Smet and Mayer [2], the main challenge of GRC is to have an approach which is as integrated as possible. Integrated GRC was developed to manage the increasing business complexity due to new legal requirements enforced as a result of various financial scandals and business failures. Racz [3] proposed the first scientific definition to the term stating that "GRC is an integrated, holistic approach to organization-wide governance, risk and compliance ensuring that an organization acts ethically correct and in accordance with its risk appetite, internal policies and external regulations, through the alignment of strategy, processes, technology and people, thereby improving efficiency and effectiveness". This definition however, does not consider the security aspect of GRC and so we will consider other suitable definitions too as security is an important aspect of GRC, but it has failed to be mentioned by most researchers exploring GRC topics and concepts. A GRC approach does assist organisations in their approach for IT security and IT Security can benefit from an integrated GRC view [4]. Security vulnerabilities have risks which must be constantly monitored and evaluated in order to reduce the opportunity of a breach [1]. Managing the security

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 14–24, 2018.
https://doi.org/10.1007/978-3-319-77818-1_2

architecture is important in managing a global risk and compliance platform, IBM acknowledges this and provides their own solutions to GRC which considers the security aspect and also fills the gap missed by most researchers. Many other consulting companies have also proposed similar solutions, identifying that there is indeed a strong, inseparable link between GRC and security capabilities. Vicente and Da Silva [5] have identified the young age of scientific research around GRC. In more recent studies, Racz [6] has also mentioned that there is a lack of a scientifically grounded definition, stating that most GRC related definitions are published by software vendors and consultants and are suited to their products and services. During the time of this writing, Racz [3, 6] claim is supported by the research contributed by De Smet and Mayer [2] who have identified that more research is still needed to define the integration between various terms.

In the next section we define the significant terms in GRC and security before discussing IT GRC in Sect. 3 and IT security in Sect. 4. Section 5 introduces the integration between IT GRC and IT security. Then, in Sect. 6, an Integrated IT GRC Security (GRCS) framework synthesizing ideas, theories, and models from these two concepts is presented. The paper concludes in Sect. 7.

2 GRC and Security

In order to better understand the integration between GRC, we need to first define each individual term and so the following provides a brief definition of governance, risk management and compliance.

Governance/Corporate Governance: Defined as a set of processes, policies and laws affecting the way an enterprise or corporation is directed or controlled. Corporate governance principles which are well defined and enforced provide a structure that suits all stakeholders concerned, ensuring the company follows regulations, ethical standards and best practices [7]. It deals with internal and external aspects of an organization [8]. The past failure of many large organisations has prompted policy makers to initiate legislative reforms which require disclosure and reporting of organisational risks. The Sarbanes-Oxley Act, for example, was the government's response to the Enron scandal, the large US energy company which collapsed due to a reduced perception of debt and risk and overstatement of revenues as a result of undisclosed ownership structures [9].

Risk Management: An enterprise wide risk management approach supports corporate governance. The Committee of Sponsoring Organizations of the Treadway Commission (COSO) provide a suitable definition for enterprise wide risk management which is widely accepted, defining ERM as "a process, effected by an entity's board of directors, management and other personnel, applied in strategy setting and across the enterprise, designed to identify potential events that may affect the entity, and mange risks to be within its risk appetite, to provide reasonable assurance regarding the achievement of objectives" [10].

Compliance: According to Fowler-Rians [11], regulatory compliance is achieved through meeting expected behaviors in processes and practices. It refers to adherence to

established guidelines, internal policies, regulations or legislative obligations by an organization. i.e. company compliance with the Sarbanes-Oxley legislation and a growing body of other regulations and laws.

Integration of GRC: As the number of legislative rules and regulations increase, organisations have to deal with increased risks. These concerns lead companies to approach governance, risk management and compliance functions in a separate manner [8]. Growth in each specific area led to cost concerns which initiated an integrated governance, risk and compliance approach that would look across an organisations risk and control functions holistically and seek to improve both organisational efficiency and effectiveness of risk and control functions [12]. According to Rasmussen [13], an integrated enterprise view of risk and compliance means accountability is effectively managed and businesses have a complete system of record which subsequently provides visibility across multiple risk and compliance issues. This also introduces a sustainable view for business procedures as the increasing business risks and threats can be minimized with a holistic and integrated approach on GRC issues. Rasmussen [13] also mentions how a siloed GRC approach means there is less framework for managing risk and compliance as integrated business functions, this in turn leads to poor visibility across the organisation. Other outcomes of an unintegrated GRC approach includes: wasted resources and spending, poor visibility across the enterprise, overwhelming complexity, lack of business agility, greater exposure and vulnerability [13]. Recor and Hu [7] also mention that leveraged integration through the improvement of GRC processes can guide organisations to reach their overall objectives by ensuring that there is connectivity between risks, strategy and performance.

GRC and Security: In today's dynamics, the demand for accountability, regulatory compliance and security are increasing as these are mandatory areas of business which need to be covered, this leads to GRC of information security becoming a high priority goal [14]. Asnar and Massacci [14] have also identified that a process to govern security is missing at an organisational level. In their research [14], have developed on the link between GRC and information security, describing the importance of a GRC management process for information security. However, whilst there is a strong relationship between GRC and security, it is suitable to say that there is otherwise a lack of research in terms of the integration between the two topics. In contrast, there are a wide variety of organisational and industry articles mentioning the importance of integration between GRC and security. For example Rashid [15] has mentioned that GRC programs allow security professionals to gain visibility into organisational risks. Security professionals often work very closely with risk managers and both the risk and security functions interlink. Risk managers who look after GRC initiatives may be misinformed when they aren't fully briefed about information security, leading to conflicting situations [16].

AMR Research [17] shows that security purposes were fourth in reasons for companies investing in GRC solutions, this is a clear example of how GRC closely initiates with security and there is an opportunity to cover this gap in research literature. While there is a lack of research linking GRC and security together, it is easy to see how information security is involved in each aspect of the GRC components. Governance needs to be incorporated into the organizations IT security frameworks in order to ensure the effectiveness of information security governance [18].

3 IT GRC

According to Racz et al. [9], IT GRC is the term used for when GRC activities are restricted for IT operations. Risks and controls are interconnected with IT activities, resulting in a number of benefits for the organisation. The GRC integration process is streamlined through the use of technology, and IT can be a driver or enabler of integration among governance, risk management and compliance [16]. IT GRC has expanded throughout the years as technology replaces more and more manual processes. [3], found that at the time of writing their research piece, there was a lack of research on integrated approaches to IT GRC. More recent studies, however, still support the fact that there is a lack of attention on IT GRC, especially from the scientific community [2]. It is also mentioned that the link between IT governance and risk management is neglected [2].

The main reason for implementing IT GRC strategies was historically due to increasing regulatory pressure and a drive to lower the costs which were originally gained from the siloed approach [7]. Success in today's business environment requires that organisations integrate, build and support business processes which are built on a common technology backbone [13]. Information technology can streamline the GRC integration process, making it more cost effective [16]. Properly aligning IT with business strategies can enable technology to be used for value creation and competitive advantage. An IT GRC program also contributes further to each component in GRC. According to Linkous [19], an integrated IT GRC program provides value to the compliance processes and can improve the information assurance efforts. Each component of IT GRC is interrelated to each other, and therefore an IT GRC program is more effective rather than implementing just one or two of the components. For example, the attention on IT governance is captured through enforcing compliance measures. IT Governance also governs IT RM and IT Compliance activities. Through a critical analysis on prior research, Racz [3] found that none of the chosen models claiming to integrate GRC had fully covered all aspects, on top of that, none of the models elaborated on IT GRC specifically. After identifying this gap, Racz [3] proposed a detailed scientific model for integrating IT governance, risk and compliance management.

Through this research it is identified that there is a lack of research articles with an IT GRC focus within specifically the banking sector. This identifies that there is an opportunity to contribute in this area, and also contribute to IT GRC applications in various other industry-specific areas.

4 IT Security

With the adoption of IT security being a mandatory task for most, if not all, organisations in today's environment, experts are finding it increasingly difficult to apply holistic measures across different domains. Adopting a risk management perspective is not enough to completely eliminate the security risk, hence the reason we are not integrating security within GRC, but rather taking a separate approach to consider security on its own. Very often, there is insufficient knowledge about the security

domain, threats, countermeasures and company infrastructure, leading to wrong decision making [20]. Ekelhart et al. [20] identify that the main reasons for this happening is due to the vaguely defined security terminology and because managers who make decisions are often not understanding the complexity of underlying IT infrastructure [20]. Damianides [21] also identifies how there is little consideration given to organisational requirements and priorities and in the past, information security would be dealt as a solely technological issue. Damianides recommends that information security should be addressed in all phases of a project. According to Grob et al. [22], Information security management (ISM) is focused on organisations information systems operating at a faultless service level. Traditionally, ISM focuses on the consideration of technical systems, such systems can cause operational business risks and therefore these IT related risks must be identified and adequate countermeasures must be defined. Analyzing threats within the scope of ISM is occasionally defined as risk management [22]. Grob et al. [22] have also identified that there needs to be a functional alignment between operational risk management (ORM) and ISM as ISM has more of a system-based focus and therefore can capture possible threats better, whereas ORM focuses more on the overall amount of damage impacting business processes. The perception of risks in an organisation is influenced by the lack of security culture and training. Grob et al. [22] have depicted the misalignment between ORM and ISM.

The human element which challenges information security involves a number of aspects. Firstly, security risks not only need to be effectively communicated to stakeholders but also require a mutual understanding between the stakeholders. Human errors also threaten best security practices. Human errors are defined by Kraemer and Carayon [23], as non-deliberate accidental cause of poor computer and information security. Kraemer and Carayon [23] have also identified the main factors which causes errors in information security, these errors can be traced back to poor communication, security culture and security policy, including a number of other issues which the authors have identified through their study. Humans are the cause for many information security breaches, and decision makers can make decisions which contribute to risk and impact an organisations response to threats. In fact, the biggest IT security risk is the human element [2] and many prior events such as the Enron and WorldCom scandals reaffirmed this.

The organisational element refers to factors such as organisational size, top management support and type of industry which has an influence on how effective information security controls are within organisations [24]. Other factors such as uncertainty of environmental elements, rapid change of technology, competitors' behaviours and customers' security requirements, and changes in legislation also have an impact on the way security is managed in an organisation [24]. Top management support has been identified as an important factor which is critical for implementing security controls within organisations [25]. Werlinger et al. [26] have identified through their own research how a lack of security culture in an organisation makes it difficult to change existing security practices.

The technological complexities are another challenge which contribute to not being able to maximize full security efforts. Testing security systems are a costly, lengthy and a complex process which is why many organisations have difficulty in this area. Werlinger et al. [26], have identified that network and system complexity is challenging

for organisations who are even wanting to implement security controls. Other IT complexities involve decentralization of IT management, mobility and distribution of user access, security updates and consistent installation and a lack of support for using security tools [26] which all contribute to the complexity of IT security related changes.

Regardless of all the available frameworks, many organisations are struggling with implementing IT security measures for two reasons: (1) they may not have a comprehensive security strategy, (2) their security strategy isn't updated to reflect changes in their business, cyber security practices and IT platforms [1]. The resulting threat to IT security includes a costly security breach.

5 IT GRC and IT Security

Executive boards and management have a number of fundamental responsibilities associated with information security governance, including understanding why information security need to be governed, and ensuring it fits in the IT governance framework [21]. IT GRC is similar to GRC in the sense that it has been identified that there is minimal research articles conducted on the integration of IT GRC and IT security. However, when looking at articles outside of the research field, we are able to identify that there is in fact integration between IT GRC and security in the current business world. According to PwC (2017), IT GRC is defined as "Combining disciplines for better enterprise security. Adopting a unified IT governance, risk management and compliance (IT GRC) approach, and managing the associated activities coherently will create efficiencies, provide a holistic view of the IT environment and ensure accountability". An IT GRC program links with security in a number of ways and in order to support effective communications, the IT GRC program should provide the ability to allow different categories of users to view risk and compliance data in their own relevant ways, these users may range from IT operations, risk managers, auditors and even security operations [19]. While security is a distinct function, it is still very much interrelated with risk-related functions and so it is important to consider security as a distinct part of IT GRC functions too.

IT Governance and IT Security: IT governance and information security are linked through the development of information security governance practices. According to Da Veiga and Eloff [27], Information security governance can be defined as the overall manner in which information security is deployed to mitigate risks. The concept arises when it was found that communication of the information security culture and control frameworks is the responsibility of company executives. Da Veiga and Eloff [27] also mention that organisational risks can only be addressed when a governance framework for information security is in place. While there is a large link between the two concepts, there is a lack of research on the integration of IT governance and IT security management elements, while IT governance is viewed as a component of the wider IT management model [2]. Certain characteristics of IT governance and security governance contribute to more effective alignment and execution of IT programs. In relation to certain regulations, for example the SOX, security is no longer just an IT issue, an effective IT and security governance program is essential. Security and risk management are a key part of the IT governance framework, but more research is still needed

to guide how this integration should occur [2]. In order to meet the Sarbanes Oxley requirements, it should not be considered as just a compliance process, but also an opportunity to develop strong governance models.

IT Risk Management and IT Security: The relationship between risk and IT security is inseparable, in essence, IT security is solely performed to mitigate risks [4]. According to Parent and Reich [28], there are three primary areas which IT risk management targets: the security of data and information, the integrity of hardware and systems and IT project implementations [28]. The management of technology risk is synonymous with information security, leading to an under appreciation of both concepts. [2] Have also proposed through their research, that integrating IT risks in the decision making framework will accommodate for information security aspects. As Grob et al. [22] have identified, the IT risk analysis function within IT risk management serves as a basis for identifying and implementing measures for risk governance. Risk governance is achieved by avoiding, passing, decreasing or accepting risks and in the context of information systems, IT security experts can conduct such measures for risk governance due to their competencies [22]. A number of standards and best practices for IT security management have been established and offer extensive improvements within IT risk management efficiency [22].

IT Compliance and IT Security: IT security can be driven by IT compliance and appears with regulations which assist with data protection and privacy. Frameworks such as HIPAA, COBIT and ISO17799 help organisations establish a comprehensive approach to both privacy compliance management and information security [19]. Linkous [19] also mentioned how the SOX helped organisations adapt a holistic approach to security and privacy compliance as having SOX in effect as boards of directors began to be interested in security compliance. In essence, as the landscape for information security becomes more complex, organisations have to ensure their compliance requirements address any regulatory and non-regulatory changes. Many employees in IT security departments are acting without the knowledge of the regulatory requirements and what these require in terms of regulatory compliance, hence the reason it is important to strengthen the connection of IT security and compliance requirements [4]. In this environment, information security initiatives are faced with increasing regulatory and compliance pressures, this is leading to the development of security-specific compliance frameworks. Such actions are directing security managers into more IT GRC based activities.

6 An Integrated IT GRCS Framework

As recommended by [2], more research is still needed to define how well to integrate both security and risk management into organisations IT governance frameworks. In contrast to this however, they are many organisational resources which can be useful in identifying the link between IT GRC and security. Most organisations adopting an IT GRC program are often missing the security component, therefore addressing this

problem through the development of their own IT GRC/IT Security based solution. We can see from this that it is not possible to separate the two, and often, if not mentioned as a separate topic, IT security is already embedded into IT GRC in one way or another. Past research has already begun to demonstrate how effective compliance initiatives are linked to direct benefits with company revenue, profits and customer retention, therefore it has been predicted that a baseline for security activities will include information security moving towards mandated and standardized frameworks. Based on our findings, we propose our own framework which addresses some of the identified gaps in our research. The bottom line is that there are not enough research papers that address GRC and security given the very important and blatant link between the two, especially in the context of IT. Therefore, we firstly present a high level framework for IT GRCS in Fig. 1.

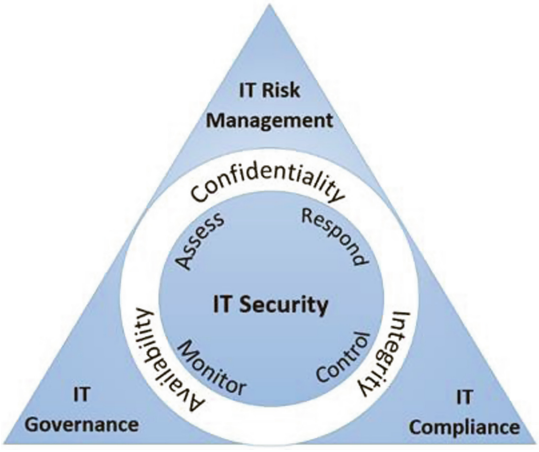


Fig. 1. Context adaptive IT GRCS framework

This framework incorporates all elements of IT GRCS into a simplified model, with IT security being in the middle as it is incorporated in each pillar for IT GRC. The CIA (confidentiality, integrity, availability) concept is a vital dimension in the model, it guides policies for IT and information security in organisations to protect all organisational assets. The process involving assess, respond, control and monitor, identified in our IT security Framework, was what we referred to when developing this model. However, we noticed that a similar process can be applied across all pillars of GRCS. These four steps helps an organisation to adapt to situation depending on context. Next we also propose a more detailed model (Fig. 2) which digs deeper into each pillar of IT GRCS and we are able to see how this framework can be applied in an organisational context. And every aspect adapts as the context changes and reacts to changes in the other elements.

Firstly, for the IT Security pillar we can see that there is an additional component which incorporates people, data, information, applications, network and infrastructure with our process model for IT security. This component has been derived from IBMs Security framework and is a good reference model as we can see that protecting IT within all these areas is vital for IT security. From the IT Security pillar, there are feedback loops to the IT GRC pillars, which shows the incorporation of IT GRCS now. The process model for IT Governance has been derived from Cobit 4.1 and has been chosen as it is both suitable and simple for our model. The process model for IT Risk Management has been derived from ISACAs Risk IT framework which includes a set of guiding principles for effective management of IT risk. It also complements COBIT

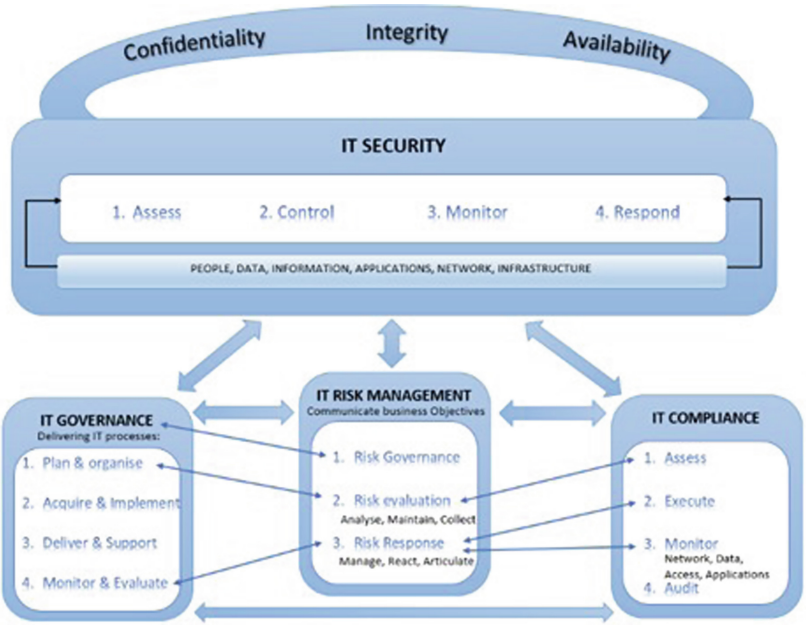


Fig. 2. Detailed IT GRCS framework

and therefore is suitable to link with our IT Governance pillar. Finally, the process model for IT Compliance is derived from a compliance process framework again by ISACA. We chose this model as it is the model suitable for IT compliance, as in our research there was a lack of frameworks and models specifically for IT compliance. We can see the link with our identified process model to IT as the monitor step refers to components from the IT security section, and also there is an audit process, which is vital for IT compliance.

7 Conclusion

In conclusion, we have identified in our research that while IT GRC has been around for a number of years now and has been an widely researched especially since the collapse of major financial organisations, there is very little literature from both academia and industry articles which propose frameworks for incorporating GRC along with IT, and especially including the IT security component. We have identified that while security is an inadmissible component in each pillar of IT GRC, it is often not mentioned – perhaps because of the assumption that it is already incorporated. Therefore we propose a framework which incorporates both IT GRC and IT Security in order to form IT GRCS. While the framework is generic, it can be applied in various sectors and there are many potential areas where further research can be done such as seeing the suitability of the framework in specific types of industries.

References

1. IBM: SAP Security and GRC Services (2015). https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SES03016USEN
2. ISACA: The Risk IT Framework – Excerpt (2009). http://www.isaca.org/knowledge-center/research/documents/risk-it-framework-excerpt_fmk_eng_0109.pdf
3. De Smet, D., Mayer, N.: Integration of IT governance and security risk management : a systematic literature review, no. 1, pp. 143–148 (2016)
4. Racz, N., Seufert, A., Weippl, E.: A process model for integrated IT governance, risk, and compliance management. In: Proceedings of the Ninth Baltic Conference on Databases and Information Systems (DB&IS 2010), p. 155 (2010)
5. Kuppinger, M.: IT GRC and IT Security - Where is the link? (2010). https://www.kuppingercole.com/blog/kuppinger/grc_it_security_link180210
6. Vicente, P., Da Silva, M.M.: A business viewpoint for integrated IT governance, risk and compliance. In: 2011 IEEE World Congress on Services, pp. 422–428 (2011)
7. Racz, N., Weippl, E., Seufert, A.: A frame of reference for research of integrated governance, risk and compliance (GRC). In: De Decker, B., Schaumüller-Bichl, I. (eds.) CMS 2010. LNCS, vol. 6109, pp. 106–117. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13241-4_11
8. Recor, J., Xu, H.: GRC technology introduction. In: Tian, W. (ed.) Commercial Banking Risk Management, pp. 305–331. Palgrave Macmillan US, New York (2017). https://doi.org/10.1057/978-1-137-59442-6_14
9. Racz, N., Weippl, E., Seufert, A.: Governance, risk & compliance (GRC) software – an exploratory study of software vendor and market research perspectives, pp. 1–10 (2011)
10. Smith, R.: Seven things you need to know about IT controls. SOX Committee Integration Consortium (2004). www.integrationconsortium.org
11. COSO (2004). https://www.coso.org/documents/COSOBoardsERM4pager-FINALRELEASEVERSION82409_001.pdf
12. Fowler-Rians, K.: Determinants of federal regulation compliance: a study of the employee trip reduction program. Unpublished Doctoral Dissertation, University of Houston (1997)
13. Frigo, M.L., Anderson, R.J.: A strategic framework for governance, risk, and compliance. Strateg. Financ. **90**(8), 20–61 (2009)
14. Rasmussen, M.: Value of a Common Architecture for GRC Platforms Business Burdened by Varying Risk & Value of a Common, pp. 1–8 (2010)
15. Asnar, Y., Massacci, F.: A method for security governance, risk, and compliance (GRC): a goal-process approach. In: Aldini, A., Gorrieri, R. (eds.) FOSAD 2011. LNCS, vol. 6858, pp. 152–184. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23082-0_6
16. Rashid, F.: How to Leverage GRC for Security (2013). http://www.bankinfosecurity.com/how-to-leverage-grc-for-security-a-6164
17. Anand, S.: Technology and the Integration of Governance, pp. 57–59, December 2010
18. AMR Research: November 2009 GRC in 2010 : $ 29.8B in Spending Sparked by Risk, Visibility, and Efficiency (2010)
19. Linkous, J.: Put the "i" in IT compliance. Commun. News **45**(12), 26 (2008)
20. Ekelhart, A., Fenz, S., Klemen, M., Weippl, E.: Security ontologies: improving quantitative risk analysis. In: Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 1–7 (2007)
21. Damianides, M.: Sarbanes-Oxley and it governance: new guidance on it control and compliance. Inf. Syst. Manag. **22**(1), 77–85 (2005)

22. Grob, H.L., Strauch, G., Buddendick, C.: Applications for IT-risk management – requirements and practical evaluation, pp. 758–764 (2008)
23. Kraemer, S., Carayon, P.: Human errors and violations in computer and information security: the viewpoint of network administrators and security specialists. Appl. Ergon. **38**, 143–154 (2007)
24. Chang, S.E., Ho, C.B.: Organizational factors to the effectiveness of implementing Information security management. Ind. Manag. Data Syst. **106**(3), 345–361 (2006)
25. Kankanhalli, A., Teo, H.-H., Tan, B.C., Wei, K.-K.: An integrative study of information systems security effectiveness. Int. J. Inf. Manag. **23**, 139–154 (2003)
26. Werlinger, R., Hawkey, K., Beznosov, K.: An integrated view of human, organizational, and technological challenges of IT security management. Inf. Manag. Comput. Secur. **17**(1), 4–19 (2009)
27. Da Veiga, A., Eloff, J.: An information security governance framework. Inf. Syst. Manag. **24** (4), 361–372 (2007)
28. Parent, M., Reich, B.: Governing information technology risk. Calif. Manag. Rev. **51**(3), 134–152 (2009)

Hybrid Classifier by Integrating Sentiment and Technical Indicator Classifiers

Nguyen Duc Van[1(✉)], Nguyen Ngoc Doanh[2,4], Nguyen Trong Khanh[3,4], and Nguyen Thi Ngoc Anh[1,4]

[1] Hanoi University of Science and Technology, No. 1, Dai Co Viet, Hanoi, Vietnam
vanndkstnk57@gmail.com, anh.nguyenthingoc@hust.edu.vn
[2] ThuyLoi University, No. 175 Tay Son, Dong Da, Hanoi, Vietnam
[3] Post and Telecommunications Institute Technology, Ho Chi Minh City, Vietnam
[4] IRD, Sorbonne Universités, UPMC Univ Paris 06 Unité Mixte Internationale de, Modélisation Mathématique et Informatiques des Systèmes Complexes (UMMISCO), 32 Avenue Henri Varagnat, 93143 Bondy Cedex, France

Abstract. Classifiers in stock market are an interesting and challenging research topic in machine learning. A large research has been conducted for classifying in stock market by using different approaches in machine learning. This research paper presents a detail study on integrating sentiment classifier and technical indicator classifier. The research subject is investigated to classify a stock into one of three labels being top, neutral or bottom. First, using technical indicators such as relative strength index (RSI), money flow index (MFI) and relative volatility index (RVI) to classify stock, then using bagging of learning machine to classify the stock. Second, using sentiment data to classify the stock. Third, integrating technical indicator and sentiment classifiers to build hybrid classifier. In this study, hybrid machine learning by combining sentiment and technical indicator classifiers is proposed. We applied this proposal hybrid classifier for five stocks in VN30. The empirical results show hybrid classifier stock has more power than single technical indicator classifier or sentiment classifier.

Keywords: Machine learning · Stock market · Classifier · Sentiment analysis Hybrid classifier · Technical indicator

1 Introduction

Recently, more and more researchers concentrate on analysing sentiment factors of stock market. This paper tests whether hybrid classifier integrating sentiment factors and technical indicator more power than single classifier.

In the fact that, hybrid machine learning has been studied by some other researches such as Gao and Yang [1]. They integrated sentiment factors and price volume factors. Gao and Yang [1] show that mixed-frequency stock index combining sentiment factors and price volume factors have positively predictive power statistically. Moreover,

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 25–37, 2018.
https://doi.org/10.1007/978-3-319-77818-1_3

mixed-frequency stock index futures sentiment and mixed-frequency stock index sentiment has greater positively predictive power in high sentiment period [1].

Moreover, marching learning is applied in stock market by Ballings et al. [2]. They studied the benchmark ensemble methods (Random Forest, AdaBoost and Kernel Factory) against single classifier models (Neural Networks, Logistic Regression, Support Vector Machines and K-Nearest Neighbor) [2]. They gathered data from 5767 publicly listed European companies and used the area under the receiver operating characteristic curve as a performance measure. The results indicate that Random Forest is the top algorithm followed by Support Vector Machines, Kernel Factory, AdaBoost, Neural Networks, K-Nearest Neighbors and Logistic Regression [2].

Sentiment data is important information related to news can be good, bad or neutral. The statistical analysis of relatively simple sentiment cues can provide a surprisingly meaningful sense for investors. Thus, integrating sentiment analysis and technical indicator in classifying stock is investigated in this paper. We used classifier that is called ensemble. First, we used technical indicator to classify a stock into one of three labels top, neutral and bottom. Then, sentiment data was used to classify by method that Lagarde and Arnaud was presented [4]. Last, a new classifier method to classify stock was proposed by integrating technical indicator classifier and sentiment classifier.

The research problem of this paper is applied the integrating technical indicator and sentiment classifier to label any stock. This proposal classifier predict a stock should buy or sell in the future so it could be supported investors in their decision.

After assigning the label for each stock, evaluation the results have some methods such as vote classifiers, Naïve Bayes in Ranking, min max classifiers [4]. The contribution of this paper is not only proposing new hybrid classifier but also using vote classifier and max classifier to evaluate the classifier results.

A case study in applying the proposal new hybrid is five stocks in VN30 of Vietnamese stock market. VN30 includes 30 stocks that are the most importance in Vietnamese stock market by capitalization and liquidity.

The rest of this paper is organized as follows. Section 2 – research methodology includes research methods, data collection and data analysis methods chosen. In Sect. 3, the proposal approach is applied in a stock. Discussions and ideas for further work and a short summary of the paper and the conclusions are presented in Sect. 4.

2 Research Methodology

2.1 Sentiment Definition

The market is driven by emotion of investors thus market sentiment is about feelings and emotion. Sentiment measures the positivity and negativity of references about the

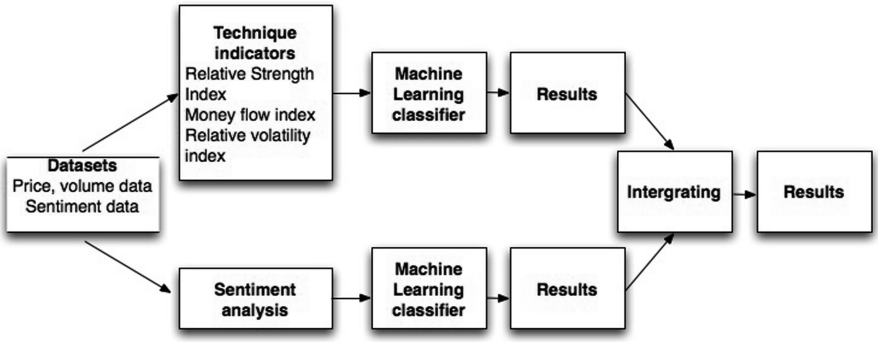


Fig. 1. Block diagram of proposal research methodology.

specific stock. The higher the measure is, the better the view of the stock is. On the other hand, market sentiment is generally described as bearish or bullish which is considered below (Fig. 1).

2.1.1 Bullish and Bearish

Returns: The close-to-close daily returns of stock i at the day t denoted $R_{t,i}$ are calculated as follows [1]:

$$R_{t,i} = 100 * \ln\left(\frac{S_{t,i}}{S_{t,i-1}}\right) \tag{1}$$

where $S_{t,i}$ is price of the stock i at the day t.

Bullish: Bullish sentiment is defined by expectations of investors who believe that stock prices will rise over time [7].

Neutral: Neutral sentiment is defined by expectations of investors who believe that stock prices will stay essentially unchanged over time [7].

Bearish: Bearish sentiment is defined by expectations of investors who believe that stock prices will fall over time [7].

2.1.2 Sentiment Ratio

Weekly, Investor's Intelligence that uses information polled directly from market professionals publish market sentiment indicator [1]. This index expresses the sentiments of investors that deal daily within the financial markets [1].

The high/low sentiment indicator compares the number of stocks making n-day highs to the number of stocks making n-day lows.

$$Ratio_{t,i} = \frac{\sum_{t=1}^{n} H\left(RT_{t,i}\right)}{\sum_{t=1}^{n} L\left(RT_{t,i}\right)} \tag{2}$$

where $H\left(RT_{t,i}\right) = \begin{cases} RT_{t,i}, & RT_{t,i} \geq 0 \\ 0, & RT_{t,i} < 0 \end{cases}$, $L\left(RT_{t,i}\right) = \begin{cases} 0, & RT_{t,i} \geq 0 \\ RT_{t,i}, & RT_{t,i} < 0 \end{cases}$

When stock prices are trading at their lows (highs) across the board, it means traders have a bearish (bullish) market sentiment.

2.1.3 Sentiment Trading Strategies (STS)

Sentiment trading strategies (STS): we compute the time-t returns based on the sign of the past cumulative sentiment from time t − n − 1 to t − 1. For each stock i and day t, we consider whether the past cumulative sentiment over the past n days is positive or negative. If the past cumulative sentiment over the past n days is positive, we go buy the stock. If the past cumulative sentiment over the past n days is negative, we sell the stock. We calculate a single time series of daily returns [1].

Sentiment trading strategies:

$$\sum_{t=1}^{n} RT_{t,i} \tag{3}$$

The fact that sentiment analysis is used to classify the stock. Concretely, this paper sentiment data such as bullish sentiment, neutral sentiment, bearish sentiment, sentiment ratio and sentiment trading strategies are used in classifier the stocks that will be presented in the next section.

2.1.4 Sentiment Classifier Formula

Lagarde and Arnaud proposed the effective method to classify a stock using sentiment data [4]. A strategy using sentiment data is derived. Thus, we reuse the idea in classifying sentiment that is shown by the formula as follows:

$$\text{Target}(\text{A}, \text{t}) = \ln\left[\frac{\text{Max}\left(\text{A}_{\tau\in(\text{t}\leq\tau\leq\text{t}+182)}\right)}{\text{A}_{\text{t}}}\right] + \ln\left[\frac{\text{Min}\left(\text{A}_{\tau\in(\text{t}\leq\tau\leq\text{t}+182)}\right)}{\text{A}_{\text{t}}}\right] \tag{4}$$

- $\text{Target}(\text{A}, \text{t}) \geq 0.0953$ will be considered as top which points to strong positive sentiment. Thus, the stock is placed into top class.
- $\text{Target}(\text{A}, \text{t}) \leq -0.223$ will be considered as bottom which points to strong negative sentiment. Thus, the stock is placed into bottom class.
- $0.0953 \geq \text{Target}(\text{A}, \text{t}) \geq -0.223$ will be considered as bottom which points to strong negative sentiment. Thus, the stock is placed into bottom class.

2.2 Technical Indicator

Technical indicators compound mathematical formulas that are used price time series data to create another time series data. These are 24 basis technical indicators are used of trading rules such as Moving average, moving average envelopes, Triple Exponential Moving Average, Average True Range (ATR), Moving Average Convergence Divergence (MACD), Bollinger Bands, bandwidth, Relative Strength Index (RSI), Relative Volatility Index (RVI), stochastic oscillator, ultimate oscillator, rate of change, relative vigor index, Ease of Movement (EMV), On Balance Volume (OBV), Accumulation Distribution Line (ADL), Chaikin Oscillator, Chaikin Money Flow (CMF) and Money Flow Index (MFI) [9]. Technical indicators can be classified into three groups that are trend, momentum and volatility based indicators. Indicators in each group are similar with each other [9] thus, this paper includes three technical indicators RVI (volatility), MFI (trend) and RSI (momentum) chosen to be on behaft of each group.

2.2.1 Relative Volatility Index (RVI)

The relative volatility index (RVI) is a volatility indicator that was developed by Donald Dorsey to indicate the direction of volatility [9]. The RVI is plotted in a range from 0 to 100 and is often used as a confirmation for other indicators such as moving average or Money Flow Index indicators.

– $S_{t,i}$ is close price of stock i at the day t, we denote

$$\mathrm{High}_{t,i} = \begin{cases} S_{t,i}; & S_{t,t} > S_{t-1,i} \\ 0; & \text{otherwise} \end{cases}$$

$$\mathrm{Low}_{t,i} = \begin{cases} S_{t,i}; & S_{t,t} < S_{t-1,i} \\ 0; & \text{otherwise} \end{cases}$$

$$\overline{\mathrm{High}_{t,i}} = \frac{1}{n}\sum_{t=1}^{n} \mathrm{High}_{t,i}, \overline{\mathrm{Low}_{t,i}} = \frac{1}{n}\sum_{t=1}^{n} \mathrm{Low}_{t,i} \tag{5}$$

$$u_{t,i} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n} \left(\mathrm{High}_{t,i} - \overline{\mathrm{High}_{t,i}}\right)^2},$$

$$d_{t,i} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n} \left(\mathrm{Low}_{t,i} - \overline{\mathrm{low}_{t,i}}\right)^2} \tag{6}$$

$$\mathrm{Dsum}_{t,i} = \sum_{t=1}^{n} d_{t,i}, \mathrm{Usum}_{t,i} = \sum_{t=1}^{n} u_{t,i}$$

The RVI formula is defined as follows

$$RVI = 100 \times \frac{Usum}{Usum + Dsum} \tag{7}$$

- RVI reading above 60 (or 70) would be considered as top which points to strong up-trend and the possibility of overbought condition. Thus, if RVI of a stock is higher than 60 (or 70), then it is placed into top class.
- RVI reading below 40 (or 30) would be considered as bottom which points to strong down-trend and the possibility of oversold condition. Thus, if RVI of a stock is lower than 40 (or 30), then it is then it is placed into bottom class.
- RVI reading between 40 to 60 (or 30 to 70) would be considered as neutral. Thus, if RVI of a stock is lower than 40 (or 30), then it is then it is placed into neutral class.

RVI is behalf volatility technical indicator group, in the next subsection we will present about volatility MFI technical indicator.

2.2.2 Money Flow Index (MFI)

MFI is an index used to measure the strength of cash flow in the market. They use both price and volume to measure sales [8]. The line is positive when the price rises, and negative when the price falls. MFI represents a momentum relative to mass, which is appropriate for determining the direction of prices and signals in the market [8]. Let $high_{t,i}, low_{t,i}, close_{t,i}, volume_{t,i}$ denote high price, low price, close price and volume of the stock i at the day t respectively. Let n be the size of time series data.

$$Typical\ price_{t,i} = \frac{high_{t,i} + low_{t,i} + close_{t,i}}{3} \tag{8}$$

$$Money\ flow_{t,i} = typical\ price_{t,i} \times volume_{t,i} \tag{9}$$

$$Positive\ money\ flow_{t,i} = \begin{cases} Money\ flow_{t,i}; & Money\ flow_{t,i} > Money\ flow_{t-1,i} \\ 0; & otherwise \end{cases} \tag{10}$$

$$Negative\ money\ flow_{t,i} = \begin{cases} Money\ flow_{t,i}; & Money\ flow_{t,i} < Money\ flow_{t-1,i} \\ 0; & otherwise \end{cases} \tag{11}$$

$$money\ ratio = \frac{\sum_{t=1}^{n} Positive\ money\ flow_{t,i}}{\sum_{t=1}^{n} Negative\ money\ flow_{t,i}} \tag{12}$$

$$MFI = 100 - \frac{100}{1 + money\ ratio} \tag{13}$$

- MFI can become overbought (>70) and prices can simply continue higher when the uptrend is strong [8]. The stock is placed into top class.

– If MFI is between 30 and 70, The stock is placed into neutral class.
– MFI can become oversold (<30) and prices can simply continue lower when the downtrend is strong [8]. The stock is placed into bottom class.

2.2.3 Relative Strength Index (RSI)

RSI is a popular momentum indicator; it represents the strength of the stock based on price fluctuations [8]. RSI is used to provide forecasts for future price fluctuations. An RSI is in the range of 0 to 100, when the RSI is less than 30, giving us a signal from the market that is concentrating on selling [8]. When the RSI is above 70 they give us a signal that the stock is being bought [8].

$$RSI = 100 - \frac{100}{1 + RS} \tag{14}$$

$$Uclose_{t,i} = \begin{cases} close_{t,i}; & close_{t,i} > close_{t-1,i} \\ 0; & otherwise \end{cases} \tag{15}$$

$$Lclose_{t,i} = \begin{cases} close_{t,i}; & close_{t,i} < close_{t-1,i} \\ 0; & otherwise \end{cases} \tag{16}$$

$$RS = \frac{\sum_{t=1}^{n} Uclose_{t,i}}{\sum_{t=1}^{n} Lclose_{t,i}} \tag{17}$$

2.3 Classification Algorithms

2.3.1 Technical Indicator Classifiers

Let $w_{x,t} = \left(w_{x,t}^1, w_{x,t}^2, w_{x,t}^3\right)$ be the vector label of the stock x at time t that

$$w_{x,t}^1 = \begin{cases} 1; & A_{label} = \text{"}top\text{"} \\ 0; & otherwise \end{cases}$$

$$w_{x,t}^2 = \begin{cases} 1; & A_{label} = \text{"}neutral\text{"} \\ 0; & otherwise \end{cases}$$

$$w_{x,t}^3 = \begin{cases} 1; & A_{label} = \text{"}bottom\text{"} \\ 0; & otherwise \end{cases}$$

$$x \in X = \{RSI, MFI, RVI, Bullish, Bearish, Sentiment\ ratio, Sentiment\ trading\ strategies\}$$

Let $\mathcal{A} = \{A_1, A_2, ..A_m\}$ is set of stocks.

Algorithm 1: Technical indicator classifiers

Input: Stock Dataset ($S_{t,A}, w_{RVI,t}, w_{MFI,t}, w_{RSI,t}, N$)

$X_1 = \{RVI, MFI, RSI\}$

Output: $A_{t,label}$

For i=1 **to** N

Using formula (7) **return** RVI_{label}.

If (RVI_{label} ="top") **return** $w_{RVI,t} = (1,0,0)$.

Else if (RVI_{label} ="neutral") **return** $w_{RVI,t} = (0,1,0)$.

Else return $w_{RVI,t} = (0,0,1)$.

Using formula (13) **return** MFI_{label}

If (MFI_{label} ="top") **return** $w_{MFI,t} = (1,0,0)$.

Else if (MFI_{label} ="neutral") **return** $w_{MFI,t} = (0,1,0)$.

Else return $w_{MFI,t} = (0,0,1)$.

Using formula (14) **return** RSI_{label}.

If (RSI_{label} ="top") **return** $w_{RSI,t} = (1,0,0)$.

Else if (RSI_{label} ="neutral**") return** $w_{RSI,t} = (0,1,0)$.

Else return $w_{RSI,t} = (0,0,1)$.

$$w_t = \sum_{x \in X_1} w_{x,t}$$

If($Max_{k=1,2,3}(w_t^k) = w_t^1$) ***then*** $A_{t,label} = "top"$

Else If($Max_{k=1,2,3}(w_t^k) = w_t^2$) ***then*** $A_{t,label} = "neutral"$

Else $A_{t,label} = \text{bottom}.$

Return $A_{t,label}$.

End.

2.3.2 Sentiment Classifiers

Let $Se_{t,A}$ denote the sentiment data value of stock A at time t.

Algorithm 2: Sentiment classifiers.

Input: Stock Dataset ($S_{et}, w_{bull,t}, w_{bear,t}, w_{ratio,t}, w_{STS,t}, N$)

$X_2 = \{Bullish, Bearish, Sentiment\ ratio, STS\}$

Output: A_{label}

For i=1 **to** N

Using formula (2) (3) and return S_{et} sentiment data.

Using sentiment data with formula (4) return Sentiment label.

While ($x \in X_1$)

If ($A_{t,label}$ ="top") **return** $\mathbf{w}_{x,t} = (1,0,0)$.

Else if ($A_{t,label}$ ="neutral") **return** $\mathbf{w}_{x,t} = (0,1,0)$.

Else return $\mathbf{w}_{x,t} = (0,0,1)$.

$$\mathbf{w}_t = \sum_{x \in X_2} \mathbf{w}_{x,t}.$$

If($\mathrm{Max}_{k=1,2,3}(w_t^k) = w_t^1$) **then** $A_{t,label}$ = "top"

Else If($\mathrm{Max}_{k=1,2,3}(w_t^k) = w_t^2$) **then** $A_{t,label}$ = "neutral"

Else $A_{t,label}$ = "bottom".

Return $A_{t,label}$.

End.

2.3.3 Hybrid Classifiers

Algorithm 3: Hybrid classifiers.

Input: Dataset ($w_{x,t}, N$).

Output: A_{label}

For i=1 to N

Using Algorithm 1 return $w^{ind}_{x,t} = w_{x,t},\ x \in X_1,\ A^{ind}_{label,t} = A_{label,t}$.

Using Algorithm 2 return $w^{se}_{x,t} = w_{x,t}, x \in X_1,\ A^{se}_{label,t} = A_{label,t}$.

If ($A^{ind}_{label,t} = A^{se}_{label,t}$) **then return** $A_{label,t}$.

Else {

$$w_t = \sum_{x \in X_2} w_{x,t}.$$

If($Max_{k=1,2,3}(w^k_t) = w^1_t$) then $A_{t,label}$ = "top".

Else If($Max_{k=1,2,3}(w^k_t) = w^2_t$) then $A_{t,label}$ = "**neutral**"

Else $A_{t,label}$ = "**bottom**".

}

Return $A_{t,label}$.

End.

3 Experiments and Results

3.1 Datasets

To evaluate the performance of our proposed hybrid classifier, we perform five stocks of Vietnamese stock market [6]. Vietnamese stock market is considered potential for investors. VN30 stocks being highest capitalization and liquidity in Vietnamese stock market are 80% capital of total market and 60% liquidity of total market [6]. Thus, five

stocks BID (Joint Stock Commercial Bank for Investment and Development of Vietnam), BVH (BAOVIET Holdings), CTG (Vietnam Joint Stock Commercial Bank for Industry and Trade-Main Operation Center), FPT (The Corporation for Financing and Promoting Technology) and VIC (Vingroup joint stock company), are chosen from stocks in the Top 30 stocks to apply hybrid classifier.

3.2 Experiments

The data stock prices of BID, BVH, CTG, FPT and VIC are **considered** in one year from 4th January 2016 to 30th December 2016. Thus, each stock has 252 data, the total data are 1260.

Let $S_{t,i}$ denote the price of the stock i at the day t. The price can be high, close, low, volume that depend on whatever technical indicator is used.

$$\mathcal{A} = \{VIC, FPT, CTG, BID, BVH\}.$$

N = 199.
n in the indicator classifiers is applied $n = n_{ind} = 14$.
n in the sentiment classifiers is applied $n = n_{sen} = 52$.

3.3 Results

The averages error rate and their variance for 3 algorithm are shown below (Table 1). The bold results show the best mean and variance of error rates comparing the three algorithm classifiers. Thus, the empirical results show that in hybrid classifier has not only low average error rate but also low variance of error rate which confirms its outstanding performances comparing with the left two classification methods (Fig. 2 and Tables 2, 3 and 4).

Table 1. Classification error rates of three algorithm classifiers.

	Indicator		Sentiment		Hybrid	
	Mean	Variance	Mean	Variance	Mean	Variance
BID	0.3568	0.5699	0.3568	0.3657	**0.2060**	**0.2160**
BVH	0.3065	0.5867	0.4874	0.5347	**0.2010**	**0.2928**
CTG	0.3719	0.5212	0.5477	0.3772	**0.2814**	**0.2702**
FPT	0.4372	0.7243	0.6884	0.5040	**0.3065**	**0.2915**
VIC	0.3970	0.6747	0.6030	0.5931	**0.2915**	**0.3410**

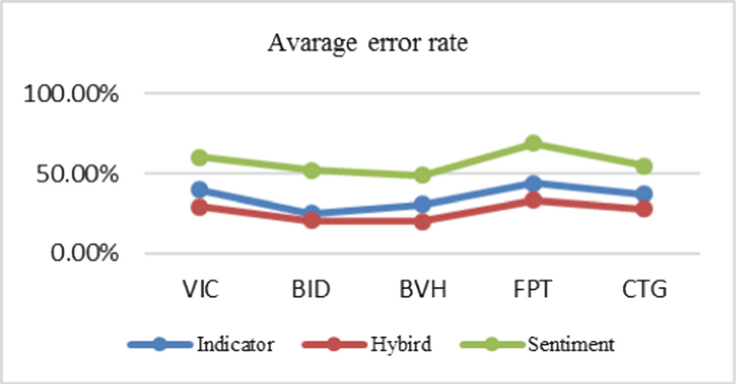


Fig. 2. Percentage of average error rates of three algorithm classifiers.

Table 2. Correct percentage results of sentiment classifier.

Sentiment	Target				
		Top	Neutral	Bottom	
Output	Top	35	157	17	16.75%
	Neutral	109	484	182	62.45%
	Bottom	4	5	2	18.18%
	% Sentiment classifier correct	23.65%	74.92%	1.00%	

Table 3. Correct percentage results of indicator classifier.

Indicator	Target				
		Top	Neutral	Bottom	
Output	Top	34	141	34	16.27%
	Neutral	157	391	227	50.45%
	Bottom	4	4	3	27.27%
	% Indicator classifier correct	17.44%	72.95%	1.14%	

Table 4. Correct percentage results of hybrid classifier.

Hybrid	Target				
		Top	Neutral	Bottom	
Output	Top	60	133	16	28.71%
	Neutral	45	676	54	87.23%
	Bottom	4	4	3	27.27%
	% Hybrid classifier correct	55.05%	83.15%	4.11%	

4 Conclusion and Discussion

Classification in stock market is a potential and a beneficial area of research for investors. The classification of a stock into three classes that are top, neutral and bottom will be supportive for business decision makers. In this paper, historical data in Vietnamese stock market are taken over a period of one years. We attempt to classify five stocks in VN30. Different classifiers of machine learning have been used to classify future trend in these datasets. The empirical results show hybrid classifier stock have more power significantly in-sample test than single technical indicator classifier or sentiment classifier. In fact, hybrid classifier has shown good performance in terms of accuracy and low variance.

In the future work, it would be interesting to apply this classifier on a large group of stocks from different stock markets. Moreover, we consider social media data, fundamental data in addition to historical data in stock market classifier. The new data will improve our classifier results.

References

1. Gao, B., Yang, C.: Forecasting stock index futures returns with mixed-frequency sentiment. Int. Rev. Econo. Finan. **49**, 69–83 (2017). ISSN 1059-0560
2. Ballings, M., Van den Poel, D., Hespeels, N., Gryp, R.: Evaluating multiple classifiers for stock price direction prediction. Expert Syst. Appl. **42**(20), 7046–7056 (2015). ISSN 0957-4174
3. Krawczyk, B., Minku, L.L., Gama, J., Stefanowski, J., Woźniak, M.: Ensemble learning for data stream analysis: a survey. Inf. Fusion **37**, 132–156 (2017). ISSN 1566-2535
4. Lagarde, A.: Can Sentiment Indicators Signal Market Reversals? 27 May 2016. SSRN: https://ssrn.com/abstract=2785462
5. Nguyen, T.T., Nguyen, T.T.T., Pham, X.C., Liew, A.W.-C.: A novel combining classifier method based on Variational Inference. Pattern Recogn. **49**, 198–212 (2016)
6. Hoang, A.L.: Corporate governance impacts to the level of accounting information disclosure – evidence on VN30 listed companies of Vietnam. Eur. J. Account. Audit. Finan. Res. **4**(6), 1–9 (2016)
7. Mehmood, Y., Hanif, W.: Impact of bullish and bearish market on investor sentiment. Int. J. Innov. Appl. Stud. **9**(1), 142–151 (2014). ISSN 2028-9324
8. Khaidem, L., Saha, S., Dey, S.R.: Predicting the direction of stock market prices using random forest. Appl. Math. Finan. (2016)
9. Ozturk, M., Toroslu, I.H., Fidan, G.: Heuristic based trading system on Forex data using technical indicator rules. Appl. Soft Comput. **43**, 170–186 (2016). ISSN 1568-4946

Visualizing Space-Time Map for Bus

Hong Thi Nguyen[1,3(✉)], Diu Ngoc Thi Ngo[2], Tha Thi Bui[3], Cam Ngoc Thi Huynh[3], and Phuoc Vinh Tran[2]

[1] Rubber Industrial College (RIC), Dong Xoai, Binhphuoc, Vietnam
Hongnguyen1611@gmail.com
[2] Thudaumot University (TDMU), Thu Dau Mot, Binhphuoc, Vietnam
{diuntn,Phuoctv}@tdmu.edu.vn, Phuoc.gis@gmail.com
[3] University of Information Technology (UIT), Vietnam National University – HCMC, Ho Chi Minh City, Vietnam
bththa@gmail.com, camhuynhit@gmail.com

Abstract. Current bus maps are created by tracing ground trajectories of bus routes on available base maps. These bus maps provide with the spatial information of ground trajectories and bus stops; some websites have associated departure time table with each ground trajectory on the map. It is considered that bus users need not only spatial but also temporal information. This paper integrates visualization with temporal geography to represent spatio-temporal data of bus network as a visualized-space-time bus map. The visualized-space-time bus map displays visually the movement data of bus as graphs of bus routes and bus trips on screen. The graphs are classified as visual classes according to their attributes. The visual features of the graphs are enhanced by integrating graph classes with retinal variables. The visualized-space-time bus map enables users to analyze visually the spatio-temporal data of bus network for their travels.

Keywords: Visual · Bus map · Spatio-temporal data · Space-time map

1 Introduction

Bus map of a city is means necessary for bus users, not only residents but also tourists. They utilize bus map to design the travels according to their individual demands. Almost current bus maps are created from tourist maps or road maps by tracing additionally bus trajectories, bus stops, and bus stations of routes. On the map, several bus trajectories are drawn densely along with several road segments. The visual features of the map are strongly reduced because of the weakness of selective and associative features with respect to viewers. Moreover, current bus maps do not display the time departing from or arriving at end stations, the time coming to or leaving from bus stops.

We proposed to apply the concept of space-time cube in temporal geography to represent bus network as a space-time map of bus network [1]. However, the limitation of screen on size and resolution makes it thick in displaying bus network. The space-time bus map is quite difficult to use because of its low visual features. In that, the difference between bus routes and each other as well as between bus route and bus trip is quite difficult to be perceived by human vision.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 38–47, 2018.
https://doi.org/10.1007/978-3-319-77818-1_4

The main idea of this paper is to enhance the visual features of space-time bus map by classifying visual graphs as visual classes according to each route to integrate them with retinal variables. Each visual class is composed of route graph (ground trajectory), trip graphs (space-time trajectories), and elements relating route such as the time axis, lines associating time with trip graphs. The visual classes are integrated with retinal variables to differentiate route graphs from each other and route graphs from trip graphs. The integration of visual graphs with retinal variables results in the significant enhancement of the utilization of space-time bus map in applications and analysis.

This paper is structured as follows. The next section refers to the capacity of human vision and visual features of maps in cognizing and understanding visualized maps. The Sect. 3 approaches temporal geography and space-time cube to making space-time bus map. In Sect. 4, visualized-space-time bus map is upgraded from space-time bus map, where only one normalized trip graph is displayed for a route to reduce the thickness of screen and all visual graphs are classified as visual classes according to each route to integrate them with retinal variables. The final section summarizes the results of the work.

2 Visual Features

2.1 The Capacity of Human Vision

The people perceive real world by collecting information through five senses, vision, hearing, touch, taste, and smell. Each sense has an organ to collect data, excite brain, and transmit information to mind. Vision stimulates strongly human perception as well as acquires fast and numerously information [2]. Human vision perceives an object in real world through light rays reflecting from the object to eyes which is an important organ of sight.

Human detects entities in real world by visional perception of light rays reflecting from the entities and focusing on the back of eyes. For two objects of the same size, closer one seems bigger. The visional perception of object size depends on visional angle which is formed by the pencil of light rays from the object to eye. For two objects of the same size, the closer one is perceived to be bigger because its visional angle is bigger. As a result, a smaller object is perceived to be further from eye. The visional resolution is the capacity of human eye for distinguishing two different points; it is also smallest visional angle to differentiate two points. The resolution depends on individual specification of biology of human eye and on the brightness, the luminance, and environment of objects.

The depth of an object perceived by human eye is due to the difference of lengths of light rays from the points on the object to eye. The distinction between the lengths of light rays coming to eye from points of different distance to eye results in the difference of phase of light rays at human eye. Accordingly, human vision can perceive the depth of the object. The specification enables painters to feel the depth in aerial perspective images on plane. The 3D orthogonal coordinate is used in visualization to makes feeling the depth on plane.

Human vision has capacity for perceiving different colors. The perception of color difference depends on brightness, luminance, environment, and biology specification of each person. In visualization, human capacity for distinguishing colors affects the selective feature as the color variable is utilized to represent visual graphs. Similarly, the brightness resolution refers to human capacity for detecting the distinction between different values as the brightness variable is used to represent values.

2.2 The Visual Features

The visual features result from studying the properties of human vision and the requirements of users in viewing visual graphs. In visual analysis, analysts have to answer various questions, elementary and synoptic [3], by their visional perception in viewing visual graphs representing data. The visual features applying for visual graphs enable human to extract information as well as answer questions as far as possible. Bertin proposed four following features to visualize a data set [4].

- *Selective feature:* The selective feature refers to the possibility of human vision distinguishing graphs representing different data variables or graphs representing different value intervals of one or more data variables.
- *Associative feature:* The associative feature refers to the possibility of human vision grouping variables or values of the same category or attribute.
- *Ordered feature:* The ordered feature refers to the possibility of human vision comparing data values of variables to differentiate between big one and small one or arranging them in order as an increasing or decreasing sequence.
- *Quantitative feature:* The quantitative feature refers to the possibility of human vision determining a quantitative value of data.

2.3 The Retinal Variables

Data visualization is a mapping of data onto graph representing visually data significance on planar screen. In that, a visual variable representing a data variable is displayed as an axis on screen, a visual variable representing tuples of relational data variables is shown as a coordinate system. The marks representing values of data or data-tuples constitute visual graphs at positions indicated on the axis or on the coordinate system. Retinal variables are used to enhance visual features of visual graphs. The works by Bertin discovered six retinal variables, shape, size, symbol, brightness, color, and direction (Table 1) [4, 5]. The development of graphic technique provides additionally with new retinal variables of frequency flicker and phase flicker [6].

3 Approaching Temporal Geography to Making Space-Time Bus Map

Since prehistoric period, maps have been created to record by drawing natural things, objects in real world [7]. Map depicts symbolically the natural surface of the globe,

Table 1. Retinal variables

Retinal Variables	Visual Marks
Shape ***S***	
Size ***Z***	
Symbol ***L***	
Brightness ***B***	
Color ***C***	
Direction ***D***	

indicates positions of entities on ground as well as represents the relationship between objects and their activities. In addition of drawings of real world, statistical graphs and data visualization are also utilized on maps to depict the relationship among things as well as between objects and nature. Cartography, statistics, and visualization have been relating to each other in their growth history [8]. The basic theory of cartography has strongly developed since 19th century, from the coordinate system indicating positions of objects to the symbols depicting entities and their relations. Besides that, cartographers classified different kinds of maps, general and thematic. For each type of map, there are common and individual codes to show and display things on map frame of two dimensions, where time is shown as a data table if necessary. The limitation of dimension of traditional map can not show time in the coordinate system.

3.1 Space-Time Map

Space-time map is an application of space-time cube [1, 9–12]. It is constituted by a 3D orthogonal coordinate system, where two axes serve as ground coordinate system indicating ground positions of objects, the rest is used to indicate time (Fig. 1). In that, the ground coordinate system is used to indicate the ground positions based on the geographic coordinate system or a geodetic coordinate system; the time unit suitable for data set is defined on the time axis. In other words, space-time map joins a 2D traditional map with a time axis perpendicular to the plane of the 2D map. With this structure, space-time map has also many types similar to 2D maps.

According to temporal GIS (Geographic Information Science), every object exists at a location on ground associating with a time point or a time interval [13]. A moving object has its ground position changing over time. In reality, each object moves on a continuous ground trajectory recorded by digital instruments as points on ground associating with time. Technically, the ground trajectory of a moving object is a polyline time-ordered connecting ground points recorded by positioning instruments. In that, the positions of moving object are recorded by following ways: time-based, location-based, change-based, event-based, combined [14].

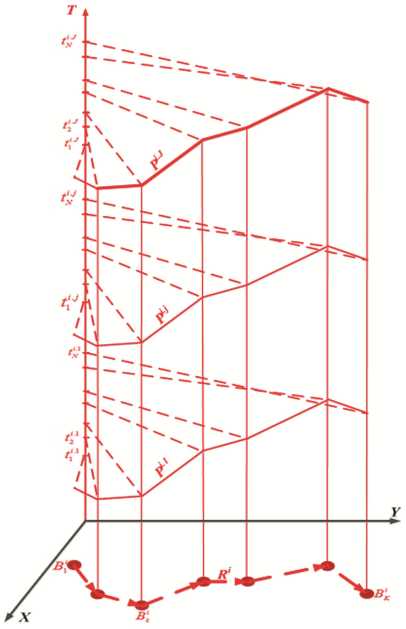


Fig. 1. Space-time bus map for the route i consists of a route graph R^i (ground trajectory) joining bus stops and several trip graphs $P^{i,j}$ of the route i.

3.2 Space-Time Map for Representing Bus Network

Bus network of a city is composed of several routes. Hochiminh city has over 150 bus routes, each of which connects end stations, one is departure and another is arrival, and is coded by a natural number, titled by the names of two end stations. Bus stops along the route are the locations to pick up and drop out passengers. The information of network is shown on 2D map. Traditionally, bus network is shown on maps of street. A lot of digital bus maps are supplied on website by different institutions with various features. Ground trajectories of bus routes are traced along streets of the city base map, where all bus stops from departure to arrival station are joined with trajectories. On the base map, user may select a bus route by clicking at the name or the code of a route on the route table of bus network (http://xe-buyt.com/ban-do-xe-buyt). Another bus map provides user with soft tool to find bus route for an individual travel from departure to arrival location. Some of cases, this bus map can not display any bus route for a selected travel (http://bando24h.com/FindBus.aspxv).

This study proposes to represent bus network on space-time map, where each bus route is shown on the base map of 2D coordinate system indicating ground locations as a ground trajectory of all buses moving on the route. The ground trajectory may be drawn along roads where all busses of the route move, or be simulated as a polyline connecting straight bus stops to each other and to bus stations, called route graph. Hundreds busses move on a route in a day, the movement of a bus is represented as a space-time trajectory on space-time map, called bus trip. The space-time trajectories are simulated as polylines time-ordered

connecting straight spatio-temporal points which are the spatial positions of bus stops associated with time instants when the bus of the trip comes to or leaves from the stops or the stations. A polyline simulating space-time trajectories is called trip graphs.

Route graphs are mathematically represented by the visual variable $X \times Y$ that is a geographic or geodesy coordinate system indicating ground locations. Data of route graphs are tuples $<i, x, y>$, where $i|i = 1, 2, \ldots, I$ is route code, $(x_n^i, y_n^i) \in X \times Y|n = 1, 2, \ldots, N$ is the ground locations of the bus stop n of the route i, $B_n^i|n = 1, 2, \ldots, N$ is the title of the bus stop at $(x_n^i, y_n^i) \in X \times Y$. Trip graphs are mathematically represented by visual variable $X \times Y \times T$, where $X \times Y$ indicates ground locations and T indicates time. Data of a trip graph are tuples $<i, j, x, y, t>$, where $i|i = 1, 2, \ldots, I$ is route code, $j|j = 1, 2, \ldots, J$ is trip code, $(x_n^{i,j}, y_n^{i,j}, t_n^{i,j}) \in X \times Y \times T|n = 1, 2, \ldots, N$ is the space-time location of the trip j of the route i at the of bus stop B_n^i, and $t_n^{i,j}$ is the time when the bus of the trip j of the route i calls at the stop $B_n^i(x_n^i, y_n^i)$.

4 Approaching Temporal Geography and Visualization to Creating Visualized-Space-Time Bus Map

For visualization, the visual features of space-time map of bus network may be improved by reducing the thickness of screen as displaying route and trip graphs and enhancing the visual features of graphs of different categories. In subsequent discussion, let:

R^i: Route graph of the route $i|i = 1, 2, \ldots, I$;
$P^{i,j}$: Trip graph $j|i = 1, 2, \ldots, J$ of the route i;
Q^i: Normalized trip graph of the route i;
T^i: The time axis applying for the route i.

4.1 Reducing the Thickness of Screen

A bus network is constituted to depict hundreds of routes, each of which consists of hundreds of trips. It is impossible to display a lot of graphs on a screen of limited size and resolution. The following reasoning aims at reducing the number of trip graphs of a route. Studying the time table of all bus routes, we discover that all trip graphs of a route are similar; they have the same projection on ground, which is just the graph of the route. Indeed, the trip graphs of a route coincide with each other as they are moved according to the direction of the time axis. The study represents only one trip graph for a route, which is called normalized trip graph of the route.

All trip graphs of a route may be generated from the normalized trip graph of the route. The generation of trip graphs of a route is created according to two ways. First, in order to generate a trip graph at an indicated departure time, user moves the normalized trip graph along the direction of the time axis so that its departure point associates with the departure time of the trip on the time axis. Second, user moves the time axis along itself so that the departure time on the time axis associates with the departure point of

the normalized trip graph. In the second way, each route associates with an individual time axis to make trip graphs and only one trip graph displays for a route at a moment.

The approach of the normalized trip graph reduces significantly the thickness of screen. The number of route graphs does not change but the number of trip graphs reduces from hundreds to one graph for a route. The approach of the normalization of trip graph results in the significant improvement in displaying space-time map of bus network, where there are only two graphs for a route, one for route graph and another for normalized trip graph (Fig. 2).

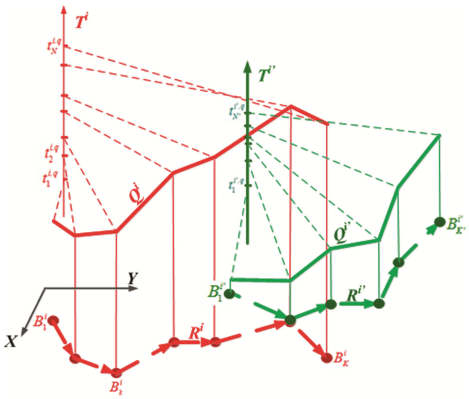


Fig. 2. Visualized-space-time bus map is created with the visual class of the route i and the visual class of the route i'. The visual class of red color represents data of the route i, the visual class of green color represents data of the route i'. The bold dashes are route graphs and thin lines are normalized trip graphs. The round points symbolize bus stops. (Color figure online)

4.2 Creating Visualized-Space-Time Bus Map

Integrating the concept of space-time cube in temporal GIS [9, 13, 15] with visualization for analysis, we constitute visualized-space-time bus map. For a space-time bus map, data relating to bus and its activities are organized on a base map [1]. For a visualized-space-time map, data of bus network are classified according to each route as visual classes, each of which is composed of graphs representing data concerning a route of bus network. Visualized-space-time map is the integration of visual classes of routes with base map. Each visual class comprises (Fig. 2):

- A polyline approximating ground trajectory connects bus stops (including two end stations) located on base map, called route graph.
- A normalized trip graph is space-time trajectory of some bus travelling on the route according to the regulated time table.
- Bus stops and two end stations of the route, commonly called bus stops, are also located on base map.
- Time axis marked linear-time points can be moved along the direction perpendicular to the base map to convert the normalized trip graph to a convenient trip graph.
- The lines associate the time axis with the points on trip graphs

Based on the base map, visual classes may be joined each other by graphic soft tools. For applications of visual analysis of spatio-temporal data of bus, users may show visual classes necessary for their problems of visual analysis according to flexible ways.

4.3 Improving Visual Features of Visualized-Space-Time Bus Map

The approach of visualization is applied for space-time bus map to strengthen the capacity of human vision for perceiving visual graphs and processing bus data. The improvement is studied according to the visual features including selective, associative, ordered, and quantitative. For a route of space-time bus map, route graph, trip graphs and normalized trip graph, and time axis are considered as the elements of the visual class of the route i, $G^i = R^i \cup Q^i \cup T^i$.

Selective feature: The feature of visual selection of visualized-space-time bus map refers to the difference between a graph class and others as well as between route graph and trip graphs of a route. The feature may be enhanced by integrating retinal variables into the graphs, for example:

- For distinguishing between a visual class and others, the color variable is integrated into graph classes. The product is defined as follows.

$$C \times G^i = \{(c_1, G^1), (c_2, G^2), \ldots, (c_I, G^I)\}$$

$$C \times G^i = \{(c_1, R^1), \ldots, (c_I, R^I)\} \cup \{(c_1, Q^1), \ldots, (c_I, Q^I)\} \cup \{(c_1, T^1), \ldots, (c_I, T^I)\}$$

  where $C = \{c_1, c_2, \ldots, c_I\}$ is the color variable, e.g. $C = \{red, green, blue, pink, \ldots\}$.

Technically, the number of colors is limited and the human capacity for distinguishing different colors is also limited. The limitation is made good by only selecting graph classes necessary for application to display simultaneously.

- For distinguishing between route graph, normalized trip graph, and time axis of a route, the symbol variable is integrated into graph class. The product is defined as follows.

$$L \times G^i = \{l_1, l_2, l_3\} \times \{G^1, G^2, \ldots, G^I\}$$

$$L \times G^i = \{l_1, l_2, l_2\} \times \{R^1, R^2, \ldots, R^I\} \cup \{Q^1, Q^2, \ldots, Q^I\} \cup \{T^1, T^2, \ldots, T^I\}$$

$$L \times G^i = \{(l_1, R^1), \ldots, (l_1, R^I)\} \cup \{(l_2, Q^1), \ldots, (l_2, Q^I)\} \cup \{(l_3, T^1), \ldots, (l_3, T^I)\}$$

where $L = \{l_1, l_2, l_3\}$ is the symbol variable, and l_1, l_2, l_3 are symbols, e.g. l_1 is dashes, l_2 is thin lines, l_3 is arrows.

The marks of bus stops are visually distinguished from route graph by the difference of symbols according to the concepts of GIS, where bus stops are represented as objects of point meanwhile route graphs as objects of lines.

Associative feature: The feature of visual association refers to the visual similarity of the graphs of the same category. In above examples, the graphs of a route including route and trip graphs are visually associative by the display of the same color; the route or trip graphs of different routes are visually associative by the display of the same symbol such as line, dashes or arrow (Fig. 2). The marks of bus stops are visually associated with route graph by the color of the graph. In other words, the color of bus stops of a route is similar to the color of its route.

Ordered feature: The feature of visual order refers to the order of trip graphs and bus stops. The order of trip graphs of a route is visually perceived by associating each trip graph with time axis (Fig. 1). The order of bus stops which all busses of a trip call at is cognized by the arrows representing the direction of route graph.

Quantitative feature: The feature of visual quantity refers to the locations of bus stops and the time associating with trip graphs. The quantitative values relating to a bus stop are cognized by lines associating its location with coordinate axes. The quantitative values relating to trip graph are cognized by lines associating trip graph with time axis and route graph.

5 Conclusion

Space-time bus map is formed by the combination of visual variables representing bus data with a base map. The map provides visually with not only ground data such as ground trajectories and bus stops of different routes but also regulated temporal data of bus trips. The map enables to solve the problem of visual analysis of spatio-temporal data concerning bus network in a city, e.g. the approach to visualization method to finding bus routes and trips for a travel with given data of departure and arrival location, and arrival time.

Visualized-space-time bus map is upgraded from space-time bus map by replacing hundreds of trip graphs for a route with only one normalized trip graph and integrating retinal variables into space-time bus map to enhance the visual features of the map. Visualized-space-time bus map is structured with separated graph classes. Each class represents data of a route, including a route graph, a normalized trip graph, bus stops, a time axis, and lines associating time with normalized trip graph. This structure is suitable for storing as well as displaying fast and analyzing flexibly bus data or finding bus travels. The proposed visualized-space-time map may apply for the systems of transportation of regular routes such as metro, train, and so on.

References

1. Nguyen, H.T., Duong, C.K.T., Bui, T.T., Tran, P.V.: Visualization of spatio-temporal data of bus trips. In: Presented at the IEEE 2012 International Conference on Control, Automation and Information Science, ICCAIS 2012, Hochiminh City, Vietnam (2012)
2. Alexandre, D.S., Tavares, J.M.R.S.: Introduction of human perception in visualization. Int. J. Imaging Robot. **4**, 60–70 (2010)
3. Andrienko, N., Andrienko, G.: Exploratary Analysis of Spatial and Temporal Data - A Systematic Approach. Springer, Heidelberg (2006). https://doi.org/10.1007/3-540-31190-4

4. Bertin, J.: Semiology of Graphics (1983)
5. Bertin, J.: General theory, from semiology of graphics. In: Dodge, M., Kitchin, R., Perkins, C. (eds.) The Map Reader. Theories of Mapping Practice and Cartographic Representation, pp. 8–16. Wiley (2011)
6. Green, M.: Toward a perceptual science of multidimensional data visualization: Bertin and beyond. ERGO/GERO Hum. Factors Sci. (1998)
7. Smith, C.D.: Cartography in the prehistoric period in the old world: Europe, the Middle East, and North Africa. In: Harley, J.B., David, W. (eds.) The History of Cartography: Cartography in Prehistoric, Ancient and Mediaeval Europe and the Mediterranean, Chicago, vol. 1, pp. 54–101 (1987)
8. Friendly, M.: Milestones in the history of thematic cartography, statistical graphics, and data visualization (2009)
9. Bach, B., Dragicevic, P., Archambault, D., Hurter, C., Carpendale, S.: A review of temporal data visualizations based on space-time cube operations. In: Eurographics Conference on Visualization (EuroVis), pp. 1–19 (2014)
10. Kristensson, P.O., Dahlbäck, N., Anundi, D., Björnstad, M., Gillberg, H., Haraldsson, J., et al.: An valuation of space time cube representation of spatiotemporal patterns. IEEE Trans. Visual. Comput. Graph. **15**, 696–702 (2009)
11. Kraak, M.J.: The space-time cube revisited from a geovisualization perspective. In: Presented at the 21st International Cartographic Conference (ICC) "Cartographic Renaissance" (2003)
12. Tran, T.V., Tran, P.V., Bui, T.T.: Representing uncertain time on space-time cube for bus movement. In: Presented at the IEEE 2013 International Conference on Control, Automation and Information Science, ICCAIS 2013, Nha Trang, Vietnam (2013)
13. Hagerstrand, T.: What about people in regional science? In: Presented at the Ninth European Congress of Regional Science Association (1970)
14. Andrienko, N., Andrienko, G., Pelekis, N., Spaccapietra, S.: Basic concepts of movement data. In: Giannotti, F., Pedreschi, D. (eds.) Mobility, Data Mining and Privacy, pp. 15–38. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-75177-9_2
15. Andrienko, G., Andrienko, N.: Dynamic time transformation for interpreting clusters of trajectories with space-time cube. In: Presented at the IEEE Symposium on Visual Analytics Science and Technology, Poster (2010)

Generation of Power State Machine for Android Devices

Anh-Tu Bui[1(✉)], Hong-Anh Le[2], and Ninh-Thuan Truong[1]

[1] VNU, University of Engineering and Technology, Hanoi, Vietnam
batu@ictu.edu.vn, thuantn@vnu.edu.vn
[2] Hanoi University of Mining and Geology, Hanoi, Vietnam
lehonganh@humg.edu.vn

Abstract. Power consumption is a major problem on mobile devices. When an application runs, it causes the mobile device to reach a specified state of power consumption. We can determine energy consumption states of mobile devices by analyzing source code of the application. In this paper, we introduce a new approach to modeling energy consumption states due to the impact of Android applications using state machines. The approach takes into account the power states at specific time of the running application. The paper also proposes to construct a finite automata of power states extracted from the source code of the application. We have implemented a plug-in (called PSA) which can be integrated in Android Studio and InteliJ to visualize the finite automata of power states.

Keywords: Formal analysis · Power automata · Power consumption
Mobile devices

1 Introduction

The mobile devices plays an indispensable role in nowadays life. The biggest advantage of these devices is that they are portable thanks to their small size and lightness. Along with the efforts to reduce the size and the weight of devices, their battery capacity must be limited. As a consequence, hardware producers try to find new technologies in order to increase battery capacity, meanwhile software developers try to optimize mobile applications to reduce energy consumption.

Research in estimating the energy usage of mobile devices has investigated in a wide variety of techniques, ranging from specialized hardware, cycle-accurate simulators and operating system level instrumentation. However, these approaches existed their limitations in practice, especially, the ability to early support developers to estimate and calculate energy consumption of apps.

It is clear that making software acting affects strongly to power consuming, such as: software can not able to turn off the screen when people stop using mobile devices, 3G signal transceivers do not turn off when devices are connected Wifi or GPS always acts when devices stay in the same position is the main cause making useless power consuming.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 48–59, 2018.
https://doi.org/10.1007/978-3-319-77818-1_5

When analyzing the effect of control process by software, we realized that statements in program affects powerfully to working mode of devices, it changed power consumption level of devices in a time unit. This changing appeared when statements required hardware components working such as: when the command *mediaplayer.Start()* appeared, speakers of mobile devices changed from **off** to **on** mode and the power consumer level of devices in **on** mode is higher than **off** mode.

In this paper, we propose an approach to generating Power State Machine (PSM) for Android mobile devices. The approach takes into account the power states at specific time of the running application. The paper also proposes to construct a finite automata of power states extracted from the source code of the application. We have also implemented a plug-in (called PSA) which can be integrated in Android Studio and InteliJ to visualize the finite automata of power states.

The remainder of the paper is structured as follows. Section 2 describes the approach of modeling the power consumption level in mobile devices, using formal methods to model and analyze power consumption level. Section 3 gives several algorithms and the implementation of the support tool (PSA). In Sect. 4, we compared our approach with related works. Finally, we concludes the paper with the main point contribution in Sect. 5.

2 Formal Analysis of Android Applications' Power Consumption

In this paper, we consider a model of one mobile device included only popular hardware components [10]: Audio, GPS, LCD Screen, Wifi, 3G Cellular.

This mobile device is installed Android operating system and the application is run on Android operating system. We just assessed the effect of an individual software to power consumption of device as well, we contemporarily ignored the effect of operating system and other softwares.

To follow the effect of software to the power consumption, we focused on researching the effect of source code to each hardware component, analyzed commands which could affect to the performance of hardware, so that we could detect the different consumption level on these hardware components. And finally, we assessed the whole effects of the application on a mobile device.

2.1 Analysis of the Power State in Mobile Applications for a Hardware Component

Each hardware component had the different active state, such as Audio device had 2 power states: On and Off. We defined a power state for a hardware component as follows.

Definition 1. *A power state of a hardware component is the level of energy consumption in a unit of time, corresponding to the activity levels of the hardware.*

Hardware components are controlled by source code in software. These controlled statements would affect to change active state of hardware, they changed power state of hardware in the application.

Example, whenever users want to turn on the music, the program will carry out the command *Start()* of *MediaPlayer* class, and power state of sound generator Audio will change from **off** into **on**. When carrying out command *Stop* of *MediaPlayer* class, the power state of Audio device will change from **on** into **off**, we describe in Fig. 1.

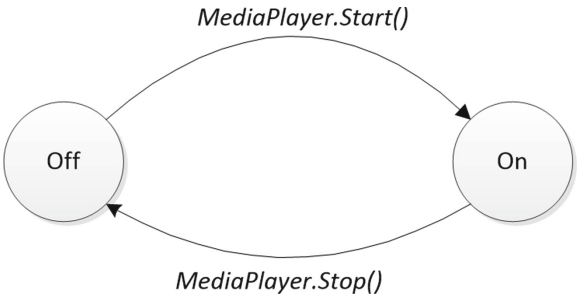


Fig. 1. Power state of Audio device

To model the changing of power state of Audio device, we used a finite automata [4], defined below:

Audio Automata:

$$A_{Audio} = (Q_{Audio}, \Sigma_{Audio}, \delta_{Audio}, q_{0Audio}, F_{Audio})\ [6]$$

where:

$Q_{Audio} = \{off, on\}$
$\Sigma_{Audio} = \{\text{“}Start()\text{”}, \text{“}Stop()\text{”}\}$
$q_{0Audio} = \text{“}off\text{”}$
$F_{Audio} = Q_{Audio}$
δ_{Audio} describles state transition of audio hardware, illustrated in Table 1.

Table 1. State transition table of audio hardware

State-input	Start()	Stop()
Off	On	
On		Off

To do the same actions with other hardware components [2], we give finite automata performing for GPS hardware component as follow:

GPS Automata:

$$A_{GPS} = (Q_{GPS}, \Sigma_{GPS}, \delta_{GPS}, q_{0GPS}, F_{GPS})$$

where:

$Q_{GPS} = \{off, idle, on\}$
$\Sigma_{GPS} = \{$"$PutExtra(String, true)$", "$PutExtra(String, false)$", $RequestLocationUpdates()$"$\}$
$q_{0GPS} =$ "off"
$F_{GPS} = Q_{GPS}$
δ_{GPS} is described by the transition Table 2.

Table 2. State transition table of GPS hardware

State-input	PutExtra(String, true)	PutExtra(String, false)	RequestLocation Updates()
Off	Idle		
Idle		Off	On
On		Off	

LCD Automata:

$$A_{LCD} = (Q_{LCD}, \Sigma_{LCD}, \delta_{LCD}, q_{0LCD}, F_{LCD})$$

where:

$Q_{LCD} = \{off, on\}$
$\Sigma_{LCD} = \{$"$LockNow()$", "$Acquire()$"$\}$
$q_{0LCD} =$ "off"
$F_{LCD} = Q_{LCD}$
δ_{LCD} is describle by a transition Table 3.

Table 3. State transition table of LCD hardware

State-input	LockNow()	Acquire()
Off		On
On	Off	

3G Cellular [8] *Automata:*

$$A_{3GCellular} = (Q_{3GCellular}, \Sigma_{3GCellular}, \delta_{3GCellular}, q_{03GCellular}, F_{3GCellular})$$

where:

$Q_{3GCellular} = \{off, idle, on\}$
$\Sigma_{3GCellular} = \{$"$SetMobileDataEnabled(true)$", "$SetMobileDataEnabled(false)$", $Execute(httpget)$"$\}$
$q_{03GCellular} =$ "off"
$F_{3GCellular} = Q_{3GCellular}$
$\delta_{3GCellular}$ is describle by a transition Table 4.

Table 4. State transition table of 3G cellular hardware

State-input	SetMobileData Enabled(true)	SetMobileData Enabled(false)	Execute(httpget)
Off	Idle		
Idle		Off	Transmitting
Transmitting		Off	

Wifi Automata:

$$A_{Wifi} = (Q_{Wifi}, \Sigma_{Wifi}, \delta_{Wifi}, q_{0Wifi}, F_{Wifi})$$

where:

$Q_{Wifi} = \{off, idle, on\}$
$\Sigma_{Wifi} = \{$"$SetWifiEnabled(false)$", "$SetWifiEnabled(true)$", $Execute(httpget)$"$\}$
$q_{0Wifi} =$ "off"
$F_{Wifi} = Q_{Wifi}$
δ_{Wifi} is describle by a transition Table 5.

Table 5. State transition table of Wifi hardware

State-input	SetWifi Enabled(false)	SetWifi Enabled(true)	Execute(httpget)
Off		Idle	
Idle	Off		Transmitting
Transmitting	Off		

2.2 Analysis Power State in Mobile Applications

Considering all hardware components in mobile devices, we realized that each hardware component had a defined power state in a certain time. Such as, in any

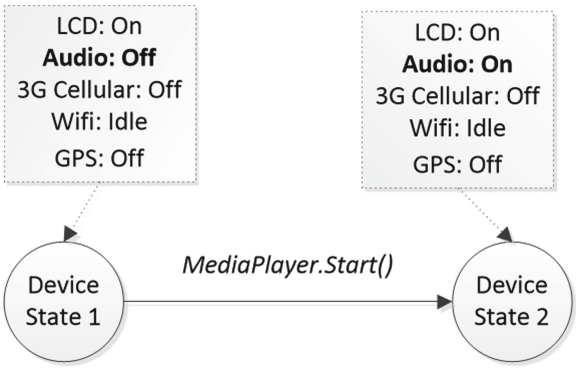


Fig. 2. Power state of mobile application was changed when Audio turn on.

time, LCD Screen was **on**, GPS was **off**, Audio was **on**), Wifi was **idle**, and 3G Cellular was **off**, the power consumption level would be defined and unchanged quantity, called the power state for mobile device.

Definition 2. *A power state of a mobile application (PS_{App}) is a combination of single power states of the hardware components.*

$$PS_{App} = (PS_{LCD}, PS_{GPS}, PS_{Audio}, PS_{wifi}, PS_{3GCellular})$$

The power state of an application in above example (PS_{App}) is:

$$PS_{App} = (on, off, on, idle, off)$$

When there was any hardware component changing state by the effect of program source code, device would change from current power state into other power state. Example, when users turned on the music, power state of device would be describe through Fig. 2.

To specify the whole power states of mobile device, we plan to combine all of states of hardware components, calculate all of the cases that can affect to device. With the recommended method, in the case of having 5 hardware components: LCD, Audio, GPS, 3G, Wifi, we could find out 108 power states of device. These are the power state of mobile device that may appear in applications. To generally model the power state in mobile application, we built a generate automata to perform power states by merging component automata, as below:

$$A = (Q, \Sigma, \delta, q_0, F)$$

Conversion graph is built according to the Algorithm 1.

Algorithm 1. Merge 5 single-automatas

Input:

$A_{audio} = (Q_{audio}, \Sigma_{audio}, \delta_{audio}, q_{0\_audio}, F_{audio})$
$A_{GPS} = (Q_{GPS}, \Sigma_{GPS}, \delta_{GPS}, q_{0\_GPS}, F_{GPS})$
$A_{LCD} = (Q_{LCD}, \Sigma_{LCD}, \delta_{LCD}, q_{0\_LCD}, F_{LCD})$
$A_{cellular} = (Q_{cellular}, \Sigma_{cellular}, \delta_{cellular}, q_{0\_cellular}, F_{cellular})$
$A_{wifi} = (Q_{wifi}, \Sigma_{wifi}, \delta_{wifi}, q_{0\_wifi}, F_{wifi})$

Output:

$A = (Q, \Sigma, \delta, q_0, F)$

1: $Q = \{q | q = (q_{audio}, q_{GPS}, q_{LCD}, q_{cellular}, q_{wifi})\}$
2: $\Sigma = \Sigma_{audio} \bigcup \Sigma_{GPS} \bigcup \Sigma_{LCD} \bigcup \Sigma_{Celluar} \bigcup \Sigma_{wifi}$
3: $q_0 = (q_{0\_audio}, q_{0\_GPS}, q_{0\_LCD}, q_{0\_cellular}, q_{0\_wifi})$
4: $F = Q$ ▷ δ is calculated by following algorithm:
5: **for each** $\{q_{audio} \rightarrow aq_{1\_audio}\} \in \delta_{audio}$ **do**
6: **for each** $(q_{audio}, q_{GPS}, q_{LCD}, q_{cellular}, q_{wifi}) \in Q$ **do**
7: **for each** $(q_{1\_audio}, q_{GPS}, q_{LCD}, q_{cellular}, q_{wifi}) \in Q$ **do**
8: $\delta = \delta \bigcup \{(q_{audio}, q_{GPS}, q_{LCD}, q_{cellular}, q_{wifi}) \rightarrow a(q_{1\_audio}, q_{GPS}, q_{LCD}, q_{cellular}, q_{wifi})\}$
9: **end for**
10: **end for**
11: **end for**

2.3 Optimize Power Consumption Model for Each Specific Application

In the Subsect. 2.2, we introduced general power model for a mobile device performed by finite automata, however, when we work with a particular automata we realize that some of them were not appeared.

They could not appear because there was no control statement in source code of program thus hardware components could not perform corresponding power states. Such as, if there was no Start() command, *on* power states of Audio automata might not appear, this means some of the power state on general automata for device could not be able to appeared.

Therefore, to exactly perform power state for a particular application, we must omit power states that could not be able to appeared. To do this, we have had two ways to perform the combination of component automata models:

- Optimize all component automata before combining them into general automata.
- Combine all component automata and then optimize power states that could not be able to appeared.

In this paper, we choose the first way because it can reduce execution performance. We optimize component automata before combining them into a general automata using the Algorithm 1 presented in Subsect. 2.2. We also used the Algorithm 2 to optimize a single automata in which the input of the algorithm consisted of a general automata and a group of statements chosen from source code that we anticipated that they could affect to power state.

Algorithm 2. Simplify automata by removing useless symbols

Input:

$A = (Q, \Sigma, \delta, q_0, F)$

$\Sigma' = \{a \mid a \text{ is a statement in the program}\}$

Output:

$A' = (Q', \Sigma', \delta', q_0', F')$

$q_0' = q_0$
$Q' = \emptyset$
$newQ = q_0 \bigcup \{q \mid q_0 \rightarrow aq, \{q_0 \rightarrow aq\} \in \delta, \forall a \in \Sigma'\}$
$\delta' = \{q_0 \rightarrow aq, \{q_0 \rightarrow aq\} \in \delta, \forall a \in \Sigma'\}$
while $(Q' \neq newQ)$ **do**
    $Q' = newQ$
    for each $q_1 \in Q'$ **do**
        for each $a \in \Sigma'$ **do**
            if $\{q_1 \rightarrow aq\} \in \delta$ **then**
                $newQ = newQ \bigcup \{q\}$
                $\delta' = \delta' \bigcup \{q_1 \rightarrow aq\}$
            end if
        end for
    end for
end while
$Q' = newQ$
$F' = Q'$

3 Implementation

Recent programming tools, such as Android Studio or IntelliJ allows programmers to analyze commands about syntax, data however it is not supporting programmers in analyzing effect of power states in a program. By inspecting the states, programmers can realize high power consumption states and adjust statement accordingly.

For that reason, we built a Plug-in running on Android Studio and IntelliJ called PSA. PSA used JavaParse to analyze all source code in software projects, it selected statements that could change state of device to optimizing component automata. The overview of PSA Plugin is presented in Fig. 3.

PSA analyze source code of a program and visualize them in a state transition diagram. In the Fig. 4, PSA visualizes Wifi automata for an mobile application.

After optimizing all component automata, PSA combined these automatas into general automata (Fig. 5).

With this state diagram, programmers could observe occurred power state by commands, as a consequences, they could adjust statements in the program to be the most suitable. Each time of adjusting source code, PSA would analyze source code again, create new automata to analyse effects of source code to power state of mobile applications.

Based on assessing the power consumption level in a time unit for every power state of each hardware component [5], we got general power consumption

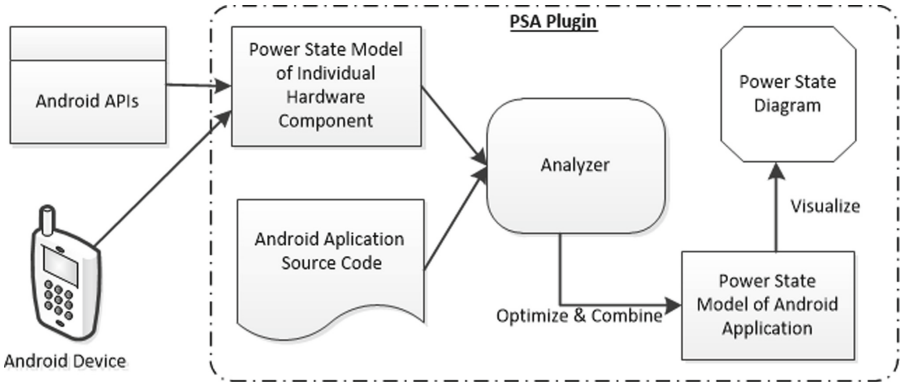


Fig. 3. Overview of PSA Plugin

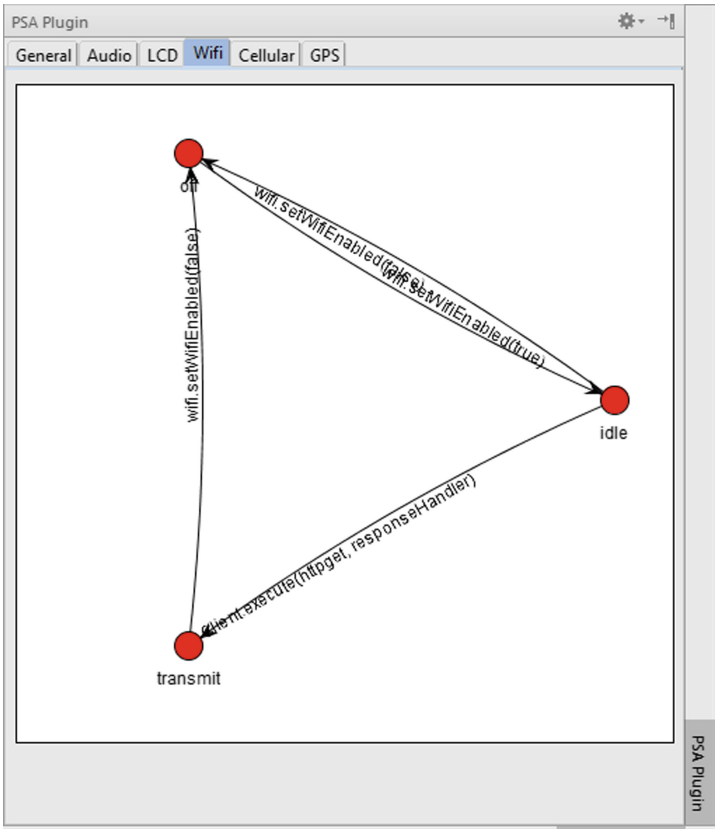


Fig. 4. Visualization of the Wi-fi automata

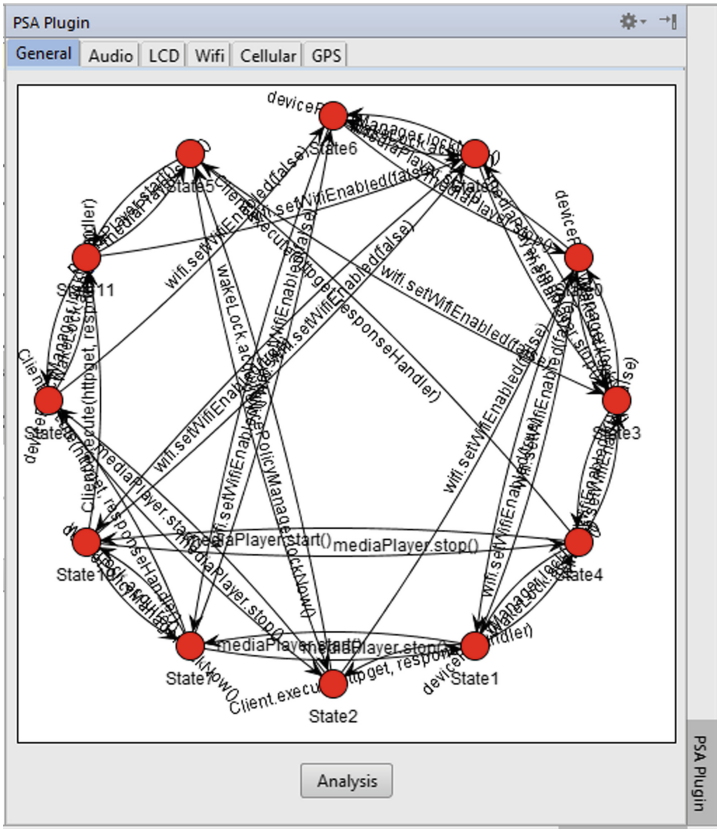


Fig. 5. Visualization of general automata

level in one time unit. This helped programmers know what power state wasted power to quickly adjust the commands, and open the researching trend about power consumption [3] later.

4 Related Work

Nakajima [9] proposed a model-based approach to the representation and analysis of the asynchronous power consumption of Android applications. They introduce a formal model, the power consumption automaton (PCA), show how the PCA is analyzed with existing tools and present some discussions based on their experience.

The paper [7] proposed an approach to estimating power consumption level by analyzing command lines. The paper introduced Elens, a tool used to visualize power consumption level on each command line. This approach permitted calculating power consumption level for command lines in a specific application

however it did not permit analyzing and checking the power constrains in general cases.

Lide Zhang [5] proposed an approach that is both lightweight in terms of its developer requirements and provides fine-grained estimates of energy consumption at the code level. It achieves this using a novel combination of program analysis and per-instruction energy modeling. The approach also provides useful and meaningful feedback to developers that helps them to understand application energy consumption behavior.

Carroll [1] presented a detailed analysis of the power consumption of the Openmoko Neo Freerunner mobile phone. They measure not only overall system power, but the exact breakdown of power consumption by the device's main hardware components. The paper proposed this power breakdown for micro-benchmarks as well as for a number of realistic usage scenarios. These results are validated by overall power measurements of two other devices: the HTC Dream and Google Nexus One. They develop a power model of the Freerunner device and analyse the energy usage and battery lifetime under a number of usage patterns.

In this paper, we work about the generation of Power State Machine for Android devices and then we provide another approach to analysis power consumption of these devices.

5 Conclusion

By assessing the effect of program statements in controlling hardware components of mobile devices, we determine the power states and the corresponding between statements in source code and the changing power states of hardware components. Thus, we built each automata model represented the power states of each hardware component in mobile devices. After finished optimizing these component automata, we combined component automata models to form a general one for a mobile application.

We have built a tool with the aim of supporting programmers on assessing the effect of source code to power consumption of mobile device called PSA. This tool analyse automatically source code and visualize exactly the power state model for each hardware component in mobile application.

The power model enabled programmers to observe the application's power states. We can extend this work by considering the power consumption level and examining the constraints about power consumption over the time.

References

1. Carroll, A., Heiser, G., et al.: An analysis of power consumption in a smartphone. In: USENIX Annual Technical Conference, Boston, MA, vol. 14, p. 21 (2010)
2. Couto, M., Carção, T., Cunha, J., Fernandes, J.P., Saraiva, J.: Detecting anomalous energy consumption in android applications. In: Quintão Pereira, F.M. (ed.) SBLP 2014. LNCS, vol. 8771, pp. 77–91. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11863-5_6

3. Datta, S.K., Bonnet, C., Nikaein, N.: Android power management: current and future trends. In: 2012 First IEEE Workshop on Enabling Technologies for Smartphone and Internet of Things (ETSIoT), pp. 48–53. IEEE (2012)
4. Grzes, T.N., Solov'ev, V.V.: Minimization of power consumption of finite state machines by splitting their internal states. J. Comput. Syst. Sci. Int. **54**(3), 367–374 (2015)
5. Hao, S., Li, D., Halfond, W.G.J., Govindan, R.: Estimating mobile application energy consumption using program analysis. In: 2013 35th International Conference on Software Engineering (ICSE) (2013)
6. Hopcroft, J.E., Motwani, R., Ullman, J.D.: Introduction to Automata Theory, Languages, and Computation, 3rd edn. Addison-Wesley, Boston (2006)
7. Li, D., Hao, S., Halfond, W.G.J., Govindan, R.: Calculating source line level energy information for Android applications. In: Proceedings of the 2013 International Symposium on Software Testing and Analysis - ISSTA 2013, p. 78 (2013)
8. Mendonça, J., Lima, R., Andrade, E., Callou, G.: Assessing performance and energy consumption in mobile applications. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 74–79. IEEE (2015)
9. Nakajima, S.: Model-based power consumption analysis of smartphone applications. In: ACESMB@ MoDELS (2013)
10. Zhang, L., Dick, R.P., Morley Mao, Z., Wang, Z.: Accurate online power estimation and automatic battery behavior based power model generation for smartphones, Ann Arbor

Modeling Self-adaptation - A Possible Endeavour?

Emil Vassev(✉)

Lero–The Irish Software Research Centre, University of Limerick, Limerick, Ireland
emil.vassev@lero.ie

Abstract. Self-adaptive systems have the capability to autonomously modify their behavior at runtime in response to changes in their internal structure or execution environment. Therefore, often self-adaptation emerges as a means to solve problems related to performance or security, to increase efficiency, or to react to various hazards. Basically, self-adaptation may emerge to solve a whole spectrum of various problems or hazards occurring in the execution environment, which implies that behavior modeling for self-adaptation requires intrinsic knowledge of the system context.

A new approach to modeling self-adaptation compliant to system goals is presented in this paper. In this approach, KnowLang, a knowledge representation language for self-adaptive systems, is used to model self-adaptive behavior. Special KnowLang policies are at the core of this approach. Ideally, KnowLang policies are specified to handle specific situations by pursuing a specific goal. A policy exhibits a behavior via actions generated in the environment or in the system itself. Specific probabilistic beliefs and generic conditions determine what specific actions shall be executed. Context properties are intrinsically embedded in the self-adaptive behavior, which makes that behavior context-reactive. To demonstrate the novelty of this approach, the paper elaborates on a self-adaptive behavior of an autonomous vehicle modeled with KnowLang.

Keywords: Self-adaptation · Autonomy · Knowledge representation
KnowLang

1 Introduction to Self-adaptation

The idea of a system that evolves and autonomously finds a solution to a problem, or suggests new ways to solve a problem, is both visionary and very challenging. Without any doubt, the term "adaptive" identifies one of the most challenging topics we currently explore in technology. It identifies systems with the property of being able to autonomously react to situations occurring during its lifetime. The question arises as to whether such a behavior is feasible, implementable, or even desirable. Note that self-adaptive systems must be aware of their physical environment and whereabouts, as well as of their current internal status. This ability enables software intensive systems to sense, draw inferences, and react by exhibiting self-adaptation.

A common understanding about the process of self-adaptation is the ability of a system to autonomously monitor its behavior and eventually modify the same according to changes in the operational environment, or in the system itself. The paradigm requires

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 60–68, 2018.
https://doi.org/10.1007/978-3-319-77818-1_6

that the system engages in various interactions where important structural and dynamic aspects of the environment are perceived. Therefore, it is of major importance for a self-adaptive system to be able acquire and structure comprehensive knowledge in such a way that it can be effectively and efficiently processed, so such a system becomes aware of itself and its environment.

One of the biggest concerns related to self-adaptive systems is how to prove that the autonomous self-adaptive behavior will not cause more safety hazards. Autonomous cars have already appeared on our streets and unfortunately due to some severe accidents they appear to be not as secure as we had hoped them to be. This paper tackles the question of achieving self-adaptive behavior through knowledge representation and awareness reasoning, at least in a certain context, with maximized safety guarantees that will help us to establish trust in autonomous systems.

2 Correct Self-adaptation

Self-adaptation is often related to non-determinism and thus, its correctness proof, if even possible, is a tedious task. As we have discussed before, 100% safety is not possible [1]. A possible approach could be to work on *probabilistic guarantees* through probabilistic model checking, which is considered a powerful technique for formally verifying quantitative properties of systems that exhibit stochastic behavior [2]. Probabilistic behavior may arise, for example, due to failures of unreliable components, a dynamic environment, etc. The problem is that non-determinism leads to *unforeseen behavior*, which basically, cannot be model-checked. It can be eventually simulated (to some extent) through a random generation of the simulated conditions and verified via testing.

Lero, the Irish Software Research Center is currently tackling a project where a special test bed (Test Bed for Adaptive Systems, or TBAS) is under R&D. TBAS targets testing of self-adaptive systems under simulated conditions in both virtual and physical testing environments. With TBAS we shall be able to efficiently test adaptive behavior by validating self-* objectives through evaluation of the system's ability to perceive both the internal and external environments and react to changes. With TBAS, we target the evaluation of features that manifest the system's awareness about situations and conditions, and the system's ability to self-adapt to those situations and conditions when adaptation is required. The foundation of TBAS is the KnowLang Framework [3] and based on this we are developing two test platforms:

(1) A fully virtual simulation environment (virtual TBAS) where multiple virtual adaptive entities (VAEs) can be tested both individually or/and as an "intelligent swarm" whereby VAEs interact not only with the environment, but also internally. Each VAE incorporates a KnowLang Reasoner along with a knowledge base (KB) operated by that reasoner. The virtual TBAS runs as a transparent distributed system comprising multiple interconnected machines, each capable of running hundreds of VAEs.
(2) A test platform based on WiFi/Bluetooth-communicating, mini-computerized and robotized platforms, yet capable of running a fully-functional VAE. Such a robotized platform (Lero Robotics Platform - LRP) is autonomously controlled by the

hosted VAE and is equipped with a GPS and a variety of plug-in sensors such as: light detectors, microphones, smoke detectors, motion detectors, humidity detectors, high-speed thermometer, barometer, etc. LRP is to be designed to be pluggable into different motion platforms (wheels-based, propeller-based, and water-motion-based) and actuator platforms (e.g., mechanical arms).

3 Modeling Self-adaptation with KnowLang

KnowLang employs special knowledge structures and a reasoning mechanism for modeling autonomic self-adaptive behavior [4]. Such a behavior can be expressed via KnowLang policies, events, actions, situations and relations between policies and situations (see Eqs. 1 through 10). Policies (Π) are at the core of autonomic behavior. A policy π has a goal (g), policy situations (Si_π), policy-situation relations (R_π), and policy conditions (N_π) mapped to policy actions (A_π) where the evaluation of N_π may eventually (with some degree of probability) imply the evaluation of actions (denoted $N_\pi \overset{[Z]}{\rightarrow} A_\pi$) (see Eq. 6). A condition (n) is a Boolean expression over an ontology (see Eq. 2), e.g., the occurrence of a certain event. Policy situations Si_π are situations (see Eq. 7) that may trigger (or imply) a policy π, in compliance with the policy-situations relations R_π (denoted by $Si_\pi \overset{[R_\pi]}{\rightarrow} \pi$), thus implying the evaluation of the policy conditions N_π (denoted by $\pi \rightarrow N_\pi$) (see Eq. 6). Therefore, the optional policy-situation relations (R_π) justify the relationships between a policy and the associated situations (see Eq. 10).

$$\begin{aligned} &\Pi := \{\pi_0, \pi_1, \ldots, \pi_m\}, m \geq 0 \quad (policies) \\ &A_\pi \subset A \qquad (A_\pi - policy\ actions; A - the\ set\ of\ all\ actions) \\ &Si_\pi \subset Si \qquad (Si_\pi - policy\ situations) \\ &R_\pi \subset R \qquad (R_\pi - policy - situation\ relations) \end{aligned} \tag{1}$$

$$n := be(O)\ (Boolean\ expression\ over\ ontology) \tag{2}$$

$$N_\pi := \{n_0, n_1, \ldots, n_k\}, k \geq 0\ (policy\ conditions) \tag{3}$$

$$s := be(O)\ (state) \tag{4}$$

$$g := \langle \Rightarrow s' \rangle | s \langle \Rightarrow s' \rangle\ (goal) \tag{5}$$

$$\begin{aligned} &\pi := < g, Si_\pi, [R_\pi], N_\pi, A_\pi, map(N_\pi, A_\pi, [Z]) > \quad (policy) \\ &N_\pi \overset{[Z]}{\rightarrow} A_\pi \qquad (N_\pi implies\ the\ evaluation\ of\ actions\ A_\pi) \\ &Si_\pi \overset{[R_\pi]}{\rightarrow} \pi \rightarrow N_\pi \qquad (Si_\pi\ trigger\ \pi) \end{aligned} \tag{6}$$

$$Si := \{si_0, si_1, \ldots, si_n\}, n \geq 0\ (situations) \tag{7}$$

$$
\begin{aligned}
&si: =< s, A_{si}^{\leftarrow}, \left[E_{si}^{\leftarrow}\right], A_{si} > \quad (situation)\\
&A_{si}^{\leftarrow} \subset A^{*} \quad \left(A_{si}^{\leftarrow} - executed\ actions\right)\\
&\qquad\qquad (A^{*} - the\ set\ of\ all\ finite\ sequences\ with\ elements\ in\ A)\\
&A_{si} \subset A \quad \left(A_{si} - posible\ actions\right)\\
&E_{si}^{\leftarrow} \subset E^{*} \quad \left(E_{si}^{\leftarrow} - situation\ events\right)\\
&\qquad\qquad (E^{*} - the\ set\ of\ all\ finite\ sequences\ with\ elements\ in\ E)
\end{aligned} \tag{8}
$$

$$
R: = \left\{r_0, r_1, \ldots., r_n\right\}, n \geq 0 \ (relations) \tag{9}
$$

$$
\begin{aligned}
&r: =< \pi, [rn], [Z], si > \quad (relation)\\
&si \in Si,\ \pi \in \Pi,\ si \overset{[Z]}{\rightarrow} \pi
\end{aligned} \tag{10}
$$

Note that in order to allow for self-adaptive behavior, relations must be specified to connect policies with situations over an optional probability distribution (Z) where a policy might be related to multiple situations and vice versa. Probability distribution (Z) is provided to support probabilistic reasoning and to help the reasoner to choose the most probable situation-policy "pair". Thus, we may specify a few relations connecting a specific situation to different policies to be undertaken when the system is in that particular situation and the probability distribution over these relations (involving the same situation) should help the reasoner decide which policy to choose (denoted by $si \overset{[Z]}{\rightarrow} \pi$) (see Eq. 10). Hence, the presence of probabilistic beliefs (Z) in both mappings and policy relations justifies the probability of policy execution, which may vary with time.

Ideally, KnowLang policies are specified to handle specific situations, which may trigger the application of policies. A policy exhibits a behavior via actions generated in the environment or in the system itself. Specific conditions determine which specific actions (among the actions associated with that policy (see Eq. 6) shall be executed. These conditions are often generic and may differ from the situations triggering the policy. Thus, the behavior not only depends on the specific situations a policy is specified to handle, but also depends on additional conditions. Such conditions might be organized in a way allowing for synchronization of different situations on the same policy. When a policy is applied, it checks what particular conditions N_{π} are met and performs the mapped actions A_{π} ($map\left(N_{\pi}, A_{\pi}, [Z]\right)$) (see Eq. 6). An optional probability distribution Z may additionally restrict the action execution. Although specified initially, the probability distribution at both mapping and relation levels is recomputed after the execution of any involved action. The re-computation is based on the consequences of the action execution, which allows for reinforcement learning.

4 Case Study

Obviously, self-adaptive and autonomous systems that replace human beings should be carefully designed in concern with safety risks stemming from the autonomous behavior.

For example, autonomous vehicles should be designed towards maximizing the safety guarantee that no pedestrian would ever be injured while operating in autonomous mode.

4.1 Modeling Self-adaptation that Adds on Safety

In this case study, we used KnowLang to model and specify self-adaptive behavior that adds on that safety guarantee. To do so, we determined multiple situations that can be considered critical because they involve an autonomous vehicle in close proximity to pedestrians. For example, such situations are "approaching a crosswalk", "passing through a school zone", "crossing uncontrolled intersection", "approaching a failed traffic light", "passing a stopped vehicle", "approaching a car accident", etc. The following is an example of specifying with KnowLang the "approaching a crosswalk" situation (see Fig. 1).

```
CONCEPT_SITUATION ApproachingCrosswalk {
....
  SPEC {
    SITUATION_STATES
{eMobility.eCars.CONCEPT_TREES.Route.STATES.InCloseDistanceToCrosswalk}
    SITUATION_ACTIONS  {
      eMobility.eCars.CONCEPT_TREES.SlowDown,
      eMobility.eCars.CONCEPT_TREES.StopCar,
      eMobility.eCars.CONCEPT_TREES.UseEngineBreaking,
      eMobility.eCars.CONCEPT_TREES.DenySpeeding,
      eMobility.eCars.CONCEPT_TREES.TurnSteeringWheelRight,
      ....
    }
    ....
  }
}
```

The sample above is a simple specification where for clarity some details are missing, but the reader can conclude that the specified situation ApproachingCrosswalk is determined by the InCloseDistanceToCrosswalk state (specified somewhere else in the specification model) and by actions that can be undertaken once the autonomous vehicle has ended up in this situation (see Eq. 8 in Sect. 3). Note that the InCloseDistanceToCrosswalk state is specified as a Boolean expression (not shown here) that determines if the car enters a section of the followed route where a crosswalk is in close proximity.

In the next step, we formalized self-adaptive behavior in the form of policies (Π) (see Eq. 1 in Sect. 3) driving the autonomous vehicles in this situation. For example, we specified a policy that handles this situation in various conditions emphasizing damages or malfunction of the driving system, e.g., flat tires, malfunctioning steering wheel, malfunctioning brakes, etc. Hence, one possible self-adaptive behavior that emerged from this exercise as adding on car safety can be described as "*automatically deny car*

speeding, turning steering wheel right and stopping the car in the case of flat tire when the car is getting in close proximity to a crosswalk".

```
CONCEPT_POLICY SafeDriveAroundCrosswalk {
....
  SPEC {
    POLICY_GOAL {eMobility.eCars.CONCEPT_TREES.SafeCrosswalkPassing}
    POLICY_SITUATIONS
{eMobility.eCars.CONCEPT_TREES.ApproachingCrosswalk}
    POLICY_RELATIONS {eMobility.eCars.RELATIONS.Situation_Policy_1}
    POLICY_ACTIONS {
      eMobility.eCars.CONCEPT_TREES.SlowDown,
      eMobility.eCars.CONCEPT_TREES.StopCar,
      eMobility.eCars.CONCEPT_TREES.UseEngineBreaking,
      eMobility.eCars.CONCEPT_TREES.DenySpeeding,
      eMobility.eCars.CONCEPT_TREES.TurnSteeringWheelRight,
      ....
    }
    POLICY_MAPPINGS {
      MAPPING {
        CONDITIONS {eMobility.eCars.CONCEPT_TREES.Vehicle.STATES.FlatTire}
        DO_ACTIONS {
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarDenySpeeding,
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarTurnSteeringWheelRight,
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarStopCar }
        PROBABILITY {0.5}
      }
      MAPPING {
        CONDITIONS {eMobility.eCars.CONCEPT_TREES.Vehicle.STATES.FlatTire}
        DO_ACTIONS {
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarStopCar,
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarDenySpeeding,
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarTurnSteeringWheelRight }
        PROBABILITY {0.3}
      }
      MAPPING {
        CONDITIONS {eMobility.eCars.CONCEPT_TREES.Vehicle.STATES.FlatTire}
        DO_ACTIONS {
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarTurnSteeringWheelRight,
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarStopCar,
eMobility.eCars.CONCEPT_TREES.Vehicle.FUNCS.CarDenySpeeding }
        PROBABILITY {0.2}
      }
....
    }
....
  }
}
```

Fig. 1. Autonomous vehicles approaching a crosswalk.

As specified, the probability distribution gives an initial designer's preference about what actions should be executed if the system ends up in running the SafeDriveAroundCrosswalk policy. Note that at runtime, the KnowLang reasoner maintains a record of all the action executions and re-computes the probability rates every time when a policy has been applied and consecutively, actions have been executed. Thus, although the system will initially execute the sequence of actions CarDenySpeeding, CarTurnSteeringWheelRight, and CarStopCar (it has the higher probability rate of 0.5), if that policy cannot achieve satisfactory its SafeCrosswalkPassing goal with this sequence of actions, then the probability distribution will be shifted in favor of another sequence, which might be executed the next time when the system will try to apply the same policy. Therefore, probabilities are recomputed after every action execution, and thus the behavior changes accordingly.

4.2 Testing the Self-adaptation Model

As part of this exercise, tests were performed with the Lero's virtual TBAS (see Sect. 2) to simulate awareness emerging when the KnowLang Reasoner operates over the specified model for self-adaptation. The exercise was performed with the initial version of TBAS where a special host application ran the KnowLang Reasoner and the communication to it went through a command line where the reasoner was fed with simulated conditions. Note that the KnowLang Reasoner iterates over an *awareness control loop* [5] where all the states expressed in KnowLang are evaluated at any loop iteration, which leads to re-evaluation of all the goals and situations expressed with states. Therefore,

when a situation is determined through the evaluation of its states, eventually, a policy will be applied to tackle this situation.

In this test, we simulated the ApproachingCrosswalk situation along with the car's FlatTire state (see Sect. 4.1). Further, by providing feedback to the reasoner from the actions' realization, we enforced self-adaptation by switching the sequence of actions due to the probability re-distribution caused by this feedback. Therefore, the reasoner attempted to find the safest sequence of actions in case of flat tire. More simulated conditions, e.g., rain, snow, ice, etc., helped to determine that the safest sequence of actions is <CarTurnSteeringWheelRight, CarStopCar, CarDenySpeeding> (see Sect. 4.1). Note that this sequence was initially granted with the lowest probability rate of 0.2 (out of 1.0), but during the simulation process the reinforcement learning made it the one that is to be selected first in case of a flat tire.

5 Conclusion

An autonomous vehicle is loaded with AI and operates in a potentially nondeterministic environment. This lack of determinism and certainty is additionally extended by requirements, business conditions, and available technology. Therefore, if we want to construct reliable autonomous vehicles, we need to plan for uncertainty by capturing the autonomous and self-adaptive behavior exhibited while operating in the nondeterministic environment. Failure to do so may result in systems that are overly rigid for their purpose, an eventuality unsafe in their autonomy and adaptation.

This paper has presented the authors' experience with the KnowLang framework mastered to capture self-adaptation for autonomous smart vehicles. The approach demonstrated that self-adaptation needs to be properly handled and formalized, so it can be processed by contemporary formal verification techniques. For example, simulation is one possible method that can be helpful in verifying safety properties, via the formalization of non-desirable system states along with the formalization of behavior that will never lead the system to these states. The paper has presented simulation and testing with a KnowLang-based test bed.

Acknowledgements. This work was supported with the financial support of the Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie).

References

1. Vassev, E.: Safe artificial intelligence and formal methods. In: Margaria, T., Steffen, B. (eds.) ISoLA 2016. LNCS, vol. 9952, pp. 704–713. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47166-2_49
2. Baier, C., Katoen, J.P.: Principles of Model Checking. MIT Press, Cambridge (2008)
3. Vassev, E., Hinchey, M.: KnowLang: knowledge representation for self-adaptive systems. IEEE Comput. **48**(2), 81–84 (2015)

4. Vassev, E., Hinchey, M.: Knowledge representation for adaptive and self-aware systems. In: Wirsing, M., Hölzl, M., Koch, N., Mayer, P. (eds.) Software Engineering for Collective Autonomic Systems. LNCS, vol. 8998, pp. 221–247. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16310-9_6
5. Vassev, E., Hinchey, M.: Awareness in software-intensive systems. IEEE Comput. **45**(12), 84–87 (2012)

Enhancement of Wu-Manber Multi-pattern Matching Algorithm for Intrusion Detection System

Soojin Lee and Toan Tan Phan(✉)

Cyber Security Lab, Department of Computer Science and Engineering,
Korea National Defense University, Nonsan, Republic of Korea
cyberkma@gmail.com, pttoan.it@gmail.com

Abstract. Intrusion Detection System (IDS) is a monitoring system that is the most commonly used today. IDS monitors and analyzes network traffic to detect and prevent malicious behaviors. The main process of IDS is pattern matching, which typically accounts for about 70% of IDS processing time [9]. Wu-Manber [11] is one of the fastest pattern matching algorithms [3] which is commonly used in IDSs. It uses hash techniques to build the hash tables based on the shortest patterns. However, the difference between patterns often degrades the efficiency of the algorithm. In this paper, we propose an improved Wu-Manber algorithm that reduces dependence on the shortest patterns by combining Bloom filters. The experimental results show that our algorithm reduces the matching time by 10% in the worst case and 78% in the best case compared to the original Wu-Manber algorithm, and also reduces the memory usage by 0.3%.

Keywords: Intrusion detection systems · Pattern matching · Network security
Wu-Manber · Bloom filters

1 Introduction

Intrusion Detection System is a monitoring system that is the most commonly used to monitor and analyze network traffic in an effort to detect and prevent malicious behaviors. IDSs are classified based on detection approach consisting of signature-based and anomaly-based. The signature-based detection uses a set of rules (or signatures) to detect the intrusions, while the anomaly-based detection uses the machine learning techniques to detect anomaly behaviors (such as zero-day attacks). Snort is an open signature-based intrusion detection system that is the most commonly used.

Several algorithms are used in IDS such as Aho-Corasick [1] (AC), Boyer-Moore [6] (BM), and Wu-Manber [11] (WM), in which Wu-Manber is one of the fastest algorithms on average. It is a hash-based algorithm which uses the concept of "shift bad characters" from BM algorithm [6] to get the maximum shift distance in case of pattern mismatch. WM [11] uses only first m characters of each pattern, with m is the length of the shortest pattern, to build three hash tables: SHIFT, HASH, PREFIX. These hash tables then are used in the searching phase. However, according to fact survey results of the currently latest Snort 2.9 database, the shortest pattern is 3 characters and the

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 69–78, 2018.
https://doi.org/10.1007/978-3-319-77818-1_7

longest pattern is 516 characters, and 51% of all the patterns is greater than 9 characters. If the algorithm is run on only the first m characters of each pattern (in this case m is 3), then it will take a lot of time to match remaining characters, assuming the first m characters matched. Most previous research about the improvements of WM did not address this issue, and they were only interested in the pattern prefix part.

In this paper, we propose an enhanced Wu-Manber algorithm that focuses on the remaining characters of the patterns (called the pattern *suffix* part). We use Bloom filters instead of the PREFIX table to evaluate quickly a few characters of the *suffix* part that are capable of matching the incoming string before searching in the HASH table. The experimental result shows that our algorithm skipped a significant number of unnecessary accesses to the HASH table. Therefore it brings high-performance improvement in terms of time and memory usage compared to WM algorithm. Section 2 reviews the related background to better understand the problem. Section 3 surveys the related works. Section 4 describes the detailed structure of the proposed algorithm. Section 5 includes our experimental results. Finally, Sect. 6 concludes this paper.

2 Background

2.1 Snort and Pattern Matching Algorithms

Snort is a signature-based intrusion detection system that is the most commonly used in defense systems today. It is an open source software-based tool using a rule-driven language, where each rule consists of headers and options. The headers consist of the protocols, IP addresses, port, etc., The option fields contain the messages, contents, sid, etc., in which the contents are the signatures of attacks that were collected from the monitoring systems. The currently latest Snort 2.9 version includes 44 rule groups with a total of about 17476 rules consisting of dos, dns, ftp, web, icmp, trojan, etc., Fig. 1 represents a sample Snort rule. Several pattern matching algorithms are used in Snorts as Boyer-Moore (BM), Aho-Corasick (AC), and Wu-Manber (WM). These algorithms are classified into either single or multiple pattern matching. The single pattern matching algorithms simply match only one pattern at the moment as BM algorithm. While the multi-pattern matching algorithms, such as AC and WM, match multiple patterns at the moment. WM is a hash-based algorithm which its average performance is better than AC.

```
alert udp $EXTERNAL_NET any -> $HOME_NET 5093 (msg:"ET DOS
Possible Sentinal LM Amplification attack"; dsize:6; content:"|7a 00 00 00 00
00|"; sid:2021172; rev:1;)
```

Fig. 1. Sample rule of Snort

2.2 Wu-Manber Algorithm

Wu-Manber [11] is a hash-based algorithm using "shift bad characters" technique in order to get the maximum shift distance when it finds a mismatch. The database of WM is a set of same length patterns which each pattern is first m characters of original pattern (where m is the length of the shortest pattern [11]). Each pattern is divided into two parts: *prefix* and *block*. WM uses three hash tables: SHIFT, HASH, and PREFIX. SHIFT table contains the maximum shift distance (called *shift-value*) of each block in pattern set. HASH is a hash table of the blocks that its *shift-value* is zero. Each index of HASH is a linked list of all patterns of the same block. PREFIX is a hash table of the *prefixes*. PREFIX is used to quickly check the appearance of a pattern in the linked list of HASH table. WM consists of two main phase: preprocessing and matching phase. In preprocessing phase, the algorithm constructs three hash tables consist of SHIFT, HASH, and PREFIX. In matching phase, WM scans on the incoming string to get blocks. The blocks then are hashed to get results by the same hash functions of the preprocessing phase. The results then are searched in the HASH table for retrieving the corresponding entities. The detailed processes are described as follow:

(a) **Preprocessing phase**

First, WM finds the minimum length m of the shortest pattern. The algorithm then uses only the first m characters for each pattern (known as a representative pattern) to build three hash tables. To construct SHIFT table, each pattern of size m is divided into multiple substrings of size B, called X. The substrings X then are computed the corresponding *shift-value*s and are stored in the SHIFT table. The shift-values are computed as following: For each pattern P*i* to compute the shift-value for each substring X, there are two possibilities. If X does not appear in any pattern, then its shift-value is (m − B + 1) characters. This value also is the default shift-value to build the table [11]. The second case, if X appears in some of the patterns, when the rightmost position of X in the pattern, called q, is located, and the *shift-value* is (m − q).

The HASH table contains only the blocks that its shift-value is zero. Each index of HASH is a linked list of (full) patterns of the same block. The PREFIX table is used to quickly check the appearance of a pattern in the linked list of HASH table. Experimentally, a block size of 2 or 3 is a favorable choice [11]. Figure 2 shows that the SHIFT, HASH, and PREFIX are constructed from a pattern set P = {GetInfo, passwd, password=, sicken, ficken}, where m is 6 and B is 3.

(b) **Matching phase**

In this phase, the algorithm uses a sliding window of size m to scan on the incoming string to get blocks. Each block then is mapped into the SHIFT table to get the *shift-value*. If the *shift-value* is greater than zero, then the window is slid by the length of *shift-value* and repeats the process. Otherwise, if the *shift-value* is zero, then the substring of the window might match to one of the patterns. When the HASH and PREFIX tables are checked to determine the matched pattern. Table 2 represents the matching phase.

Table 1. SHIFT table

Block	Shift	Block	Shift
Get	3	swo	0
etI	2	sic	3
tIn	1	ick	2
Inf	0	cke	1
pas	3	ken	0
ass	2	fic	3
ssw	1	others	4
swd	0		

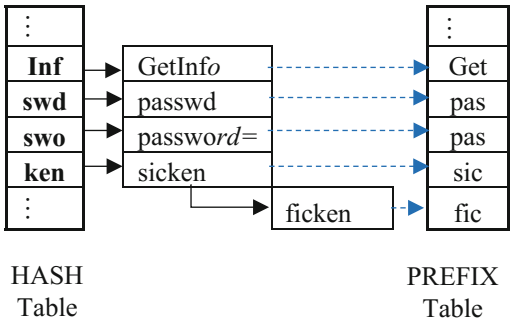


Fig. 2. HASH and PREFIX table

2.3 Bloom Filter

Bloom filter [5] is a hash vector representing for multiple strings which can easily exclude negative matches. It uses less memory space but quickly queries for membership. The Bloom filter computes several hashing functions on each string to get hash results. The results then are indexed into the Bloom vector of size m. The bits at the index position are set to 1. To check the existence of a new string in the string set, the Bloom filter uses the same hash functions to compute the hash results of this string and then checks the corresponding bits in the Bloom vector to determine whether the new string exists or not. If 100% of the hash bits are set, then the new string might exist. Otherwise, the new string does not exist. Bloom filter has no false negatives but has a probability of false positives.

Table 2. Matching phase

Step	LoggedGetInforootpassword=toor	Shift	Output
1	**Logged***GetInforootpassword=toor*	4	
2	Logg**edGetI***nforootpassword=toor*	2	
3	Logged**GetInf***orootpassword=toor*	0	GetInfo
4	LoggedG**etInfo***rootpassword=toor*	4	
5	LoggedGetIn**foroot***password=toor*	4	
6	LoggedGetInforo**otpass***word=toor*	2	
7	LoggedGetInforoot**passwo***rd=toor*	0	password=
8	LoggedGetInforootp**asswor***d=toor*	4	
9	LoggedGetInforootpassw**ord=to***or*	4	
10	LoggedGetInforootpassword=**toor**	End	

3 Related Work

Based on WM, many variants were proposed to overcome the limitation of WM algorithm. In 2009, an improved WM algorithm based on address filtering named as AFWM was proposed by Zhang et al. [4]. Based on the address pointers of the patterns, the Prefix table in AFWM is utilized to filter the linked list of possible matching patterns. The patterns in the linked list are sorted in ascending order according to the address pointers. The advantage of the address filtering algorithm is that it avoids traversing the whole linked list.

In 2015, another improved WM algorithm based on a data structure of AVL tree is implemented by Bhardwaj and Garg [10]. The improved algorithm eliminates the Prefix table, construct two Shift table and uses nonlinear data structure of AVL tree. The results show that the algorithm has better performance as compared to WM and the variants. However, the improvement is no efficiency of memory usage due to the use of two SHIFT table. Moreover, experimental data is no generality for network attacks.

In 2016, new modified WM algorithm based on Bloom filters is implemented by Aldwairi et al. [2, 3]. The algorithm uses Bloom filter instead of the PREFIX table of WM to exclude the unnecessary HASH table searches. The Bloom filters are performed by computing the hash functions on the prefix part of the patterns. In searching phase, the Bloom filter computes the same hash functions for the prefix part of each window, when the *shift-value* of the block is zero, and checking the corresponding bits in the Bloom vector to determine whether the substring might exist or not. If the hash bits appear, then the HASH table will be searched. Otherwise, the HASH table is skipped.

Generally, most previous approaches [7, 8, 12] only focused on the prefix part which is too small than the remaining characters of the patterns. Therefore they could not achieve high-efficiency.

4 Proposed Algorithm

We are interested in the remaining characters, and called suffix, in each pattern. Accordingly, a pattern consists of three parts: *prefix*, *block*, and *suffix*, as shown in Fig. 3. As explained above, WM uses only the first m characters for each pattern, including *prefix* and *block* to build three hash tables: SHIFT, HASH, and PREFIX. After searching in the SHIFT table, if the *shift-value* is zero, the HASH table is accessed to the corresponding position of the block. In fact, however, the probability of finding the matched patterns in a linked list of the patterns is very low (around 5%). There is about 95% of the patterns do not match, while they are still accessed in the HASH table when the *shift-value* is zero. The PREFIX table is used to reduce the unnecessary accesses to the HASH table. However, even if the *prefix* is found, comparing the remaining characters (*suffix*) in original patterns also take too long time. Moreover, in the worst case, when all the *prefixes* are the same (or not exist), there is almost no hope for performance enhancement.

Our proposed algorithm uses the Bloom filters instead of the PREFIX table in order to achieve higher performance. The Bloom vector selection can ensure the efficiency of the memory and the hash functions. The detailed processes of two phases are described as follows.

4.1 Preprocessing Phase

During constructing the HASH table, we insert the Bloom vectors into each index of the HASH table. Each Bloom vector consists of 16 bits and is computed by two hash functions, called *pre-hash* and *suf-hash*. When a pattern is inserted into a linked list of the HASH table: First, the *pre-hash* function hashes the *prefix* part of the pattern. The result then is modulo 5 to get the final result, called *h1*. Second, the *suf-hash* function hashes first *m* characters of the pattern *suffix* part (*from m + 1 to 2 m*). The result *r* then is modulo 11 and plus 5 as the following equation: ((r mod 11) +5) to get the final result, called *h2*. As Fig. 4 represents formatting of a Bloom vector. Finally, the bits of the Bloom vector corresponding to *h1, h2* are set to 1. There are *k* Bloom vectors corresponding to the size of the HASH table. Each Bloom vector represents a linked list of the patterns in the HASH table. Figure 5 represents a new HASH table structure.

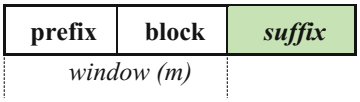


Fig. 3. Pattern format

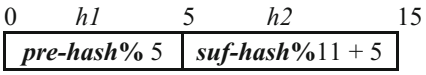


Fig. 4. Bloom vector format

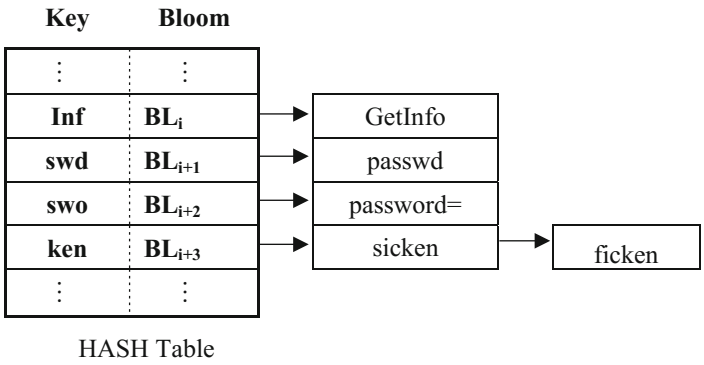


Fig. 5. New HASH table structure

4.2 Matching Phase

Similar to original WM, our algorithm also uses a sliding window of size m to scan on the incoming string T to get blocks. Each block then is mapped into the SHIFT table to get the *shift-value*. If the *shift-value* is greater than zero, then the window is slid by the length of *shift-value* and repeats the process. Otherwise, if the *shift-value* is zero, then the substring of the window might match to one of the patterns. After then, the *pre-hash* and *suf-hash* functions are used to hash the substring. The *pre-hash* computes the *prefix* of the block and *suf-hash* computes the next m characters following the block. Two hash results will then be mapped into the Bloom vector at the corresponding index of the HASH table. The *pre-hash* result will be searched in the first 5 bits of the vector and the *suf-hash* result is the remaining 11 bits. There will be three cases as follow:

- The *pre-hash* result does not appear (bit 0): ignoring all the patterns.
- The *pre-hash* result appears (bit 1), but the *suf-hash* result does not appear (bit 0): Only search the patterns of size less than 2 m characters, ignoring other patterns.
- Both *pre-hash* and *suf-hash* results appear (bit 1): searching all the patterns.

Then sliding the window by 1 character and repeating the matching process until the end of the string.

5 Experiment and Performance Evaluations

Our main goal is the improvement of WM multi-pattern matching algorithm of the Snort. Therefore, our algorithm is built on the database of the Snort and implemented in C++ using Microsoft Visual Studio 2013 as IDE. The performance evaluations were done by comparing to original WM algorithm in the aspects of the preprocessing time, matching time, memory usage and the number of the HASH table accesses. Our algorithm is described as Bloom-Wu-Manber (BWM) algorithm.

5.1 Experimental Data

Our experimental database is the rules set of Snort 2.9. It consists of 17476 rules dividing into 44 groups, including dos, dns, ftp, web, icmp, trojan, etc., Each rule consists of *headers* and *options*. The attack signatures are the strings following keyword "*content*" of the *options*. There are totals of 40767 patterns, in which the shortest pattern is 3 characters, the longest pattern is 516 characters and 51% of the patterns are greater than 9 characters. In the case of the patterns that are shorter than size B (with B = 3) of the block, we concatenate that pattern with the previous pattern from the same rule by one space character as a delimiter. The matching process of a rule with an incoming packet is performed on both the *headers* and *options*. In our experiment, we assume that the headers were already matched, and therefore our algorithm only matches the signatures of the *options*.

We use four sample attack payload datasets that are often used to evaluate the IDSs. It consists of DEFCON20 Capture the Flag (CTF) game packet traces released in 2012, the Information Security Talent Search (ISTS12) in 2015, the all_attack_win and all_attack_unix files of FuzzDB (is like an application security scanner). Table 3 describes Snort database and the experimental datasets.

Table 3. Experimental and test dataset

Snort rule database		Test data	
Number of rules	17476	*Sample set*	*Payloads*
Number of patterns	40767	all-attacks-unix	510
Max_length pattern (ch)	516	all-attacks-win	530
Min_length pattern (ch)	3	DEFCON20	3644
Patterns greater than 9 chars	51%	ISTS12_2015	228030

5.2 Experimental Results

As the time and memory of the program slightly change in each running time, we executed each algorithm (consist of WM and BWM) 100 times on each test dataset to get the average results. To evaluate the effect of the Bloom filter, we compared with each other for the number of the HASH table accesses, processing time and memory usage of each algorithm. A more efficient algorithm should have fewer the HASH accesses, using fewer system resources while ensures better detection result. The detailed evaluation results are shown in the Tables 4, 5, 6. Table 4 shows the comparison results of the number of the HASH table accesses. On average, the number of the HASH table accesses of our algorithm is fewer than WM by 13.45%. However, there is a big difference between the minimum and maximum access counts due to the conflict of the hash functions in the Bloom filter.

Table 4. Number of HASH accesses

Test Data	Found patterns		HASH table accesses		
	WM	BWM	WM	BWM	Performance
all-attacks-win	39364	39364	59186	51980	**12.18%**
all-attacks-unix	47540	47540	72453	66967	**7.57%**
DEFCON20	25032	25032	434294	332972	**23.33%**
ISTS12_2015	2.0E+08	2.0E+08	3.5E+08	3.1E+08	**10.74%**

Table 5 shows that our algorithm has better processing time compared to WM. On average, the preprocessing time of our algorithm is reduced by 10% and the matching time is reduced by 9.2% compared to WM. Table 6 shows that memory usage of our algorithm is also reduced by 0.34%.

Table 5. Preprocessing time and matching time

Test data	Preprocessing time (ms)			Matching time (ms)		
	WM	BWM	Performance	WM	BWM	Performance
all-attacks-win	448	406	**9%**	139	124	**10.79%**
all-attacks-unix	465	406	**13%**	163	146	**10.43%**
DEFCON20	547	500	**9%**	980	847	**13.57%**
ISTS12_2015	563	500	**11%**	1616750	1584380	**2.00%**
Average:			**10%**			**9.20%**

Table 6. Memory usage

Test data	Memory usage (KB)		
	WM	BWM	Performance
all-attacks-win	85012	84688	**0.38%**
all-attacks-unix	85040	84712	**0.39%**
DEFCON20	84957	84678	**0.33%**
ISTS12_2015	85172	84940	**0.27%**
Average:			**0.34%**

The above experimental results are based on the real dataset of Snort 2.9. As the size of the shortest pattern is equal to the size B of the block. Therefore the prefix does not appear, and this is also the worst case. In order to extend the experimental results, we modify the pattern dataset by gradually increasing the size of the shortest patterns from 4 to 6 characters. In each case, we use the test dataset of *all-attacks-win* to compare the corresponding parameters of both algorithms. Table 7 shows that our algorithm is more efficient than WM in all cases. Especially, the matching time heavily depends on the size m. When the larger the size m is, the faster the matching time is. On average, the matching time of our algorithm is reduced by 66%, and in the best case (when m is 6) it can be reduced by 78.49% compared to WM.

Table 7. Algorithm performance depend on the size of m

Size of m (ch)	Num of payloads	Found patterns	Hash table access		Preprocessing time (ms)		Matching time (ms)		Memory usage (KB)	
			WM	BWM	WM	BWM	WM	BWM	WM	BWM
3	40767	39364	59186	↓ 51980	448	↓ 406	139	↓ 124	85012	↓ 84688
4	37465	14292	55292	↓ 52947	420	↓ 375	224	↓ 109	84440	↓ 84180
5	32724	13073	36008	↓ 29447	397	↓ 353	244	↓ 073	83632	↓ 83388
6	29715	10595	24115	↓ 21772	391	↓ 343	265	↓ 057	83088	↓ 82836

6 Conclusions

In this paper, we propose an enhanced Wu-Manber algorithm for intrusion detection systems. Our algorithm uses the Bloom filters instead of the PREFIX table to reduce the number of unnecessary HASH table accesses. We focus on the suffix of the pattern because its size is very large compared to the first m characters of the pattern. The experimental results show that our algorithm is more efficient than WM in both time and memory usage. More specifically, the matching time is reduced by 10% in the worst case and reduced by 78% in the best case compared to WM. The memory usage is also reduced by 0.3% on average.

References

1. Aho, A., Corasick, M.: Efficient string matching: an aid to bibliographic search. Commun. ACM **18**, 333–340 (1975)
2. Aldwairi, M., Al-Khamaiseh, K.: Exhaust: optimizing Wu-Manber pattern matching for intrusion detection using bloom filters. IEEE (2015)
3. Aldwairi, M., Al-Khamaiseh, K., Alharbi, F., Shah, B.: Bloom filters optimized Wu-Manber for intrusion detection. J. Digit. Forensics Secur. Law **11**(4), Article 5 (2016)
4. Zhang, B., Chen, X., Pan, X., Wu, Z.: High concurrence Wu-Manber multiple patterns matching algorithm. In: Proceedings of the International Symposium on Information Process, p. 404 (2009)
5. Bloom, B.H.: Space/time trade-offs in hash coding with allowable errors. Commun. ACM **13** (7), 422–426 (1970)
6. Boyer, R.S., Moore, J.S.: A fast string searching algorithm. Commun. ACM **20**(10), 762–772 (1977)
7. Kacha, C., Shevade, K.A., Raghuwanshi, K.S.: Improved Snort intrusion detection system using modified pattern matching technique. Int. J. Emerg. Technol. Adv. Eng. **3**(7), 81–88 (2013)
8. Yang, D., Xu, K., Cui, Y.: An improved Wu-Manber multiple patterns matching algorithm. In: The 25th IEEE International Performance, Computing, and Communications Conference (IPCCC), pp. 680–686 (2006)
9. Antonatos, S., Anagnostakis, K., Markatos, E.: Generating realistic workloads for network intrusion detection systems. SIGSOFT Softw. Eng. Notes **29**(1), 207–215 (2004)
10. Bhardwaj, V., Garg, V.: Efficient Wu Manber string matching algorithm for large number of patterns. Int. J. Comput. Appl. **132**(17), 29–33 (2015)
11. Wu, S., Manber, U.: A fast algorithm for multi-pattern searching. Technical report TR94-17. University of Arizona at Tuscon (1994)
12. Zhang, W.: An improved Wu-Manber multiple patterns matching algorithm. In: Proceedings of the 2016 IEEE International Conference on Electronic Information and Communication Technology (ICEICT 2016), pp. 286–289 (2016)

Goal-Capability-Commitment Based Context-Aware Collaborative Adaptive Diagnosis and Compensation

Wei Liu[1,2(✉)], Shuang Li[2], and Jing Wang[2]

[1] School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China
liuwei@wit.edu.cn
[2] Hubei Province Key Laboratory of Intelligent Robot, Wuhan, China
l46108697@qq.com, 8783763390@qq.com

Abstract. Agent commitments as a protocol among the participants to melding collaborative patterns typically defined at design-time may not fit the need of collective adaptive systems (CASs). In this paper we propose a Goal-Capability-Commitment (GCC) based context aware collaborative adaptive diagnosis and compensation approach. First, we give a formal and semantic representation of GCC model and define semantic match relations between goals and capabilities/commitments. Second, diagnosing rules are proposed to determine the failures of capabilities and commitments. Third, a generation algorithm apply the capabilities in commitment generation and optimization to verify whether the agent and its partner can collaborate in order to achieve the desired goal under current environment. At last, two experiments over the simulated scenario of automated hospital cart transportation system provide an empirical evaluation of GCC approach.

Keywords: GCC · Agent capability · Semantical match · Context state AGV

1 Introduction

The indispensable aspects of collective adaptive systems (CASs) are adaptation and collaboration [1]. The autonomous agents as executing units in CASs must be capable of adapting their individual behaviors or coordinating their social behaviors over time to achieve a set of goals under dynamically changing and uncertain environment.

Goal models for self-adaptive system to provide an effective mechanism to represent possible changes and making adaptation decision. Baresi et al. [2] propose FLAGS requirements models which are based on the KAOS framework and are targeted at adaptive systems. Ali et al. [3] proposes an automated analysis to reason about contextual goal models that extends Tropos goal model introducing variation points where the context may influent the choice among the available variants of goals satisfaction. Lamsweerde et al. [4] proposes a framework to capture the effects of domain variability on a system using a single context-parameterized i* model and to automatically produce variations of that model based on the currently active context. Such

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 79–89, 2018.
https://doi.org/10.1007/978-3-319-77818-1_8

approaches usually assume that the goals are decomposed until they reach a level of granularity which can be represented as task. In other words, the above assumption is often not true in open CASs for two major reasons: (1) *Unpredictable agents*. The appropriate adaptation decision depends on not only requirements and contexts but also depends on the current status of the autonomous agent configured in CASs. In that sense, there can be impossible to decide which autonomous agent can be thought of as achieving one or more objectives until the system run. (2) *Unpredictable collaboration.* Owing to the autonomous agent's unrestrained joining in or quitting from CASs, the system works as dynamical network of interactions and collaborations.

There are two lines of work that are closed related to the new problems in agent-oriented CASs. One line of work studies agent capability modeling. The capability concept was introduced into the adaptation methodology for it could represent the dynamic and collaborative characteristics [5, 6]. Most previous works focused on action representation formalism of agent capability. Wickler [7] used expressive and flexible action representations to reason about the capabilities for agent cooperation. Cassandra et al. [8] presented an approach to agent capability description and matching that is expressive enough to capture complicated agent functionality. Sycara et al. [9] defined and implemented a language Larks for agent advertisement and request and a match-making process. These approaches assume that the decisions about which agent could achieve the goal by their own capabilities are determined at design time. The second line of work studies agent commitment modeling. The commitment as a protocol to be created among the participants to melding selected collaborative patterns. Chopra et al. [10] develop a semantic relationship between goals and commitments. Akin et al. [11] study in developing dynamic protocols, that can be modified by agents while interacting according to the situation they are in. The previous works assume that a participant agent has a commitment or creates a commitment to another participant to help achieve its goal. However, these approaches generates commitments uses only the locally available knowledge revised by their beliefs. In our case, every agents could able to understands and interact with each other. Hence, we need to propose a flexible diagnosis and compensation approach at runtime that not only decide which agent can be thought to achieve the goal but also allow an agent could make dynamical determination.

The medical waste AGV (Automated Guided Vehicle) transportation in San Antonio Hospital [12] is used as our example scenario. In the system there is three agents: Pickup sensor (denoted as A_1), AGV (denoted as A_2) and Cart elevator (denoted as A_3). The hospital building has two floors and the waste dump is on the ground floor. Every floor has some cart pickups. When a cart loaded medical waste is pushed in the cart pickup, a cart sensor in the ceiling of cart pickups detects the cart reads the RFID tag and delivers the calls for AGV. If an AGV takes the cart to the waste dump in different floor, then a cart elevator with sensors transits the AGV automatically to the waste dump floor.

In this paper we define a Goal-Capability-Commitment (GCC) ontology modeling framework for CASs development. To achieve the targets mentioned above, we develop a GCC model based diagnosis and reconcile approach for CAS. First, we give a formal and semantical representation of GCC model and define semantic matching relations between goals and capabilities/commitments. Second, the diagnose algorithm is proposed to determine the failures of capability and commitment at runtime according to three rules. Third, the compensation algorithm applies the capabilities in

commitment generation and optimization to verify whether the agent and its partner can collaborate in order to achieve the desired goal under current environment. We experimented our approach over the simulated scenario of automated hospital cart transportation system in our GCC-SAR prototype tool. The experiments over the simulated scenario of automated hospital cart transportation system provide an empirical evaluation of GCC model and the algorithms.

2 Technology Base

2.1 Capability-Commitment-Goal Representation

2.1.1 Capability

Capability was proposed as an abstraction for specific plans that an agent may execute to achieve the goal for agent early matchmaking approach. The capability representation could not only modify agents' execution for achieving new goals, but also derive possible candidate commitments for agents' collaboration.

Capability describes how a process is to be executed within what current states are and what the states will be. A capability can be represented as follows:

$$Capability = <Agent, In\text{-}constraints, Plan, Out\text{-}constraints>$$

Agent is specified as 'who' has the capability. *In-constraints* denotes the preconditions of the capability could execute. *Plan* describes the sequence of actions for realizing the capability. *Out-constraints* describes the effects after the performance of capability. *In-constraints* and *Out-constraints* consists of the context states which are defined for representing the value of context factors in our previous work [13]. Context state is denoted as $cs = <f, p, v>$. f means a factor, p is a predicate and v represents a value of factor. For example, "detect_cart" (c_1) is a capability of the pickup sensor (A_1). If the cart is in the pickup, then the capability c_1 could be triggered. After c_1 is executed, the pickup position information is obtained. c_1 = <A_1, {cs_1}, p_1, {cs_2}>. cs_1 = <cart, in, cart_pickup> and cs_2 = <pickup_position, is, available> are two context states.

2.1.2 Commitment

Commitment is a promise made by the debtor commits to the creditor to bring about the consequent provided the antecedent holds. The form of commitment has defined as follows:

$$Commitment = <Debtor, Creditor, Antecedent, Consequent> .$$

Commitment means that a debtor is committed to a creditor for the consequence if the antecedent holds [11]. *Debtor* and *creditor* are cooperative partners who have their own capability and they could execute in order as an alliance. In our approach, the contractual relationship between *debtor* and *creditor* is in accordance with the collaboration of their capabilities. Comparing with previous representation of agent commitment, in commitment antecedent and consequent are not propositional variables but the set of context states. For example, a commitment between AGV (A_2) and Cart elevator (A_3) is denoted

as co_1 = <c_{11}, c_{10}, {cs_{10}, cs_{16}}, {cs_{17}}>. AGV's capability "move_to_elevator" (c_{10}) represents that if the floor destination is different from the floor of AGV then the AGV will move to the cart elevator. Cart elevator's capability "transport_AGV" (c_{11}) represents that if the elevator call is available and the AGV is in the area of elevator sensor then the cart elevator will transport the AGV to the destination floor. In co_1, cs_{10} = <AGV, hasState, busy), cs_{16} = <AGV_floor, notSame, destination_floor and cs_{17} = <same (AGV_floor, destination_floor) are three context states.

2.1.3 Adaptation Goal

Goals in GoalSPEC are specifies system goals, under the hypothesis the domain is described as a set of states [14]. GoalSPEC is a language suitably created for supporting evolution and self-adaptation. Our approach adds some details about the collaborations of goal satisfaction into the extended system goals. A goal can be represented as follows:

$$Goal = <Actor, TrigConditions, FinalStates>$$

Actor is a set of agents specified as 'who' is/are the main responsible to address the given goal; *TrigConditions* must be held in order to activate the goal; *Final States* must be true in order to declare and the goal is finally satisfied. *TrigConditions* and *FinalStates* consists of the context states. In our approach, the actor of goal will not be determined until the system runs and the actor may be changed under the changing environments. The determined actor may be invalid if the agent could not perform in particular condition. A goal could be achieved by an actor or a list of actors. For example, there are two goals "get cart's position" (g_1) *and* "transport AGV to destination floor" (g_2) in the scenario. The goal g_1 = <{A_1}, {cs_1}, {cs_2}> which is achieved by cart elevator (A_1). The goal g_2 denoted as g_2 = <$\varnothing$, {cs_{10}, cs_{16}}, {cs_{17}}> at first, which is denoted as g_2 = <{A_2, A_3}, {cs_{10}, cs_{16}}, {cs_{17}}> after achieved by AGV (A_2) and cart elevator (A_3).

2.2 Semantic Matching

Semantic matching provides a machine-readable way to decide whether the goal is achieved by agent's capabilities or commitments between agents. Semantic imply relationship between sets of context states is analyzed according to contains relationship between context states. In our approach, context states with SWRL atoms formats are translated into PROLOG clause. A clause cs_i θ-subsumes another clause cs_j, written $cs_i \preccurlyeq_\theta cs_j$, if there is a substitution θ such that $cs_i\theta \subseteq cs_j$ holds [15].

Definition 1 (Semantic imply). Given two sets of context states S^i_{cs} and S^j_{cs}, if $\forall\, cs_i \in S^i_{cs}, \exists\, cs_j \in S^j_{cs}$ and $cs_j \preccurlyeq_\theta cs_i$, then S^i_{cs} semantic implies S^j_{cs}, which is denoted as $S^i_{cs} \Rightarrow S^j_{cs}$.

S^i_{cs} semantic implies S^j_{cs} means that if all the context states in S^i_{cs} are satisfied then all the context states in S^j_{cs} must be satisfied too.

There are two types of constraint-oriented plug-in matching: capability matching and commitment matching.

Definition 2 (Semantic matched by capability). Given a goal g_i = <$actor_i$, $trigConditions_i$, $finalStates_i$> and a capability c_j = <$agent_j$, $in\text{-}constraints_j$, $plan_j$, $out\text{-}constraints_j$>. If $trigConditions_i \Rightarrow in\text{-}constraints_j$ and $out\text{-}constraints_j \Rightarrow finalStates_i$, then the goal g_i is semantic matched by the capability c_j, denoted as $c_j \vDash g_i$. If g_i does not be semantic matched by c_j, denoted as $c_j! \vDash g_i$.

A goal can be semantic matched by a capability, which means that the agent should be capable of achieving the goal by itself. If a goal cannot be semantic matched by anyone capability in the system, which means that a agent need to collaborate with other agent in order to achieve the goal. A commitment between the agent and its partner should be generated under current environment. A then, the goal will be semantic matched by the commitment.

Definition 3 (Semantic matched by commitments). Given a goal g_i = <$actor_i$, $trigConditions_i$, $finalStates_i$> and a commitment co_j = <$debtor_j$, $creditor_j$, $antecedent_j$, $consequent_j$> generated by $debtor_j$ and $creditor_j$. If $antecedent_j \Rightarrow trigConditions_i$ and $finalStates_i \Rightarrow consequent_j$, then the goal g_i is semantic matched by co_j, denoted as $co_j \vDash g_i$. If g_i does not be semantic matched by co_j, denoted as $co_j! \vDash g_i$.

The $debtor_j$ and $creditor_j$ as different capabilities can not belong to the same agent. If two capabilities belong to an agent could work together to achieve a goal, not a new commitment but a new capability with greater granularity consisting of the two capabilities will be generated.

3 Our Approach

3.1 Failure Diagnosis

We define three rules to diagnose the capability failure and commitment failure, as is described in Table 1. The states of capability and commitment were defined in the work of relating goal and commitment semantics [16].

Table 1. The rules diagnose failures

Rules	Representation (context state is denoted as *cs*)
Rule I	$\frac{c.state = Active \wedge (\exists cs \in c.Inc = false \vee \exists cs \in c.Outc = false)}{c.state = Invalid \wedge g.state = Terminated}$
Rule II	$\frac{co.state = Active \wedge \exists cs \in co.Ant = false}{co.state = Violated \wedge g.state = Terminated}$
Rule III	$\frac{co.state = Active \wedge (ca.state = Invalid \vee cb.state = Invalid)}{co.state = Terminated}$

Capability failure means that the executing capability cannot achieve the goal under current environment. The capability failure is diagnosed by Rule I. As described in

Rule I, when a capability c is *Active,* if not all the context states in the in-constraints of c ($c.Inc$) or the out-constraints of c ($c.Outc$) are satisfied, then the state of c translates from *Active* to *Invalid* and the state of the matched goal g translates from *Active* to *Terminated.* If the current value of context is different from the value described in GCC model, then the context state is not satisfied. Commitment failure means that the executing commitment has no ability to achieve the goal under current environment. The commitment failure is diagnosed by Rule II and Rule III. As described in Rule II, when a commitment *co* is *Active*, if not all the context states in the *antecedent* of *co* (*co. Ant*) are satisfied, then the state of *co* translates from *Active* to *Violated* and the state of the matched goal g translates from *Active* to *Terminated.* Rule III directed against the commitment *co* created by two capabilities c_a and c_b. c_a is owned by a creditor a and c_b is owned by a debtor b. Rule III happens if c_a or c_b is Invalid and *co* is Active, then the state of *co* translates from *Active* to *Terminated*.

As the result of the diagnosis, the goal terminated invalidation will trigger capability and commitment compensation to search for an alternative capability or commitments and make sure that the new adaptive solution can be executed.

3.2 Capability and Commitment Compensation

Capability compensation is to search whether there are some alternative capabilities for achieving the terminated goals. If there is no capability can semantic match the goals indirectly, commitment compensation will try to generate new commitments to achieve the terminated goals. There are two types of commitments: cooperated commitment and assisted commitment.

Definition 4 (Commitment generation). Given a goal g_i = <$actor_i$, $trigConditions_i$, $finalStates_i$>, two capabilities c_i = <$agent_i$, $in\text{-}constraints_i$, $plan_i$, $out\text{-}constraints_i$> and c_j = <$agent_j$, $in\text{-}constraints_j$, $plan_j$, $out\text{-}constraints_j$>, if:

(1) $trigConditions_i \Rightarrow in\text{-}constraints_i \cup in\text{-}constraints_j$ and $out\text{-}constraints_i \cup out\text{-}constraints_j \Rightarrow finalStates_i$, then a **cooperated commitment** $co_p = <debtor_p, creditor_p, antecedent_p, consequent_p>$ is generated, and $antecedent_p = in\text{-}constraints_i \cup in\text{-}constraints_j$ and $consequent_p = out\text{-}constraints_i \cup out\text{-}constraints_j$;

(2) $trigConditions_i \Rightarrow in\text{-}constraints_i$, $trigConditions \cup out\text{-}constraints_i \Rightarrow in\text{-}constraints_j$ and $out\text{-}constraints_i \cup out\text{-}constraints_j \Rightarrow finalStates_i$, then an **assisted commitment** co_q = <$debtor_q$, $creditor_q$, $antecedent_q$, $consequent_q$> is generated, and $antecedent_q = in\text{-}constraints_i$ and $consequent_p = out\text{-}constraints_i \cup out\text{-}constraints_j$.

For example, there is a goal "transport AGV to different floor" (g_3) which is denoted as $g_3 = <\varnothing, \{cs_{10}, cs_{16}\}, \{cs_{17}\}>$, and two capabilities "move_to_elevator" $c_{10} = <A_2, \{cs_{10}, cs_{16}\}, \{cs_{14}, cs_{15}\}>$ and "transport_AGV" $c_{11} = <A_3, \{cs_{14}, cs_{15}, cs_{16}\}, \{cs_{17}\}>$. cs_{14} = <elevator_call, is, available> and cs_{15} = <AGV, in, elevatorSensor_area> are two context states. According to the definition of commitment generation, a assist commitment is generated $co_1 = <c_{11}, c_{10}, \{cs_{10}, cs_{16}\}, \{cs_{17}\}>$. The goal g_3 is denoted as $g_3 = <\{A_2, A_3\}, \{cs_{10}, cs_{16}\}, \{cs_{17}\}>$.

We present the commitment generation algorithm in Algorithm 1. The algorithm takes two parameters as inputs: the queue of goals to be supported (I_{tg}), the queue of inactive capabilities (I_c).

Algorithm 1 GENERATION(I_{tg}, I_c)

Require: I_{tg}, the queue of goals to be supported
Require: I_c, the queue of inactive capabilities
1: **for each** g in I_{tg}
2: **do for each** c in I_c
3: **do if** $c \models g$ **then**
4: $c.state \leftarrow active$ and $g.state \leftarrow active$
5: g.ADDMATCHEDCOMPONENT(c)
6: I_{ag}+={g}
7: **if** $g.state = terminated$ **then**
8: tc=GETTRIGCONDITION(g) **and** fs=GETFINSTATES(g),
9: **for each** c_i in I_c
10: **do** $in_i \leftarrow$GETINCONSTRAINTS(c_i) and $out_i \leftarrow$GETOUTCONSTRAINTS(c_i)
11: **for each** c_j in I_c
12: **do** $in_j \leftarrow$GETINCONSTRAINTS(c_j) and $out_j \leftarrow$GETOUTCONSTRAINTS(c_j)
13: **if** $tc \Rightarrow in_i \cup in_j$ **and** $out_i \cup out_j \Rightarrow fs$ **then**
14: $co \leftarrow$GENERATECOMMITMENT(c_i, c_j)
15: co.SETANTECEDENT($in_i \cup in_j$) and co.SETCONSEQUENT($out_i \cup out_j$)
16: $I_{co}^{'}$+={co}
17: $co.state \leftarrow active$ and $g.state \leftarrow active$
18: g.ADDMATCHEDCOMPONENT(co)
19: I_{ag}+={g}
20: **else if** $tc \Rightarrow in_i$ **and** $tc \cup out_j \Rightarrow in_j$ **and** $out_i \cup out_j \Rightarrow fs$ **then**
21: $co \leftarrow$GENERATECOMMITMENT(c_i, c_j)
22: co.SETANTECEDENT(in_i) and co.SETCONSEQUENT($out_i \cup out_j$)
23: $I_{co}^{'}$+={co}
24: $co.state \leftarrow active$ and $g.state \leftarrow active$
25: g.ADDMATCHEDCOMPONENT(co)
26: I_{ag}+={g}
27: **return** I_{ag}

Line 1 diagnoses each terminated goal g in I_{tg}. Line 2–6 search a new capability to semantic match the goal. If there is no capability in I_c could support the goal, then the algorithm will try to generate a new commitment to semantic math the goal according to Definition 4. Line 6–11 get the variables used in the following analysis. If the conditions of the first situation mentioned above are satisfied, then a cooperated commitment co is generated and which is added into $I_{co}^{'}$ (Line 13–19). If the conditions of the second situation mentioned above are satisfied, then an assisted commitment $co^{'}$ is generated and which is added into $I_{co}^{'}$ (Line 20–26). Line 27 returns the queue of goals which are achieved under the current executable environment I_{ag}.

4 Experiments and Discussion

In order to justify the above discuss and test the execution performance of our algorithms we developed a GCC based self-adaptive reconfiguration (GCC-SAR) prototype tool as a plug-in of protégé 3.4.4 (http://protege.stanford.edu/). We experiment our approach over the simulated scenario of automated hospital cart transportation system in the GCC-SAR prototype tool. We performed our experiences on an Intel Core(TM) i5-62002.4 GHz processor with 8 GB memory running Windows 10 professional edition.

In the first experiment we observed the effect of α parameter on the execution performance of Algorithm 1. This parameter control the maximum depth between a domain class and its ancestors in semantic matching that has a direct impact on the rate of the goals could semantical matched by capabilities or commitment. In order to investigate this issue we conducted an experiment in which we took the initial value of α as 0 and then increased it up to 5 and measure the matching rate of goals with the algorithm. We assumed that all the goals are to be supported in the algorithm initially.

Figure 1 shows the results of the experiment in which we observe the effect of α. In the figure the x-axis is the number of goals that need to be semantical matched and the y-axis is the matching rate of all goals. The six plots represent the six conducted experiments. The chart shows that the matching rate is grows with the growth of α until $\alpha = 3$ and changes little with $3 < \alpha < 6$. Note that in experiment we would expect to see that it is not the bigger the value of α is the higher the matching rate is. If the value of α is less than the maximum depth of domain model, then the matching rate is grows with the growth of α. In addition, if the number of capabilities is limited, then the more the number of goals is the slightly lower the matching rate is.

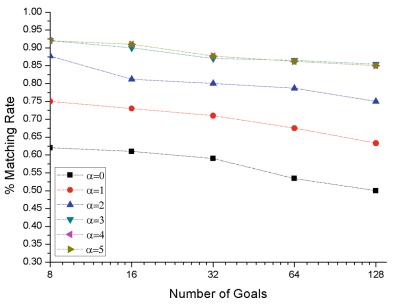
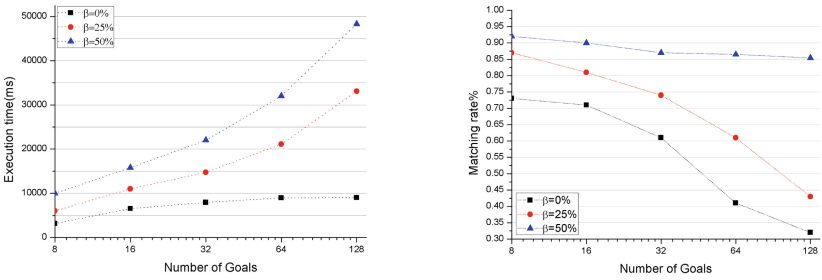


Fig. 1. Comparison of matching rate of the achieved goals based on the maximum depth

In the second experiment we observed the effect of the β parameter on the execution time of the optimization method. This β parameter is the percentage of goals matched by commitments to all matched goals. Since we realize the capability compensate and optimization for achieving each goal, this parameter has a direct impact on the time of the Algorithm 1. We presents the result of the second experiment (we fix the parameter as $\alpha = 3$) in which we observe the effect of β in Fig. 2.



(a) (b)

Fig. 2. Comparison the execution time and matching rate based on the rate of achieved goals

Figure 2(a) presents the execution time of Algorithm 1 for different values of β as 0, 25% and 50% by increasing the number of goals (denoted as x) need to be semantical matched.

The value of β is 0 means that all the goals are achieved by the capabilities. With the increasing the number of goals, the very slow growth of the execution time shows that the execution time of semantic match by capabilities does not grow exponentially. When the values of β are set as 25% and 50%, the results illustrate two important phenomenons. (1) The speedy increasing of the execution time illustrates that the semantic match by commitments could spend much more time than by the semantic match by capabilities. (2) When the numbers of goals achieved by commitments are the same in different case, such as (x = 16, β = 50%) and (x = 32, β = 50%), (x = 32, β = 25%) and (x = 64, β = 25%), (x = 50%, β = 25%) and (x = 128, β = 25%), the similar results show clearly that the number of goals achieved by commitments is the critical factor in the execution times.

Figure 2(b) presents the he matching rate of all goals with the different values of β as 0, 25% and 50% by increasing the number of goals (denoted as x) need to be semantical matched. When the value of β is set as 50%, the very slow decrease of matching rate with the exponential increasing the number of goals shows that the semantic match by commitments could maintain the matching rate in a high level. When the values of β are set as 25% and 0%, the speedy decrease of matching rate illustrates that the less commitments is generated the lower matching rate is.

5 Conclusion

GCC framework based self-adaptive diagnosis and compensation approach could reduces the gap between the requirements model and the executable model for collaborative adaptive system. The main contribution of the paper include: (1) the semantic representation for goal, capability and commitment could provide foundation of heterogeneous agents recognize the executive behavior information of each other at

runtime; (2) capabilities are applied in commitment generation and optimization to verify whether the agent and its partner can collaborate in order to achieve the desired goal under current environment.

We are currently integrating this proposal with other multi-agents interaction modeling techniques based on the agent commitments. As such, more characteristics of social behavior such as competition and disposition need to be considered. We are also beginning to study capability-based components collaborative for CAS.

Acknowledgment. Project supported by the National Natural Science Foundation of China under Grant (No. 61502355), the Doctor foundation for Science Study Program of Wuhan institute of technology (No. K201475).

References

1. Ferscha, A.: Collective adaptive systems. In: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, pp. 893–895. ACM (2015)
2. Baresi, L., Pasquale, L., Spoletini, P.: Fuzzy goals for requirements-driven adaptation. In: 18th IEEE International Requirements Engineering Conference (RE), pp. 125–134. IEEE (2010)
3. Ali, R., Dalpiaz, F., Giorgini, P.: A goal-based framework for contextual requirements modeling and analysis. Requir. Eng. **15**, 439–458 (2010)
4. Lamsweerde, A.V.: Requirements Engineering: From System Goals to UML Models to Software Specifications. Wiley, Hoboken (2009)
5. Cassandra, A., Chandrasekara, D., Nodine, M.: Capability-based agent matchmaking. In: International Conference on Autonomous Agents and Multiagent Systems, pp. 201–202 (2000)
6. Katia, S., Klusch, M., Widoff, S., Lu, J.: Dynamic service matchmaking among agents in open information environments. ACM Sigmod Rec. **28**, 147–153 (1999)
7. Wickler, G.: Using expressive and flexible action representations to reason about capabilities for intelligent agent cooperation (2000). http://www.dai.ed.ac.uk/students/gw/phd/story.html
8. Cassandra, A., Damith, C., Marian, N.: Capability-based agent matchmaking. In: The Fourth International Conference on Autonomous Agents. ACM (2000)
9. Sycara, K., Widoff, S., Klusch, M., Lu, J.: LARKS: dynamic matchmaking among heterogeneous software agents in cyberspace. Auton. Agents Multi-Agent Syst. **5**(2), 173–203 (2002)
10. Chopra, A.K., Dalpiaz, F., Giorgini, P., Mylopoulos, J.: Reasoning about agents and protocols via goals and commitments. In: International Conference on Autonomous Agents and Multiagent Systems, vol. 1, pp. 457–464 (2010)
11. Günay, A., Winikoff, M., Yolum, P.: Dynamically generated commitment protocols in open systems. Auton. Agents Multi-Agent Syst. **29**(2), 192–229 (2015)
12. Hospital Automatic Guided Vehicle Systems (2015). http://www.agvsystems.com/wp-content/uploads/2015/03/AVANT_AGV_AGC_HOSP.pdf
13. Liu, W., Feng, Z.W.: Context-based requirement modeling for self-adaptive service software. J. Comput. Inf. Syst. **8**(24), 10131–10140 (2012)

14. Sabatucci, L., Ribino, P., Lodato, C., Lopes, S., Cossentino, M.: GoalSPEC: a goal specification language supporting adaptivity and evolution. In: Cossentino, M., El Fallah Seghrouchni, A., Winikoff, M. (eds.) EMAS 2013. LNCS (LNAI), vol. 8245, pp. 235–254. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-45343-4_13
15. Sycara, K., Widoff, S., Klusch, M., Lu, J.: LARKS: dynamic matchmaking among heterogeneous software agents in cyberspace. Auton. Agents Multi-Agent Syst. **5**, 173–203 (2002)
16. Telang, P.R., Singh, M.P., Yorke-Smith, N.: Relating goal and commitment semantics. In: Dennis, L., Boissier, O., Bordini, R.H. (eds.) ProMAS 2011. LNCS (LNAI), vol. 7217, pp. 22–37. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31915-0_2

Traffic Incident Recognition Using Empirical Deep Convolutional Neural Networks Model

Nam Vu(✉) and Cuong Pham

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
{namvh,cuongpv}@ptit.edu.vn

Abstract. Traffic incident detection plays an important role for a broad range of intelligent transport systems and applications such as driver-assistant, accident warning, and traffic data analysis. The primary goal of traffic incident detection systems in real-world is to identify traffic violations happening on the road in real-time. Although research community has made a significant attempt for detecting on-road violations, there are still challenges such as poor performance under real-world circumstances and real-time detection. In this paper, we propose a novel method which utilizes the powerful deep convolutional neural networks for vehicle recognition task to detect traffic events on the separate lane. Experimental results on real-world dataset videos as well as live stream in real-time from digital cameras demonstrate the feasibility and effectiveness of the proposed method for identifying incidents under various conditions of urban roads and highways.

Keywords: Convolutional neural network · Traffic incident
Vehicle detection

1 Introduction

Automatic incident detection (AID) is a proper solution for traffic incident in intelligent transportation systems (ITS). Recently, AID has attracted research community as it is an indispensable component in modern frameworks of ITS [1]. The existing approaches of AID in literature can be categorized into direct detection and indirect detection [2]. The former determines whether vehicles crash or obstacles occur based on the information acquired by sensors pre-installed on the road. Although these algorithms are simple and relatively effective, their false alarm rate (FAR) is often dependent on the density of the traffic means. The performance is especially not high at rush hours (dense vehicles). The later, however, indirectly detects traffic incidents by analyzing traffic data collected from monitoring stations. This approach might achieve higher detection rate and lower FAR than the former. The detection methods of traffic incident can also be divided into the classical method and the modern intelligent methods.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 90–99, 2018.
https://doi.org/10.1007/978-3-319-77818-1_9

Classical detection mainly focuses on traffic data captured from digital cameras under resource-constraints such as the limitations of time computation and memory. Due to resource limitation, amount of data might not highly available, classical approaches might struggle with building accurate models for traffic incident detection. Typically AID relies on pattern recognition techniques such as neural networks or support vector machines (SVM). For example, [3] proposed an approach using fuzzy inference for solving AID problem. While artificial neural networks model (ANN) was implemented in [4] by Dipti Work by Xiao et al. [5] built an AID system using multiple kernel Support vector machine (SVM). Somehow similar to [3] on the use of fuzzy logic as inference schema, work by Ren et al. [6] proposed a fuzzy-identification method that is combined with SVM to detect and position traffic incident, and later on for analyzing traffic states. Few works such as [7] exploits hybrid technologies to enhance detection performance.

The recent explosion of big data and sensor technology brings an entirely new approach to this problem. Big data can be collected through various ways such as user-wearable devices, social network data, etc. In [8] GPS data of travelers is utilized to classify anomalous traffic behavior into a different type of traffic incidents, while [9] implemented time and location of traffic congestion detection system by using on-board GPS mounted on probe vehicle. [10] proposed a method applying text mining from Twitter to extract vehicles incident information both highways and arterials as an efficient and low-cost alternative to traditional data sources. With the increasing number of mobile phone users, while participating in traffic, mobile phone usage data has become another source of useful information to detect traffic incidents as reported in [11]. Although big data is becoming a trend in computer science with a huge achievement. In big data, a lot of complicated noise may be included in the real data and need to be handled very carefully.

Although the existing AIDs have made significant progression, the detection accuracies are still not fulfilling for many practical applications. Two main reasons for this challenge are various conditions of lightning, illumination, weather, environment etc. under real-world settings and suitable machine learning techniques. With the rapid increase of computing resources such as GPU and Memory, deep learning approach has become a promising tool effectively used in many computer vision domains including AID. In this work, we propose Convolutional Neural Networks (CNN) model for automatically detecting and positioning of traffic incidents in road lane by using a camera system installed at the traffic lamp sited at the road. Each lane is calibrated using an array of predefined zones to create inputs for the CNN model as well as the position where traffic incidents occur. The sizes of the predefined zones are empirically chosen according to the camera setting position and the real size of the vehicles on the road where the system is deployed. Then, each zone will be considered a single image to feed the CNN model to identify whether there is a vehicle appearing on the predefined zone or not at any time.

The remaining sections of this paper are organized as follows. Section 2 presents our proposed CNN models for traffic incident detection Experiments are present in Sect. 3. We end up with conclusions and discussions in Sect. 4.

2 Convolutional Neural Networks (CNN) Based Approach to Traffic Incident Detection

CNN represents feed-forward neural network which is the combination of various layers such as the convolutional layers, max-pooling layers, and fully connected layers. Recent studies have shown that deep learning has achieved good results in many areas including the image classification field. In this paper, we utilized the deep learning framework for the vehicle classification problem.

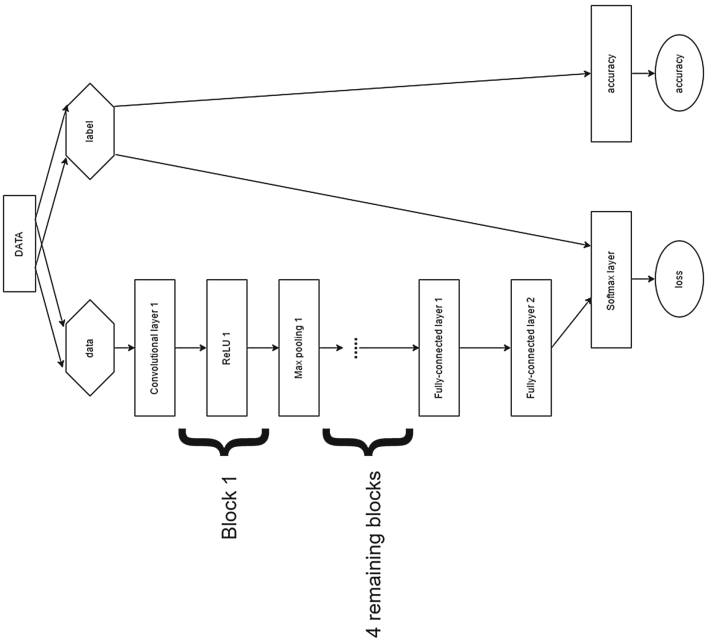


Fig. 1. 5-layers CNN.

2.1 Proposed CNN Model

In this work, we address the problem of traffic incident detection. Firstly, images of the predefined zone are classified into three different classes which are car, human (i.e. motorbike) and unknown objects to verify where vehicles (i.e. car) appear on predefined zone or not. Figure 1 illustrates a 5-layer CNN architecture that takes care of classification task. The final output of CNN model is the probability of pre-defined zone that belongs to one of three classes. These probabilities are so-called confidence scores which are then utilized to verify whether a

vehicle passes through the predefined zones. In the field of deep learning, experimentation demonstrates that the deeper the model with as many parameters as possible, the more accurate the results will be. However, the real-time traffic surveillance system always requires the processing time to be fast enough, therefore the 5-layer CNN architecture has been chosen which achieves fast processing time but remaining acceptable accuracy. The proposed CNN model is formed by two convolutional, rectified linear unit (ReLU), max-pooling, fully-connected, dropout and softmax layers.

The convolutional layer is the main part of the CNN model, consisting of a number of trainable filters. Each filter is independently convolved across the input image, calculating the dot product between the entries of the filter and the input image, then producing a feature map of that filter. A convolutional layer of the CNN model helps to detect some specific type of features at some spatial position in the image hierarchically. Therefore CNN model is able to learn feature from simple to complex manner.

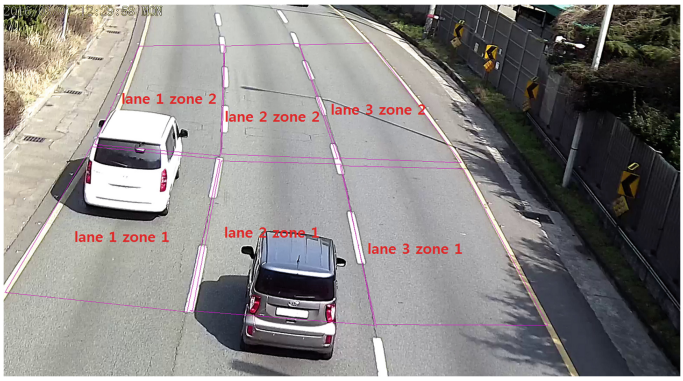


Fig. 2. Lane setting example.

Next to Convolutional layer is a ReLU activation layer, Pooling layer. While ReLU layer will apply an elementwise activation function, where all the negative pixel will be replaced by zero, the Pooling layer will reduce the spatial size of feature maps but remain the most important information.

In this CNN model, we use 5 blocks of the convolutional layer, ReLU, and pooling layers. They are stacked next to each other to extract useful features from the input image, introduce non-linearity in our network and make the feature of input image scalable and translation invariant. The final is two fully-connected layers and softmax layer. In the fully-connected layer, neurons have full connections to all activations in the previous layer, as seen in regular Neural Networks. Their activations can hence be computed with matrix multiplication followed by a bias offset. The last layer of the proposed CNN model is softmax layer. Its function is to predict the accurate class index of the vehicle, human, unknown classes based on the training dataset.

For Convolutional layers in CNN model, the higher layers usually use larger filters to process more complex part of the input image. Therefore we implement 96 filters for convolutional layer 1, 256 filters for convolutional layer 2, 384 filters for convolutional layer 3 and 4, finally 256 filters for convolutional layer 5. The dimension of two fully-connected layers is 4096 and 1000 respectively. The network is trained using mini batches, where each mini-batch contains 30 images of size 227×227. We also minimize the negative log-likelihood using stochastic gradient descent optimizer provided in Caffe.

Experimental results show that when a vehicle passes through a predefined zone, the confidence score increases gradually to the maximum value, thereafter the confidence score decreases gradually to zero. Basing on this observation, a number of the vehicle passing through specific lane can be counted, several traffic illegal incidents also can be detected such as wrong-way vehicle on one direction road, illegal parking or stopping vehicle, and walking person or motorbike enter highway road.

In order to detect wrong-way vehicle in one direction road, we define the array of predefined zones as shown in Fig. 2. If a vehicle goes through 3 consecutive predefined zones in the wrong direction, then that vehicle is alarmed as wrong way event. In order to detect illegal parking or stopping vehicle on the road, every predefined zone is checked whether there is any zone which has the constant confidence score of a car over a period of time. Walking person and motorbike event can be detected by checking every predefined zones whether their confidence score of human or motorbike is over a pre-defined threshold.

2.2 The System for Traffic Incident Detection

Our system is based on the pre-definition of zones on individual lanes; the definition of zones has been completed as given in Fig. 3. The size of the predefined zones based on the presetting of digital cameras and the actual size of the vehicle traveling on the road to ensure the results outputted from the CNN model to be highly precious. In addition, the right direction of each lane is also known in advance to be an input of the wrong way vehicle event examiner. Figure 2 shows an example of lane setting which is defined by the system operator in our real-time monitoring system.

Our system includes three main steps: first of all, each predefined zone image of the specific lane is extracted to be the inputs of CNN model, then the sequence of confidence scores of each class over previous frames are classified into two separate classes which are an event or no event resulted from the Algorithm 1. After that, in order to examine an event of wrong way car, predefined zones neighbor event has to be audited to ensure whether vehicle traveling on the entire lane is wrong way one. The method of detecting wrong way vehicle event in detail is given in Algorithm 2. Finally, based on result given by the previous step, final incident synthesis stage is utilized to give final system decision of event type.

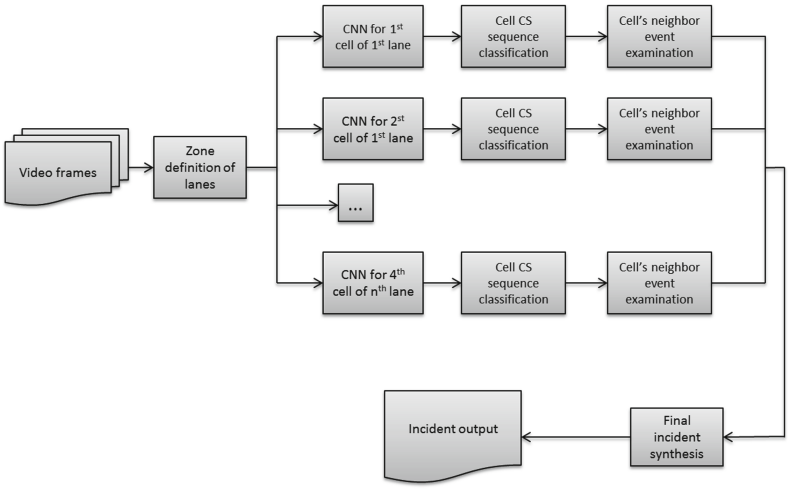


Fig. 3. Flowchart of our proposed method.

Algorithm 1 shows the way the system is implemented to detect whether or not a vehicle passing through each specific lane. This detection process is based on the confidence score of specific class such as car, human and unknown which are the output of CNN model. As given in the algorithm our confidence sequence includes three 10-confidence number arrays of each class. The main task of this algorithm is to examine those arrays separately to detect passing car event happening on each lane. In addition, experimental results show that when a vehicle passes through a predefined zone, the confidence score increases gradually to the maximum value, thereafter the confidence score decreases gradually to zero. Therefore, the examination has been done by using these observations.

Moreover, illegal stopping car and the presence of human or motorbike on the roads which allow only car are able to be detected by algorithm similar to Algorithm 2. In the algorithm, the right direction from predefined zone m to predefined zone n means the real direction which vehicles on the road are allowed to travel, and we assume that the index of predefined zones follows scheme that if a vehicle travels from smaller index zone to larger index zone, that vehicle is considered as wrong way vehicle and an alarm need to be raised. Instead of checking some neighbor predefined zones the algorithm has to check predefined zones individually. When one illegal stopping car or human or motorbike presence events occur in the specific predefined zone in some consecutive frames, then traffic incidents are detected in a specific lane. Illegal stopping car or human and motorbike presence are determined by applying the output of CNN models. Alternatively, the confidence score of human or motorbike as well as car will exceed predefined threshold for each class when human, motorbike and car occurs on the lane. After a while an event of the illegal traffic is detected, a warning message will be sent to the system to inform travelers so that they are able to get appropriate solutions when traveling.

Algorithm 1. Checking car pass

procedure
  begin:
  initialize CCS_Array (Array of car confidence score)
  initialize max _score to 0
  initialize max _threshold to 0.75
  initialize count to 0
  loop through video stream:
  At i^{th} frame:
  Calculate the car confidence score CSSi of the predefined zone.
  if $count < 10$ **then**
    CCS_Array[count] $\leftarrow$ *CCSi*
    $count++$
  else if $count \geq 10$ **then**
    Shift CCS _Array one unit to the left
    CCS _Array[9] $\leftarrow$ *CCSi*
    Checking whether CSS _Array satisfies conditions to be a car pass event
  if CCS _Array increases gradually from left-most to max _score, then decreases to zero **and** $max\_score > max\_threshold$ **then**
    return car pass event
  else
    return no event

At the last step, the events discovered through the above algorithms are rechecked and the warning messages are delivered in the form of a message associated with the location of the event. For example, the wrong car incident occurred in lane 1, the illegal parking car occurred in predefined zone 2 in lane 2.

3 Experimental Results

In order to evaluate the effectiveness of the proposed system, a set of the testing video is collected on the highway, urban and rural road of Viet Nam. Those videos are recorded under different condition of weather and lightning to accurately evaluate and compare proposed system with other existing systems. This data set includes 6 different videos which consist 66 to 81 traffic event to be detected. The resolution of the video is 1920×1080 with a mountable camera.

It is understandable that the performance of the proposed CNN model relies on object recognition accuracies. Particularly, for the traffic incident detection problem addressed in this study, the accuracy of the object recognizer has a great influence on the accuracy of traffic event identification because each type of traffic events is associated with a particular type of objects. However, this article focuses on detecting traffic events under real-world settings. The accuracy of the CNN identifier is not included in the comparison. Our system is installed using road cameras and video streams are processed on a workstation. Workstations are configured to use Caffe framework [12] with GPU support. Our system is able to work in real-time to monitor several cameras at the same time on the certain monitored track.

Algorithm 2. Wrong way car detection

procedure
begin:
Initialize $iEvent\_Array$ (Array of frame number when event happened)
Set nFrame to 20
Assume that the right direction is from zone m^{th} to zone n^{th} $(m > n)$
while frame $i^{th} \in$ streaming video **do**
  while lane j^{th} on the road **do**
    while $z \in \{1, 2, 3\}$ **do**
      Checking car pass from algorithm 1
      if Checking car pass return event **then**
        $iEvent\_Array[z] \leftarrow i$
      if zone z^{th} has passed car zone **and** $(i - iEvent\_Array[z\text{-}1]) < nFrame$ **then**
        return Wrong way event

Table 1. Accuracy comparison

Model	Accuracy	Processing time per frame (millisecond)
SVM + PHOG + GMM	74.5%	15
5-layers CNN	**89.6%**	**12**
GoogleNet CNN	86.4%	86.4
GoogleNet CNN (fine-tune)	89.8%	86.4

We have also compared our proposed method with the state of the art methods as illustrated in the Table 1. Prior to deep learning, the best results in the detection of traffic incidents often rely on manual feature extraction such as PHOG [14–16] and the SVM for object classification. The inputs of the stages based on results derived from the background modeling algorithm. Most of the algorithms used to rely on GMM [17] for background modeling to ensure real-time characteristics. Therefore, these algorithms are heavily influenced by weather conditions because when the weather changes, the conditions of the background also vary. As indicated in the Table 1, the best algorithm is that PHOG + SVM only achieves more than 70% while CNN models show their outperformance at approximately 90%. The CNN model shows that due to the lack of training data in the system, the fine-tune GoogleNet model yielded the highest result with 89.8%. However, our 5-layers model can achieve the approximate result with the faster training and testing time. As shown in Table 1 5-layers CNN model can process one frame within 12 ms compared to more than 80 ms of GoogleNet model. If the big data for training is available our model would yield the same result reported using GoogleNet fine-tune model.

As seen in Table 2, the accuracy of the algorithm is also affected somewhat by bad weather such as rain and unstable lighting conditions such as in tunnels, while the accuracy of the system has reached nearly 100% with normal weather conditions of sparse time urban road as well as highway. The worst performance video is highway traffic in the nighttime as camera setting has to change to infrared mode. In infrared mode, the input image is in the form of grayscale image instead of RGB image, therefore, making our system difficult distinguishing between vehicle and background road.

Table 2. Accuracy in various conditions

	Events detected	Total number of events	Accuracy
Highway traffic in day time	68	70	97%
Highway traffic in night time	65	80	81.3%
Highway traffic in rainy weather	67	81	82.7%
Highway traffic in sunny weather	65	66	98.5%
Urban traffic in sparse time	80	81	98.7%
Highway traffic of tunnel	59	70	84.3%

4 Conclusion

This paper proposed an approach which utilizes deep convolutional neural network to detect traffic incidents under real-world settings. Experimental results show that the proposed method is able to outperform previous methods by more than 10% while processing time is 83 FPS with my hardware configuration (Processor: Intel Core i7, GPU: GTX 1050 Ti, Ram: 8 GB). With the detection accuracies are as high as more than 81% even in nighttime and rainy weather (worst condition), and 97% in daytime and sunny have demonstrated that our proposed solution is very potential for practical applications such as driver-assistant or accident warning. Our future work should overcome few limitations such as improving accuracies of detecting and counting the group of vehicles which pass through predefined zones. Another challenge such as alleviating weather conditions impacting on the system's detection accuracies, which can possibly be solved by acquiring more data of weather conditions in training process.

References

1. Khattak, A., Wang, X., Zhang, H.: Incident management integration tool. IET Intell. Transp. Syst. **6**(2), 204–214 (2012)
2. Wang, J., Li, X., Stephen, S., et al.: A hybrid approach for automatic incident detection. IEEE Trans. Intell. Transp. Syst. **14**(3), 1176–1185 (2013)

3. Ahmed, F., Hawas, Y.E.: A fuzzy logic model for real-time incident detection in urban road network. In: Proceedings of International Conference on Agents and Artificial Intelligence, Barcelona, Spain, pp. 465–472, February 2013
4. Dipti, S., Xin, J., Ruey, L.C.: Evaluation of adaptive neural network models for freeway incident detection. IEEE Trans. Intell. Transp. Syst. **5**(1), 1–11 (2004)
5. Xiao, J., Liu, Y.: Traffic incident detection by multiple kernel support vector machine ensemble. In: Proceedings of the IEEE International Conference on Intelligent (2012)
6. Ren, J., et al.: Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment. IET Int. Trans. Syst. (ITS) **10**(6), 428–437 (2016). Transportation Systems (ITS), Anchorage, USA, September 2012
7. Tian, Q.F., Chen, Y.Z., Zhang, L.G.: Incident detection in urban freeway traffic based on adaptive algorithm. J. Trans. Eng. Inf. **8**(4), 99–103, 125 (2010)
8. Chakraborty, P., et al.: Outlier Mining Based Traffic Incident Detection Using Big Data Analytics. No. 17-05869 (2017)
9. Asakura, Y., et al.: Incident detection methods using probe vehicles with on-board GPS equipment. Transp. Res. Part C: Emerg. Technol. **81**, 330–341 (2016)
10. Gu, Y., Qian, Z.S., Chen, F.: From Twitter to detector: real-time traffic incident detection using social media data. Transp. Res. Part C: Emerg. Technol. **67**, 321–342 (2016)
11. Steenbruggen, J., Tranos, E., Rietveld, P.: Traffic incidents in motorways: an empirical proposal for incident detection using data from mobile phone operators. J. Transp. Geogr. **54**, 81–90 (2016)
12. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. ACM (2014)
13. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
14. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval. ACM (2007)
15. Pham, C., et al.: FoodBoard: surface contact imaging for food recognition. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM (2013)
16. Vu, H.N., Tran, T.A., Na, I.S., Kim, S.H.: Automatic extraction of text regions from document images by multilevel thresholding and k-means clustering. In: 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), pp. 329–334. IEEE, June 2015
17. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 2. IEEE (2004)

Check for updates

Block-Moving Approach for Speed Adjustment on Following Vehicle in Car-Following Model

Trung Vinh Tran[1], Tha Thi Bui[2], Trang Doan Thuy Nguyen[3], Cam Ngoc Thi Huynh[2], and Phuoc Vinh Tran[3(✉)]

[1] Fayetteville State University, 1200 Murchison Road, Fayetteville, NC 28301, USA
ttran1@uncfsu.edu
[2] University of Information Technology, Linhtrung, Thuduc, Hochiminh City, Vietnam
bththa@gmail.com, camhuynhit@gmail.com
[3] Thu Dau Mot University, 6 Tran Van On, Thudaumot, Binhduong, Vietnam
doantrang07@gmail.com, phuoc.gis@gmail.com, phuoctv@tdmu.edu.vn

Abstract. This paper utilizes the traffic flow model of car-following to study the movement of two consecutive vehicles travelling in the same lane of a freeway. The car-following models defined the speed and headway of a moving vehicle but have not analyzed their relation. In the proposed block-moving approach, the following vehicle is considered as a moving object obstructed by another one moving in the same direction. In that, the correlation between the speed and headway of the following vehicle, which are recorded discretely in real time by instruments installed on the following vehicle, is analyzed and processed to avoid collisions caused by the following vehicle. Additionally, the following vehicle must react accordingly to the behaviors of the leading vehicle by increasing or decreasing speed to a desired speed. This study proposed a methodology to effectively determine how much engine or brake force should be made to achieve the desired speed.

Keywords: Car-following · Traffic flow model · Two consecutive vehicles

1 Introduction

On a freeway, vehicles travel successively in lanes where the speed of each vehicle is determined by the speed of its leading vehicle. It is ideal if two vehicles in the same lane travel consecutively at the same speed and keep a constant distance from each other. However, in reality, the speed of a vehicle changes over time and this change affects the travel of the following vehicle. Consequently, the speed of the following vehicle needs to be adjusted accordingly to avoid collisions. Various car-following models have been proposed, some of them apply the data concerning each vehicle to depict the traffic flow as a group of vehicles travelling consecutively on freeway at macroscopic level, others apply the data of several consecutive vehicles to depict the movement of one vehicle at microscopic level.

For two vehicles moving consecutively, this paper considers the leading vehicle (leader) as an obstacle moving relatively with respect to the following vehicle (follower).

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 100–109, 2018.
https://doi.org/10.1007/978-3-319-77818-1_10

To avoid collisions, the follower speed has to be adjusted accordingly over time by making engine force or brake force. We proposed a methodology to detect changes of leader speed based on the data collected from the follower and to determine how much engine or brake force should be made to achieve the desired speed.

This paper is structured as follows. The subsequent section mentions briefly the concepts of car-following models related in the study. The third section analyzes the relation between the acting forces and speed of a vehicle as well as the correlation between the headway and the speed of the follower in the travel of two consecutive vehicles. The fourth section proposes a methodology to effectively determine how much engine or brake force the follower should be affected by. The conclusion section summarizes the results of research.

2 Related Works

Car-following models are considered as half-microscopic, half macroscopic models because the data used in the models not only depict the movement of one vehicle but also the movement of a traffic flow of multiple vehicles [1–3]. Several authors approached car-following models to analyze traffic flow on freeway at the macroscopic level with data related to each vehicle. Others approached car-following models to analyze the data concerning two consecutive vehicles to suppress collisions caused by the follower.

As an earlier known car-following model, the General Motors' model is also known as the stimulus-response model because it depicts the relation between stimulus, which is the difference in speed of leader and follower, and the response of follower driver. The relation of stimulus-response is represented as kinetic equations for one particle [2, 4]. The model considers that the follower drivers are taken a time interval to react against some variation of the leader. As a researching result, the model provides the follower with acceleration at the time point when a reaction is carried out. The acceleration depends on the difference of speed between two consecutive vehicles at the moment of the change of leader speed. The follower acceleration at reacting moment is positive if at the moment of the leader changes speed, its speed is slower than leader speed, and vice versa. The model results in the correlation of speed and acceleration over time between two consecutive vehicles [5]. The General Motors' model demands experiments to obtain the parameters necessary for the calculation of sensitivity coefficient.

The basic optimal velocity model (OVM) as well as its updated models provide with the space headway of a vehicle and the difference on speed of two consecutive vehicles so that the follower driver can achieve the optimal speed. This model assumes that the follower driver can perceive the safe distance based on the speed difference of his vehicle from the leader as well as the safe speed based on its headway [2]. The model is applied for a flow of multiple vehicles travelling in a lane of freeway, where each vehicle demands to receive the data concerning other vehicles of the flow. The optimal velocity model is suitable for intelligent traffic systems (ITS) because drivers can get information concerning other vehicles from the system to determine their optimal speed. The model

enables to stabilize traffic and suppress traffic jam because all drivers can find out their optimal speed [6].

The cellular automaton model separates a traffic flow into cells in space and time. It depicts space of flow as road segments of the same length, called space cells, where each space cell has only one state at a time point, occupied or not occupied by vehicles. It also divides time into equal intervals, called time cells, to update the state of space cells. The speed of each vehicle travelling on road is adjusted each time cell based on the information on the state of the space cells in front. The cellular automaton model can be applied suitably for intelligent traffic systems, where vehicles are provided the information on the state of space cells on road surface ahead. However, the model demands to install underground instruments to detect the state of space cells [2, 7].

The Newell model utilizes space-time trajectories on the ***x-t*** coordinates of two consecutive vehicles to analyze their relative space-time positions, where the slope of a space-time trajectory changes when the speed of the corresponding vehicle changes. Two space-time locations of two trajectories on the ***x-t*** coordinates at the time point they change speed constitute a rectangular defined by the necessary space headway and the time interval during which the follower driver reacts to the change of the leader. This rectangular represents the relation between the available space headway and the time interval during which the driver has to act on the follower to avoid a collision caused by the follower [7].

3 The Block-Moving Approach

The block-moving approach proposed in this paper is to adjust the speed of the follower in a sequence of two consecutive vehicles travelling in a straight line on an even road (i.e. gradient = 0) with good surface. At any moment, the safety of the two vehicles moving on road depends on their speed and the head distance between the front of follower to the rear of leader considered as a moving block ahead of the follower. As a result, the variation of the leader speed affects the head distance, which in turn affects the safety of the two vehicles. Consequently, the follower speed needs to be adjusted to ensure the head distance in a safety range avoiding collision.

3.1 The Clearance of a Moving Vehicle

The clearance of a moving vehicle is the clear distance necessary for vehicle moving ahead, it depends on the vehicle speed. In this study, a moving vehicle is considered as an object referring to time, i.e. time is the reference variable of vehicle attributes such as speed, headway, etc. [8]. Technically, the data of speed and headway are recorded and processed discretely at time points of regular interval Δt [9]. The interval Δt is so designedly small that during which, the speed of vehicles may be considered as linear.

Moving on road, a vehicle may have one of three speed states including unchanging, increasing, and decreasing. According to the second Newton's law on motion, the change of speed depends on the force acting on vehicle.

$$F(t) = m \cdot a(t) \tag{1}$$

Applying this law for a moving vehicle at the time point t_i:

$$F(t) = F_e(t_i) - F_r(t_i) - F_b(t_i) = m\frac{v(t_i) - v(t_{i-1})}{t_i - t_{i-1}} \tag{2}$$

or,

$$v(t_i) = v(t_{i-1}) + \frac{\Delta t}{m}\left[F_e(t_i) - F_r(t_i) - F_b(t_i)\right] \tag{3}$$

$$d(t_i) \approx \Delta t \cdot v(t_i) \tag{4}$$

$$d(t_i) \approx \Delta t \cdot v(t_i) = \Delta t \cdot v(t_{i-1}) + \frac{(\Delta t)^2}{m}\left[F_e(t_i) - F_r(t_i) - F_b(t_i)\right] \tag{5}$$

where

$F(t)$: the vector sum force of the forces acting on vehicle, including engine force, brake force, other resistance forces such as friction, air resistance, and so on.
m: the mass of vehicle, considered as constant.
$a(t)$: the instantaneous acceleration of the vehicle center at the time point t.
$v(t)$: the instantaneous speed of vehicle at the time point t.
$F_e(t)$: the engine force applied for vehicle at the time point t.
$F_r(t)$: the other resistance forces assumed as constant, $F_r(t) = F_r$, $\forall t$.
$F_b(t)$: the brake force controlled by controller.
$d(t_i)$: the distance in front of vehicle necessary for its movement during $[t_{i-1}, t_i] = \Delta t$, called the clearance at the time point t_i.

The engine or brake force acts on the movement of vehicle to constitute three speed states, unchanging, increasing, decreasing.

- *Unchanging speed:* At every time point, $F_e(t) = F_r$ and $F_b(t) = 0$, formula (3) becomes $v(t_i) = v(t_{i-1})$, i.e. the speed of vehicle does not change during $[t_k, t_l]$ for $\forall i | i \in [k + 1, l]$
- *Increasing speed:* It is assumed that from time point t_0, the vehicle is acted by a positive engine force F', i.e. $F_e(t_i) = F_r + F' | i = 0, 1, 2, \ldots$ while the brake force still keeps zero $F_b(t) = 0$, $\forall t \geq t_0$.

  Applying to formula (3):

$$v(t_i) = v(t_{i-1}) + \Delta t \cdot \frac{F'}{m} = v(t_0) + i \cdot \Delta t \cdot \frac{F'}{m} \tag{6}$$

  Associating formulas (4) and (6):

$$d(t_i) = \Delta t \cdot v(t_i) = \Delta t \cdot v(t_{i-1}) + (\Delta t)^2 \cdot \frac{F'}{m} = d(t_{i-1}) + (\Delta t)^2 \cdot \frac{F'}{m} \tag{7}$$

$$d(t_i) = d(t_0) + i \cdot (\Delta t)^2 \cdot \frac{F'}{m} \tag{8}$$

Acting an additional engine force F' on vehicle during the interval $\tau = n \cdot \Delta t$ from time point t_0, the vehicle needs a clearance D to travel ahead. From (8):

$$D(t_1) = d(t_1) = \Delta t \cdot v(t_0) + (\Delta t)^2 \cdot \frac{F'}{m}$$

Generally,

$$D(t_n) = d(t_1) + d(t_2) + .. + d(t_n) = \tau \cdot v(t_0) + \frac{n(1+n)}{2}(\Delta t)^2 \cdot \frac{F'}{m} \tag{9}$$

and

$$v(\tau) = v(n\Delta t) = v(t_0) + n \cdot \Delta t \cdot \frac{F'}{m} \tag{10}$$

The speed $v(t_i)$ in formula (6) is an arithmetic progression with positive common difference of $\Delta t \cdot \frac{F'}{m}$. Formula (7) provides the clearance necessary for the i^{th} period Δt (Fig. 1a). Formula (8) shows that the necessary clearance grows accordingly to a positive common difference of an arithmetic progression. If F' increases, the positive common difference increases, then the necessary clearance grows faster. Specially, formulas (6) and (8) provide the relation between speed and additional engine force as well as between clearance and additional engine force, respectively; they show that the increase of additional engine force results in the rapid increase of speed and clearance.

– *Decreasing speed:* From time point t_0, the vehicle is acted by a brake force F_b, while the engine force returns zero $F_e(t) = 0, \forall t \geq t_0$.

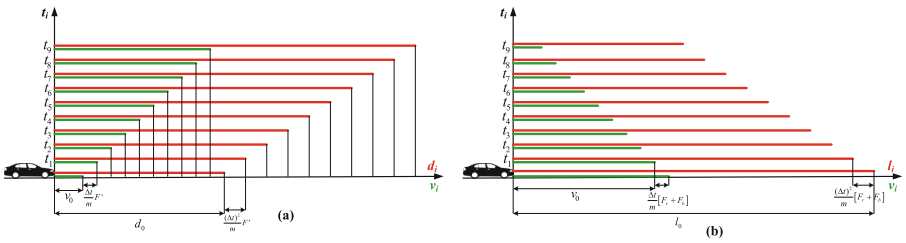


Fig. 1. The speed (green) of vehicle and the clearance (red) necessary for the movement of vehicle at the time points regular-spaced Δt: (a) increasing cumulatively over time according to arithmetic progressions as applying an additional engine force; (b) decrease cumulatively over time according to arithmetic progressions just cutting engine force and applying resistance and/or brake forces. (Color figure online)

Formula (3) becomes:

$$v(t_i) = v(t_{i-1}) - \Delta t \cdot \frac{F_r + F_b}{m} = v(t_0) - i \cdot \Delta t \cdot \frac{F_r + F_b}{m} \tag{11}$$

Formula (5) becomes:

$$d(t_i) = \Delta t \cdot v(t_i) = \Delta t \cdot v(t_{i-1}) - (\Delta t)^2 \cdot \frac{F_r + F_b}{m} \tag{12}$$

$$d(t_i) = d(t_{i-1}) - (\Delta t)^2 \cdot \frac{F_r + F_b}{m} = d(t_0) - i \cdot (\Delta t)^2 \cdot \frac{F_r + F_b}{m} \tag{13}$$

The speed $v(t_i)$ in formula (11) is an arithmetic progression with negative common difference of $\Delta t \cdot \frac{F_r + F_b}{m}$. Formula (12) provides the clearance necessary for the i^{th} period Δt (Fig. 1b). Formula (13) shows that the necessary clearance is reduced accordingly to the common difference of an arithmetic progression. If F_b increases, the negative common difference increases, then the necessary clearance is reduced faster. Specially, formulas (11) and (13) provide the relation between speed and brake force as well as between clearance and brake force, respectively. They show that the increase of brake force results in the rapid decrease of speed and clearance.

3.2 Moving-Block Approach to Adjusting Follower Speed

Studying the movement of two consecutive vehicles in the same lane of a freeway, the problem to be solved is how to adjust the follower speed in accordance with the behaviors of the leader. For solving this problem, the proposed moving-block approach considers the leader as a moving obstacle with respect to the follower to adjust the follower speed suitable for the leader behavior by using data collected from the follower, not requesting data from other sources. The controller of follower has to associate its headway with the clearance necessary for its travelling by increasing or decreasing speed to avoid a collision with the leader.

Technically, the follower can collect its instantaneous speed and the headway from instruments installed on vehicle. The control of the follower is created based on the relation between headway and speed. Based on the length of headway, the speed is controlled by acting an engine or brake force on the vehicle with the suitable intensity (Fig. 2).

$x_f(t)$: the location of the rear of the follower on the axis.
$x_l(t)$: the location of the rear of the leader on the axis.
$v_f(t)$: the instantaneous speed of follower.
$v_l(t)$: the instantaneous speed of leader.
S: the length of follower.

Δt: the sampling period, the time interval between the times recording the values of traffic variables.

$h_f(t)$: the distance from the front of follower to the bump rear of leader is called the space headway of follower, briefly called the headway of follower.

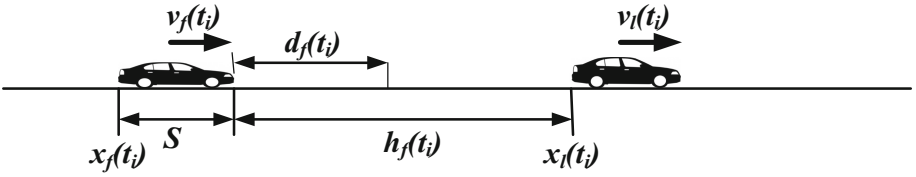


Fig. 2. The variables interacting between two vehicles moving consecutively in the same lane.

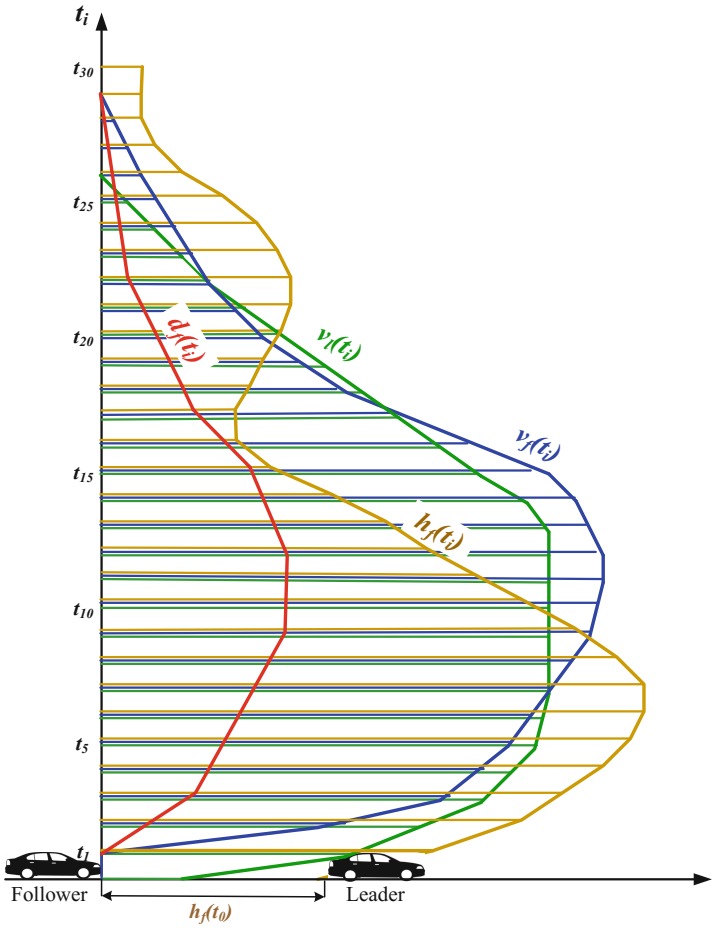


Fig. 3. Illustration of the relation among traffic variables of two consecutive vehicles moving in the same lane of freeway.

Basically, the distance passed by a moving object at regular speed is the product of the speed and the time interval during which the object takes for the distance. The traffic variables such as location, speed, and headway of moving vehicles change continuously over time, and are technically recorded discretely at regular-spaced time points [9]. During an enough small time interval $\Delta t = t_i - t_{i-1}, \forall i | i = 1, 2, ..,$ the movement of a vehicle may be considered as regular and its speed $v(t)$ is calculated as follows.

$$v(t_i) = \frac{x(t_i) - x(t_{i-1})}{\Delta t} \tag{14}$$

Applying (14) for the follower and leader at t_i:

$$\begin{aligned}
&x_f(t_i) = v_f(t_i) \cdot \Delta t + x_f(t_{i-1}) \\
&x_l(t_i) = v_l(t_i) \cdot \Delta t + x_l(t_{i-1}) \\
&x_l(t_i) - x_f(t_i) = [v_l(t_i) - v_f(t_i)]\Delta t + [x_l(t_{i-1}) - x_f(t_{i-1})] \\
&[v_l(t_i) - v_f(t_i)]\Delta t = [x_l(t_i) - x_f(t_i)] - [x_l(t_{i-1}) - x_f(t_{i-1})] \\
&[v_l(t_i) - v_f(t_i)]\Delta t = [x_l(t_i) - x_f(t_i) - S] - [x_l(t_{i-1}) - x_f(t_{i-1}) - S] \\
&v_l(t_i) - v_f(t_i) = \frac{1}{\Delta t}[h_f(t_i) - h_f(t_{i-1})]
\end{aligned} \tag{15}$$

and

$$h_f(t_i) = h_f(t_{i-1}) + [v_l(t_i) - v_f(t_i)]\Delta t \tag{16}$$

(see Fig. 3).

4 Headway-Based Controlling Follower Speed

The condition of safe travel of two consecutive vehicles is indicated by the relation between the headway and speed of the follower as follows:

- $h_f(t_i) \approx \beta \cdot \Delta t \cdot v_f(t_i)$: two vehicles are moving in critical distance. Consequently, the follower speed should not be increased nor decreased;
- $h_f(t_i) < \beta \cdot \Delta t \cdot v_f(t_i)$: two vehicles are moving in hazard distance. Consequently, the follower speed must be reduced by decreasing the current engine force $F_e = F_r + F'$ or perform resistance and brake forces $F_r + F_b$;
- $h_f(t_i) > \beta \cdot \Delta t \cdot v_f(t_i)$: two vehicles are moving in safe distance. Consequently, the follower may increase its speed by changing the engine force F_e.

where β is the safe coefficient indicating the constraint distance for safety measured from the follower front to the leader rear, β is defined by the designer of the vehicle.

Formula (15) is developed as follows:

$$v_l(t_i) - v_f(t_i) = \frac{1}{\Delta t}[h_f(t_i) - h_f(t_{i-1})]$$

$$v_l(t_{i-1}) - v_f(t_{i-1}) = \frac{1}{\Delta t}[h_f(t_{i-1}) - h_f(t_{i-2})]$$

Transforming

$$[v_l(t_i) - v_l(t_{i-1})] - [v_f(t_i) - v_f(t_{i-1})] = \frac{1}{\Delta t}[h_f(t_i) - h_f(t_{i-1})] - \frac{1}{\Delta t}[h_f(t_{i-1}) - h_f(t_{i-2})]$$

It is assumed that,
$h_f(t_{i-1}) = h_f(t_{i-2}) > \beta \cdot d(t_{i-1})$ and $v_f(t_i) = v_f(t_{i-1})$ then
The change of leader speed results in the change of headway:

$$v_l(t_i) - v_l(t_{i-1}) > 0 \Rightarrow h_f(t_i) - h_f(t_{i-1}) > 0$$

$$v_l(t_i) - v_l(t_{i-1}) < 0 \Rightarrow h_f(t_i) - h_f(t_{i-1}) < 0$$

The data of follower speed and headway are recorded and analyzed by instruments installed on the follower. In this case, the follower speed may be adjusted to the desired speed $v_f^d(t_{i+1})$:

$$v_l(t_{i+1}) - v_f^d(t_{i+1}) = \frac{1}{\Delta t}[h_f(t_{i+1}) - h_f(t_i)]$$

$$v_f^d(t_{i+1}) = v_l(t_{i+1}) - \frac{1}{\Delta t}[h_f(t_{i+1}) - h_f(t_i)]$$

In the frame of safe coefficient β, $v_l(t_{i+1})$ may be assigned by $v_l(t_i)$:

$$v_f^d(t_{i+1}) = v_l(t_i) - \frac{1}{\Delta t}[h_f(t_{i+1}) - h_f(t_i)]$$

$$v_f^d(t_{i+1}) = v_f(t_i) + \frac{1}{\Delta t}[2h_f(t_i) - h_f(t_{i-1}) - h_f(t_{i+1})] \tag{17}$$

If $[2h_f(t_i) - h_f(t_{i+1}) - h_f(t_{i-1})] > 0$, then acting on the follower an additional engine force, obtained from (6):

$$F' = m \cdot \frac{[v_f^d(t_{i+1}) - v_f(t_i)]}{\Delta t} = \frac{m}{(\Delta t)^2}[2h_f(t_i) - h_f(t_{i-1}) - h_f(t_{i+1})] \tag{18}$$

If $[2h_f(t_i) - h_f(t_{i+1}) - h_f(t_{i-1})] < 0$, then acting on the follower the resistance and brake forces, obtained from (11):

$$F_r + F_b = m \cdot \frac{[v_f(t_i) - v_f^d(t_{i+1})]}{\Delta t} = \frac{m}{(\Delta t)^2}[h_f(t_{i+1}) + h_f(t_{i-1}) - 2h_f(t_i)] \tag{19}$$

5 Conclusion

The study proposes an approach to adjusting the speed of a vehicle following another based on the speed and headway data of the follower. Technically, these data are recorded discretely with very short time interval defined by the instruments installed on the follower. The proposed concept of moving block of a moving object referring time provides the necessary engine or brake force to increase or decrease speed for safe movement. The clearance resulted from the data of follower speed at each sampling time point is compared with the technically recorded headway to determine the desired speed. The difference between headway and clearance at the same moment provides with the value of engine or brake force performed by the follower. These results can be applied for various speed controlling procedures on vehicles. The diversity of the movement of two consecutive vehicles will be presented in next studies.

References

1. Treibera, M., Kesting, A.: Microscopic calibration and validation of car-following models – a systematic approach. Proc. - Soc. Behav. Sci. **80**, 922–939 (2013)
2. Li, Y., Sun, D.: Microscopic car-following model for the traffic flow: the state of the art. J. Control Theory **10**, 133–143 (2012)
3. Brackstone, M., McDonald, M.: Car-following: a historical review. Transp. Res. Part F **2**, 181–196 (1999)
4. Arif, M., Saccomanno, F., Hellinga, B.: Evaluation of a car-following model using systems dynamics. In: The 19th International Conference of the System Dynamics Society (2001)
5. Mathew, T.V.: Car-following models. In: Transportation Systems Engineering, pp. 14.1–14.8 (2014)
6. Ge, H.X., Cheng, R.J., Li, Z.P.: Two velocity difference model for a car-following theory. Phys. A **387**, 5239–5245 (2008)
7. Zheng, J., Suzuki, K., Fujita, M.: Evaluation of car-following models using trajectory data from real traffic. In: 8th International Conference on Traffic and Transportation Studies, Changsha, China, pp. 356–366 (2012)
8. Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., Wrobel, S.: A conceptual framework and taxonomy of techniques for analyzing movement. J. Vis. Lang. Comput. **22**, 213–232 (2011)
9. Andrienko, N., Andrienko, G., Pelekis, N., Spaccapietra, S.: Basic concepts of movement data. In: Giannotti, F., Pedreschi, D. (eds.) Mobility, Data Mining and Privacy, Geographic Knowledge Discovery, pp. 15–38. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-75177-9_2

The Context-Aware Calculating Method in Language Environment Based on Hedge Algebras Approach to Improve Result of Forecasting Time Series

Minh Loc Vu[1(✉)], Hoang Dung Vu[2], and The Yen Pham[1]

[1] Giadinh University, Ho Chi Minh City, Vietnam
vuminhloc@gmail.com, yen637@gmail.com
[2] National University of Singapore, Singapore, Singapore
hoangdung@u.nus.edu

Abstract. During the recent years, many different methods of using fuzzy time series for forecasting have been published. However, computation in the linguistic environment one term has two parallel semantics, one represented by fuzzy sets it human-imposed and the rest is due to the context of the problem. Hedge algebra is the algebraic approach to the semantic order structure of domain of the linguistic variables that unifies the mentioned above two semantics of each term and therefore, there is a Context-Aware calculating method in the language environment. That is the core of the new approach we mentioned in this article to increase accuracy of solve the time series forecasting problem. The experimental results, forecasting enrollments at the University of Alabama and forecasting TAIEX Index, demonstrate that the proposed method significantly outperforms the published ones. The experimental results, forecasting enrollments at the University of Alabama, demonstrate that the proposed method outperforms the others listed methods.

Keywords: Forecasting · Fuzzy time series · Hedge algebras · Enrollments Intervals · Context-Aware

1 Introduction

Fuzzy time series originally created in 1993 by Song and Chissom [1] and applied to forecast the enrollments at University of Alabama [2, 3]. All steps in the procedure of using fuzzy time series to forecast time series fall into three phases, Phase 1: fuzzifying historical values, Phase 2: mining the fuzzy logical relationships, Phase 3: defuzzifying the output to get the forecasting values. In 1996, Chen [4] opened the new study direction of using fuzzy time series to forecast time series. In this study, Chen suggested the idea of utilize the intervals in the formula of computing the forecasting values only using the arithmetic operators. Sine then, clearly, it seem to be that Phase 1 stronger affects the forecasting accuracy rate. We can see that the step of partitioning the universe of discourse belong to Phase 1.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 110–123, 2018.
https://doi.org/10.1007/978-3-319-77818-1_11

Partitioning the universe of discourse is the essential issue in the method of using fuzzy time series as a tool for forecasting time series. Indeed, product of partitioning the universe of discourse are the intervals as the source that provide the values in the future of time series. So, the better method to partition the universe of discourse we have, the better forecasting values we get. Normally, the method of partition the universe of discourse can be divided into two types through resulted intervals, equal or not the sized intervals. From the empirical results in the list, apply the second type gives the forecasting accurate rate better than the other. Thus, the recent researches focus on the second method.

There have been pretty much method of partitioning the universe of discourse such as paper [5] is the first research confirmed the important role of partitioning the universe of discourse, this employed distribution and average based length as the way to solve the problem. In turn, [6] proposed frequency density, [7] suggested the ratios and [8] use modified genetic algorithm as basis to improve quality of intervals. Information granules are applied in [9–11] to get good intervals on the universe of discourse. By the hedge algebras approach [12, 13] presented the method of partitioning the universe of discourse. According to this approach, fuzziness intervals are used to quantify the values of fuzzy time series that are linguistic terms. These fuzziness intervals are employed as intervals on the universe of discourse. Based upon the fuzziness intervals of values of fuzzy time series, distribution of historical values of time series and adjusted fuzzy logical relationships, we can get the intervals on the universe of discourse. This is the way that the proposed method perform. The rest of this paper is organized as follows: Sect. 2, briefly introduce some basis concepts of HA; Sect. 3 presents the proposed method; Sect. 4 presents empirical results on forecasting enrollments at University of Alabama; Sect. 5 is the conclusion of this paper.

2 Preliminaries

In this section, we briefly recall some concepts associated with fuzzy time series and hedge algebras.

2.1 Fuzzy Time Series

Fuzzy time series are first introduced by Song and Chissom in 1993 [1], it is considered as the set of linguistic values that is observed by the time. Linguistic values are also called linguistic terms. It can be seen that conventional time series are quantitative view about a random variable because they are the collection of real numbers. In contrast to this, as the collection of linguistic terms, fuzzy time series are qualitative view about a random variable. There are two types of fuzzy time series, time-invariant and time-variant fuzzy time series. Because of practicality, the former are the main subject which many of researchers focus on. In most of literature, the linguistic terms are quantified by fuzzy sets. Formally, fuzzy time series are defined as following definition

Definition 1. Let $Y(t)\,(t = \ldots, 0, 1, 2, \ldots)$, a subset of R^1, be the universe of discourse on which $f_i(t)\,(i = 1, 2, \ldots)$ are defined and $F(t)$ is the collection of $f_i(t)\,(i = 1, 2, \ldots)$. Then $F(t)$ is called fuzzy time series on $Y(t)\,(t = \ldots, 0, 1, 2, \ldots)$.

Song and Chissom employed fuzzy relational equations as model of fuzzy time series. Specifically, we have following definition:

Definition 2. If for any $f_j(t) \in F(t)$, there exists an $f_i(t-1) \circ F(t-1)$ such that there exists a fuzzy relation $R_{ij}(t, t-1)$ and $f_j(t) = f_i(t-1) \circ R_{ij}(t, t-1)$ where 'o' is the max-min composition, then $F(t)$ is said to be caused by $F(t-1)$ only. Denote this as

$$f_i(t-1) \rightarrow f_j(t)$$

or equivalently $F(t-1) \rightarrow F(t)$.

In [2, 3], Song and Chissom proposed the method which use fuzzy time series to forecast time series. Based upon their works, there are many studies focus on this field.

2.2 Some Basis Concepts of Hedge Algebras

In this section we refer to paper [14] to briefly introduce some basis concepts in HA, these concepts are employed as basis to build our proposed method. HA are created by Ho et al. in 1990. This theory is a new approach to quantify the linguistic terms differing from the fuzzy set approach. The HA denoted by $AX = (X, G, C, H, \leq)$, where, $G = \{c^+, c^-\}$ is the set of primary generators, in which c^+ and c^- are, respectively, the negative primary term and the positive one of a linguistic variable X, $C = \{0, 1, W\}$ a set of constants, which are distinguished with elements in X, H is the set of hedges, "$\leq$" is a *semantically ordering relation* on X. For each $x \in X$ in HA, $H(x)$ is the set of hedge $u \in X$ that generated from x by applying the hedges of H and denoted $u = h_n \ldots h_1 x$, with $h_n, \ldots, h_1 \in H$. $H = H^+ \cup H^-$, in which H^- is the set of all negative hedges and H^+ is the set of all positive ones of X. The positive hedges increase semantic tendency and vise versa with negative hedges. Without loss of generality, it can be assumed that $H^- = \{h_{-1} < h_{-2} < \ldots < h_{-q}\}$ and $H^+ = \{h_1 < h_2 < \ldots < h_p\}$.

If X and H are linearly ordered sets, then $AX = (X, G, C, H, \leq)$ is called *linear hedge algebra*, furthermore, if AX is equipped with additional operations Σ and Φ that are, respectively, infimum and supremum of $H(x)$, then it is called *complete linear hedge algebra* (ClinHA) and denoted $AX = (X, G, C, H, \Sigma, \Phi, \leq)$.

Fuzziness of vague terms and fuzziness intervals are two concepts that are difficult to define. However, HA can reasonably define these ones. Concretely, elements of $H(x)$ still express a certain meaning stemming from x, so we can interpret the set $H(x)$ as a model of the fuzziness of the term x. With fuzziness intervals can be formally defined by following definition.

Definition 3. Let $AX = (\underline{X}, G, C, H, \leq)$ be a ClinHA. An fm: $X \rightarrow [0, 1]$ is said to be a fuzziness interval of terms in X if:

(1). $fm(c^-) + fm(c^+) = 1$ and $\sum_{h\in H} fm(hu) = fm(u)$, for $\forall u \in \underline{X}$; in this case fm is called complete;
(2). For the constants $\boldsymbol{0}$, $\boldsymbol{W}$ and $\boldsymbol{1}$, $fm(\boldsymbol{0}) = fm(\boldsymbol{W}) = fm(\boldsymbol{1}) = 0$;
(3). For $\forall x, y \in X$, $\forall h \in H$, $\frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}$, that is this proportion does not depend on specific elements and, hence, it is called *fuzziness measure of the hedge h* and denoted by $\mu(h)$.

The condition (1) means that the primary terms and hedges under consideration are complete for modeling the semantics of the whole real interval of a physical variable. That is, except the primary terms and hedges under consideration, there are no more primary terms and hedges. (2) is intuitively evident. (3) seems also to be natural in the sense that applying a hedge h to different vague concepts, the relative modification effect of h is the same, i.e. this proportion does not depend on terms they apply to.

The properties of fuzziness intervals are made clearly through following proposition.

Proposition 3. *For each fuzziness interval fm on X the following statements hold*:

(1). $fm(hx) = \mu(h)fm(x)$, *for every* $x \in X$;
(2). $fm(c^-) + fm(c^+) = 1$;
(3). $\sum_{-q\le i\le p, i\neq 0} fm(h_i c) = fm(c)$, $c \in \{c^-, c^+\}$;
(4). $\sum_{-q\le i\le p, i\neq 0} fm(h_i x) = fm(x)$;
(5). $\sum_{-q\le i\le -1} \mu(h_i) = \alpha$ *and* $\sum_{1\le i\le p} \mu(h_i) = \beta$, *where* $\alpha, \beta > 0$ *and* $\alpha + \beta = 1$.

HA build the method of quantifying the semantic of linguistic terms based on the fuzziness intervals and hedges through υ mapping that fit to the conditions in following definition.

Definition 4. Let $AX = (X, G, C, H, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \le)$ be a CLinHA. A mapping $\upsilon : X \to [0, 1]$ is said to be an semantically quantifying mapping of AX, provided that the following conditions hold:

(1). υ is a one-to-one mapping from X into [0, 1] and preserves the order on X, i.e. for all $x, y \in X$, $x < y \Rightarrow \upsilon(x) < \upsilon(y)$ and $\upsilon(\boldsymbol{0}) = 0$, $\upsilon(\boldsymbol{1}) = 1$, where $\boldsymbol{0}, \boldsymbol{1} \in C$;
(2). Continuity:
$\forall x \in X$, $\upsilon(\boldsymbol{\Phi}x) = infimum\, \upsilon(H(x))$ and $\upsilon(\boldsymbol{\Sigma}x) = supremum\, \upsilon(H(x))$.

Semantically quantifying mapping υ is determined concretely as follows.

Definition 5. Let fm be a fuzziness interval on X. A mapping $\upsilon : X \to [0, 1]$, which is induced by fm on X, is defined as follows:

(1). $\upsilon(\boldsymbol{W}) = \theta = fm(c^-)$, $\upsilon(c^-) = \theta - \alpha fm(c^-) = \beta fm(c^-)$, $\upsilon(c^+) = \theta + \alpha fm(c^+)$;
(2). $\upsilon(h_j x) = \upsilon(x) + Sign(h_j x)\{\sum_{i=Sign(j)}^{j} fm(h_i x) - \omega(h_j x) fm(h_j x)\}$,

where $j \in \{j : -q \le j \le p \ \& j \neq 0\} = [-q \wedge p]$
and $\omega(h_j x = \frac{1}{2}[1 + Sign(h_j x)Sign(h_p h_j x)(\beta - \alpha)] \in \{\alpha, \beta\}$;

(3). $\upsilon(\boldsymbol{\Phi}c^-) = 0$, $\upsilon(\boldsymbol{\Sigma}c^-) = \theta = \upsilon(\boldsymbol{\Phi}c^+)$, $\upsilon(\boldsymbol{\Sigma}c^+) = 1$, and for $j \in [-q\hat{}p]$,

$$\upsilon(\boldsymbol{\Phi}h_jx) = \upsilon(x) + Sign(h_jx)\{\sum_{i=sign(j)}^{j-sign(j)} \mu(h_i)fm(x)\} - \frac{1}{2}(1 - Sign(h_jx))\mu(h_j)fm(x),$$

$$\upsilon(\boldsymbol{\Sigma}h_jx) = \varphi(x) + Sign(h_jx)\{\sum_{i=sign(j)}^{j-sign(j)} \mu(h_i)fm(x)\} + \frac{1}{2}(1 + Sign(h_jx))\mu(h_j)fm(x).$$

The *Sign* function is determined in the following

Definition 6. A function $Sign : X \to \{-1, 0, 1\}$ is a mapping which is defined recursively as follows, for $h, h' \in H$ and $c \in \{c^-, c^+\}$:

(1). $Sign(c^-) = -1$, $Sign(c^+) = +1$;
(2). $Sign(hc) = -Sign(c)$, if h is negative w.r.t. c; $Sign(hc) = +Sign(c)$, if h is positive w.r.t. c;
(3). $Sign(h'hx) = -Sign(hx)$, if $h'hx \neq hx$ and h' is negative w.r.t. h;
$Sign(h'hx) = +Sign(hx)$, if $h'hx \neq hx$ and h' is positive w.r.t. h.
(4). $Sign(h'hx) = 0$ if $h'hx = hx$.

3 Proposed Method

For convenience to present proposed method, we name the linguistic values of fuzzy time series as the variables A_i with $i \in$ N. $Rer\upsilon(x)$ and *Revrfm(x)*, respectively, are the reversed mapping of $\upsilon(x)$ and *fm(x)* from [0, 1] to the universe of discourse of fuzzy time series, U. Denote I_k, on U, as the interval corresponding to A_k.

3.1 Rule for Adjusting the Fuzzy Logical Relationships

We can adjust the fuzzy logical relationships to improve forecasting result depending upon the concrete forecasting problem. The rule for adjusting is as follows:

With A_m is the linguistic term that we are considering. If $Rer\upsilon(A_m)$ is the semantically quantifying mapping of A_m on the universe of discourse, then one also is the semantic core of A_m. If the other values belonging to the fuzziness interval of A_m, then they are semantically equal to $Rer\upsilon(A_m)$, that mean they together reflex the meaning of A_m. If a is the value that belong to $A_{m(+)1}$ and $|Rer\upsilon(A_m) - \mathrm{a}| > |Rer\upsilon(A_{m-(+)1}) - \mathrm{a}|$, then a is more close semantic with $A_{m-(+)1}$ than A_m. So, we can extend $fm(A_{m(+)1})$ cover up a.

3.2 Method for Partitioning the Universe of Discourse

We name the proposed method is VL

Step 1:
Determine the U, the universe of discourse of fuzzy time series $F(t)$. $U = [\mathrm{Min}.F(t) - D1, \mathrm{Max}.F(t) + D2]$, where *D1* and *D2* are two proper positive numbers. Setting n is the number of intervals that we would like to divide on the universe of discourse.

Step 2:
Building the Clin HA with only two hedges, h_{-1} and h_{+1}, $AX = (X, G, H, \Sigma, \Phi, \leq)$ corresponding to linguistic variable that is considered as fuzzy time series *F(t)*. That mean determining the set of parameters of *AX*. Using above HA generate *n* linguistic terms which use to qualitatively describe time series. The way to determine these linguistic terms as follows:

Applying two hedges, h_{-1} and h_{+1}, on the primary generators c^- and c^+, from left to right to generate the linguistic terms.

If the number of linguistic terms are less than, one interval, the number of intervals that we need to divide, then find the interval that contain maximum amount of historical values, assuming that this interval corresponding to the linguistic term A_i. From A_i generating two linguistic term $h_{-1}Ai$ and $h_{+1}Ai$.

Step 3:

Calculating the average of values of $F(t)$, $\bar{F}(t)$;Calculating W and $\mu(L)$ as follows:

$$\mu(h_{+1}) = W = \frac{\bar{F}(t) - Min.F(t)}{Max.F(t) - Min.F(t)}(*),$$

Where $Min.F(t)$ and $Max.F(t)$, respectively, are the max and min of historical values of F(t).

Step 4:

Based upon the distribution of historical values, put them into the corresponding linguistic term' fuzziness interval.

3.3 Algorithm for Forecasting

Step 1:
Apply VL to partition the universe of discourse.

Step 2:
Mine the fuzzy logical relationships: $A_p \rightarrow A_q$, where A_p and A_q, respectively, are the linguistic values of *F(t)* and *F(t+1)*.
Set the group of fuzzy logical relationships having the same left side: $At \rightarrow Au(\mathrm{m}) \ldots Av(n)$, $m, \ldots, n$ are the number of iterations of fuzzy logical relationship $At \rightarrow Au$ and $At \rightarrow Av$.
Adjust the fuzzy logical relationships following rule 3.1.

Step 3:
Compute the forecasting values: Suppose that the value of the time series at $t-1$, *ft*, if *ft* belong to *Revfm*(*At*), then

The forecasting value at t is $\frac{m*Rer\upsilon(\mathrm{Au}) + \cdots n*Rer\upsilon(\mathrm{Av})}{m + \cdots n}(**)$

4 Empirical Result

4.1 We Test the Proposed Method on the Time Series that are Enrollments at University of Alabama

This time series have been used in many previous studies.

We apply proposed method for 7 and 17 intervals.

With 7 intervals

Apply the VL to partition the universe of discourse

$$Max.F(t) = 20000,\ Min.F(t) = 13000,\ \bar{F}(t) = 16194$$

Building $AX = (X, G, C, H, \Sigma, \Phi, \leq)$, Let $G = \{C^- = \text{Low(Lw)}, C^+ = \text{High(Hi)}\}$, $H = H^- \cup H^+$, $H^- = \{\text{Little(L)}\}$, $H^+ = \{\text{Very(V)}\}$

Follow (*) we have

$$W = \frac{16194 - 13000}{20000 - 13000} = 0.4563$$

Continue study the data of mathematical, we can choose $\mu(V) = 0.4563$ and we have $fm(V.Lw) = \upsilon(Lw) = 0.20821$.

Based upon the distribution of historical values we can put the historical values into the following intervals:

$A_1 = [0, \upsilon(\text{VVV.Lw}), \upsilon(\text{Lw}))$ where 0 and υ(Lw), respectively, are left and right border of the linguistic values "LVV.Lw" that mean 'Little-Very-Very-Low'. Similarly, we have:

$$\begin{aligned}
A_2 &= [\upsilon(\text{Lw}), \upsilon(\text{L.Lw}), \upsilon(\text{LVL.Lw}));\\
A_3 &= [\upsilon(\text{LVL.Lw}), \upsilon(\text{VL.Lw}), \upsilon(\text{VVL.Lw})),\\
A_4 &= [\upsilon(\text{VVL.Lw}), \upsilon(\text{VL.Hi}), \upsilon(\text{LLVL.Hi})),\\
A_5 &= [\upsilon(\text{LLVL.Hi}), \upsilon(\text{L.Hi}), \upsilon(\text{LL.Hi})),\\
A_6 &= [\upsilon(\text{LL.Hi}), \upsilon(\text{VLV.Hi}), \upsilon(\text{Hi})),\\
A_7 &= [\upsilon(\text{Hi}), \upsilon(\text{V.Hi}), 1].
\end{aligned}$$

After calculating, we have:

$$\begin{aligned}
&I_1 = [13000, 14457),\ I_2 = [14457, 15598),\ I_3 = [15598, 16029),\\
&I_4 = [16029, 16752),\ I_5 = [16752, 17750),\ I_6 = [17750, 18263),\\
&I_7 = [18263, 20000]
\end{aligned}$$

The semantically quantifying mappings: $Rer\upsilon(A_1) = 13303$, $Rer\upsilon(A_2) = 15402$, $Rer\upsilon(A_3) = 15833$, $Rer\upsilon(A_4) = 16625$, $Rer\upsilon(A_5) = 17138$, $Rer\upsilon(A_6) = 18029$, $Rer\upsilon(A_7) = 19207$.

From Table 1 we have the group of fuzzy logical relationships that show Table 2 as follows.

Table 1. Historical and fuzzified values

Years	Enrollments	Fuzzified values
1971	13055	A_1
1972	13563	A_1
1973	13867	A_1
1974	14696	A_2
1975	15460	A_2
1976	15311	A_2
1977	15603	A_3
1978	15861	A_3
1979	16807	A_5
1980	16919	A_5
1981	16388	A_4
1982	15433	A_2
1983	15497	A_2
1984	15145	A_2
1985	15163	A_2
1986	15984	A3
1987	16859	A_5
1988	18150	A_6
1989	18970	A_7
1990	19328	A_7
1991	19337	A_7
1992	18876	A_7

We have the forecasting result as well as some recent method's as follows:

With 17 intervals:

Similarly, apply the proposed method for 17 intervals on the universe of discourse we will have the forecasting result as follows:

In the field of time series research, RMSE, NE(%) and NNE(%) criterias are alway used to evaluate forecasting quality. $\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^{'} - x_i\right)^2}$, $\text{NE}(\%) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i^{'} - x_i}{x_i}\right|.100$ *and* $\text{NNE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i^{'} - x_i}{x_{max} - x_{min}}\right|.100$ where x_i' is the forecasting value, x_i is historical value and n is the number of forecasting values. This study is also used one to compare proposed method's forecasting result with some recent method's. Based upon RMSE criteria, we can see that proposed method's RMSE is less than the others. That means proposed method forecasting result give more exactly forecasting resullt than the others (Tables 3 and 4).

Table 2. Group of fuzzy logical relationships

Group 1	$A1 \rightarrow A1$ (2), $A1 \rightarrow A2$
Group 2	$A2 \rightarrow A2$ (5), $A_2 \rightarrow A_3$ (2)
Group 3	$A3 \rightarrow A3$, $A3 \rightarrow A5$ (2)
Group 4	$A4 \rightarrow A2$
Group 5	$A5 \rightarrow A4$, $A5 \rightarrow A5$, $A5 \rightarrow A6$
Group 6	$A6 \rightarrow A7$
Group 7	$A7 \rightarrow A7$ (3)

Table 3. Compare result of proposed method with some recent method's (with = 7)

Years	Historical values	Chen et al. 2013	Wang et al. [10]	Lu et al. [11]	Proposed method
1972	13563	14347	13944	14279	14003
1973	13867	14347	13944	14279	14003
1974	14696	14347	13944	14279	14003
1975	15460	15550	15328	15392	15510
1976	15311	15550	15753	15392	15510
1977	15603	15550	15753	15392	15510
1978	15861	15550	15753	16467	15510
1979	16807	16290	16279	16467	17138
1980	16919	17169	17270	17161	17186
1981	16388	17169	17270	17161	17186
1982	15433	16209	16279	14916	15402
1983	15497	15550	15753	15392	15510
1984	15145	15550	15753	15392	15510
1985	15163	15550	15753	15392	15510
1986	15984	15550	15753	15470	15510
1987	16859	16290	16279	16467	17138
1988	18150	17169	17270	17161	17186
1989	18970	18907	19466	19257	19207
1990	19328	18907	18933	19257	19207
1991	19337	18907	18933	19257	19207
1992	18876	18907	18933	19257	19207
RMSE		486.3	506.0	445.2	400.4
NE(%)		2.52	2.68	2.30	1.95
NNE(%)		6.43	6.93	5.88	4.52

4.2 Test of Forecasting TAIEX Index

Chen and Chen [13] have applied their proposed method on the experimental data sets TAIEX Index of November and December 2004. The data set consists of 44 items.

Table 4. Compare result of proposed method with some recent method's

Years	Historical values	Lu et al. [11]	Proposed method
1972	13563	13678	13582
1973	13867	13678	13582
1974	14696	14602	14457
1975	15460	15498	15443
1976	15311	15192	15447
1977	15603	15641	15447
1978	15861	15827	15371
1979	16807	16744	16752
1980	16919	17618	17031
1981	16388	16392	16517
1982	15433	15410	15433
1983	15497	15498	15447
1984	15145	15192	15371
1985	15163	15567	15470
1986	15984	15567	15470
1987	16859	16744	16810
1988	18150	17618	18156
1989	18970	19036	18973
1990	19328	19574	19297
1991	19337	19146	19059
1992	18876	19146	19059
RMSE		256.3	216.1
NE(%)		1.06	0.97
NNE(%)		2.81	2.20

The historical training data of TAIEX is fuzzified into 9 fuzzy sets (h = 9). *The accuracy* metrics of the result:

$$\mathbf{RSME} = \mathbf{56.86; NE}(\%) = \mathbf{0.8; NNE}(\%) = \mathbf{12.44}$$

Our proposed method is applied to the same TAIEX datasets. The process is as follows According to 3.2:

- **Step 1.** Determine the U, the universe of discourse of fuzzy time series $F(t)$. $U = [\min F(t) - D_1, \max F(t) + D_2]$, where D_1 and D_2 are two proper positive numbers.
- **Step 2.** Building the ClinHA with only two hedges, h_{-1}, h_{+1}, $A = (X, G, H, \Sigma, \leq)$. This means determining the set of parameters of AX model needs to be consistent with the context of the problem "forecasting TAIEX Index" mentioned above.

Let

$$\boldsymbol{G} := \{C^- = Low(Lo),\ C^+ = High(Hi)\};\ \mathrm{H} := H^+ \cup H^-;\ H^- := \{Little(L)\};\ H^+ := \{Very(V)\}$$

x_1 := 5759.61 (Actual index of day 01/11/2004 … x_{44} := 6139.69 (Actual index of day 31/12/2004).

The following equations are performed

$$S^+ := \left[\max_{\substack{(x_{i+1}-x_i)>0 \\ 1 \le i \le 43}} |x_{i+1} - x_i|\right] = 94.74 \qquad S^- := \left[\max_{\substack{(x_{i+1}-x_i)<0 \\ 1 \le i \le 43}} |x_{i+1} - x_i|\right] = 138.1$$

$$\bar{F}(t) = \frac{1}{44}\sum_{i=1}^{44} x_i = 5933.51$$

$$W = \frac{F(t) - \min F(t)}{\max F(t) - \min F(t)} = 0.52$$

$$\bar{S} = \frac{1}{43}\sum_{i=1}^{4} 3|x_{i+1} - x_i| = 94.59$$

Because $S^- > S^+$, $\mu(h_{-1}) = \frac{\bar{S}}{S^-}$ hence $\mu(L) = 0.71$.

Continue to apply the algorithms that we have recommended, the following results are achieved:

With h = 7

The values of I_i are calculated

$$I_1 = [5700.00, 5835.00) \qquad I_5 = [6051.16, 6097.36)$$
$$I_2 = [5835.00, 5918.48) \qquad I_6 = [6097.36, 6133.68)$$
$$I_3 = [5918.48, 5986.20) \qquad I_7 = [6133.68, 6150.00$$
$$I_4 = [5986.20, 6051.16)$$

$$\mathrm{Rev}(A_1) = 5743.76 \quad \mathrm{Rev}(A_2) = 5\,883.95$$
$$\mathrm{Rev}(A_3) = 5940.14 \quad \mathrm{Rev}(A_4) = 6001.92$$
$$\mathrm{Rev}(A_5) = 6083.61 \quad \mathrm{Rev}(A_6) = 6119.36 \quad \mathrm{Rev}(A_7) = 6119.57$$

The forecasted values are listed in Table 5, the accuracy metrics are

RSME = 53.87; NE(%) = 0.07; NNE(%) = 10.97

With h = 9 The values of I_i are calculated

I_1=[5700.00, 5777.00); I_2=[5777.00, 5835.00). I_3=[5835.00, 5884.00); I_4=[5884.00, 5992.00). I_5=[5992.00, 6020.00); I_6=[6020.00, 6027.00). I_7=[6027.00, 6083.00); I_8=[6083.00, 6097.00). I_9=[6097.00, 6150.00].

Table 5. Compare forecasted index result of proposed method with result of Chen

Date	Actual index	Chen' forecasted index	Our forecasted index	
		h = 9	h = 7	H = 9
2/11/2004	5759.61	5674.81	5813.86	5743
3/11/2004	5862.85	5768.14	5813.86	5852
4/11/2004	5860.73	5854.81	5892.44	5876.04
5/11/2004	5931.31	5908.26	5892.44	5876.04
8/11/2004	5937.46	5934.81	5912.05	5912.05
9/11/2004	5945.2	5943.81	5912.05	5912.05
10/11/2004	5948.49	5934.81	5912.05	5912.05
11/11/2004	5874.52	5937.12	5912.05	5912.05
12/11/2004	5917.16	5908.26	5892.44	5919.27
15/11/2004	5906.69	5934.81	5892.44	5919.27
16/12/2004	5910.85	5934.81	5892.44	5919.27
17/11/2004	6028.68	5937.12	5892.44	5919.27
18/11/2004	6049.49	6068.14	5977.41	5979.18
19/11/2004	6026.55	6068.14	5977.41	5979.18
22/11/2004	5838.42	5976.47	5977.41	5979.18
23/11/2004	5851.1	5854.81	5892.44	5876.04
24/11/2004	5911.31	5934.85	5892.44	5876.04
25/11/2004	5855.24	5934.81	5892.44	5919.27
26/11/2004	5778.65	5854.81	5892.44	5876.04
29/11/2004	5785.26	5762.12	5813.86	5797.89
30/11/2004	5844.76	5762.12	5813.86	5852
1/12/2004	5798.62	5834.85	5892.44	5876.04
2/12/2004	5867.95	5803.26	5813.86	5797.89
3/12/2004	5893.27	5854.81	5892.44	5876.04
6/12/2004	5919.17	5854.81	5892.44	5919.27
7/12/2004	5925.28	5937.12	5942.00	5912.05
8/12/2004	5892.51	5876.47	5942.00	5912.05
9/12/2004	5913.97	5854.81	5892.44	5919.27
10/12/2004	5911.63	5934.81	5892.44	5919.27
13/12/2004	5878.89	5937.12	5892.44	5919.27
14/12/2004	5909.65	5854.81	5892.44	5919.27
15/12/2004	6002.58	5934.81	5892.44	5919.27
16/12/2004	6019.23	6068.14	5977.41	5979.18
17/12/2004	6009.32	6062.12	5977.41	5979.18
20.12.2004	5985.94	6062.12	5977.41	5979.18
21/12/2004	5987.85	5937.12	5977.41	5979.18
22/12/2004	6001.52	5934.81	5977.41	5979.18
23/12/2004	5997.67	6068.14	5977.41	5979.18

(*continued*)

Table 5. (*continued*)

Date	Actual index	Chen' forecasted index	Our forecasted index	
		h = 9	h = 7	H = 9
24/12/2004	6019.42	5934.81	5977.41	5979.18
27/12/2004	5985.94	6068.14	5977.41	5979.18
28/12/2004	6000.57	5937.12	5942.00	5979.18
29/12/2004	6088.49	6068.14	5977.41	5979.18
30/12/2004	6100.86	6062.12	6119.36	6119.36
31/12/2004	6139.69	6137.12	6143.57	6143.57
	RSME	**56.86**	**53.87**	**48.02**
	NE(%)	**0.80**	**0.70**	**0.59%**
	NNE(%)	**12.44**	**10.97**	**9.17%**

$\text{Rev}(A_1) = 5740.00$ $\text{Rev}(A_2) = 5829.00$ $\text{Rev}(A_3) = 5869.00$ $\text{Rev}(A_4) = 5940.00$
$\text{Rev}(A_5) = 6002.00$ $\text{Rev}(A_6) = 6026.00$ $\text{Rev}(A_7) = 6051.00$ $\text{Rev}(A_8) = 6085.00$
$\text{Rev}(A_9) = 6119.00$

The forecasted values are listed in Table 5, the accuracy metrics are

$$\textbf{RSME} = \textbf{48.02; NE}(\%) = \textbf{0.59;NNE}(\%) = \textbf{9.17}$$

Compared to the results of [13], our method gives more accurate results and the calculating process is much simpler.

5 Conclusion

Researchers who use information granules as models to predict time series has emphasized the inherent semantics of words, e.g. "Information granulation is inherent to fuzzy time series" and "Information granules are human-centric constructs capturing the semantics of the concepts of interest, which are inherent to all ensuing processes of abstraction" [11]. Information granules are linguistic values (or terms). But the inherent semantics of term is resulted from human knowledge hence depends on context. The context here is 'the high or low level of the annual number of enrolled students at Alabama university'. 'high' and 'low' are the two main words whose semantics are used to describe the information within a context. Other words 'little' and 'very' are the impacting words which have effect on 'high' and 'low' to create mediate semantics to illustrate the 'high', 'low' levels corresponding to the annual number of enrolled students at Alabama university. Hedge algebras is an approach to the inherent semantics of words to represent the semantics of information granules by fuzzy sets with their inherent semantics. With the definitions of 'quantitative semantics mapping', 'fuzzy set based semantics of the words', etc., the hedge algebras have constructed 'a set of

weights' which $\in [0, 1]$ hence is a partition of $[0, 1]$. The normalized historical data of time series (from 0 to 1) is distributed within this set hence is the basis for the optimal partitioning. Each partition has a quantitative semantics value which can be considered as "semantics core" meaning all the historical data belongs to this partition will lie around this "semantics core". The fuzzy parameters of the HA are determined based on the analysis of the relationship between the historical values of a given time series. Consequently, the forecasting enrollment has been solved by the Context-Aware approach. The above statements fully explain our approach based on the inherent semantics of term is easy to understand and simplicity in practice to forecasting enrollment in fuzzy time series with remarkable accuracy in comparison with the other approaches has published.

References

1. Song, Q., Chissom, B.S.: Fuzzy time series and its models. Fuzzy Sets Syst. **54**, 269–277 (1993)
2. Song, Q., Chissom, B.S.: Forecasting enrollments with fuzzy time series – part I. Fuzzy Sets Syst. **54**, 1–9 (1993)
3. Song, Q., Chissom, B.S.: Forecasting enrollments with fuzzy time series – part II. Fuzzy Sets Syst. **62**, 1–8 (1994)
4. Chen, S.-M.: Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst. **81**, 311–319 (1996)
5. Huarng, K.: Efective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets Syst. **123**, 387–394 (2001)
6. Jilani, T.A., Burney, S.M.A., Ardil, C.: Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning. Int. J. Comput. Inf. Syst. Control Eng. **4**, 39–44 (2010)
7. Huarng, K., Yu, T.H.-K.: Ratio-based lengths of intervals to improve fuzzy time series forecasting. IEEE Trans. Syst. Man Cybern.—Part B Cybern. **36**, 328–340 (2006)
8. Bas, E., Uslu, V.R., Yolcu, U., Egrioglu, E.: A modified genetic algorithm for forecasting fuzzy time series. Appl. Intell. **41**, 453–463 (2014)
9. Wang, L., Liu, X., Pedrycz, W.: Effective intervals determined by information granules to improve forecasting in fuzzy time series. Expert Syst. Appl. **40**, 5673–5679 (2013)
10. Wang, L., Liu, X., Pedrycz, W., Shao, Y.: Determination of temporal information granules to improve forecasting in fuzzy time series. Expert Syst. Appl. **41**, 3134–3142 (2014)
11. Lu, W., Chen, X., Pedrycz, W., Liua, X., Yang, J.: Using interval information granules to improve forecasting in fuzzy time series. Int. J. Approx. Reason. **57**, 1–18 (2015)
12. Lan, V.N., Ho, N.C.: Hedge algebras-an order-based structure of terms-domains: an algebraic approach to human reasoning. J. Sci. Technol. **45**, 77–108 (2007)
13. Chen, S.-M., Chen, C.-D.: TAIEX forecasting based on fuzzy time series and fuzzy variation groups. IEEE Trans. Fuzzy Syst. **19**(1), 1–12 (2011)

Algebraic Operations in Fuzzy Object-Oriented Databases Based on Hedge Algebras

Doan Van Thang(✉)

FPT University, Ho Chi Minh, Vietnam
vanthangdn@gmail.com

Abstract. The article is about blurry algebraic operations for classes blurred and objects blurred in blurry oriented object databases based on a semantic approximation approach of algebras of the hedge. It also defines the blurry algebraic operations for blurry classes and blurry objects. Finally, these blurred algebraic operations are applied to the process and query data blurred on the blurry object-oriented database model.

Keywords: Fuzzy object-oriented database · Hedge algebra
Algebraic operations

1 Introduction

The blurry relational database model (FRDB) and blurry object-oriented database (FOOD) model and related problems have been widely studied in recent years by many domestic and foreign authors [1–9]. To implement blurry information in the data model, there are several basic approaches: blurry set theory-based model [7], probability and blurry model [1], etc. All these approaches are designed to reach and treat blurry values to build evaluation methods and comparison among them to manipulate data more accessible and accurately.

Based on the advantages of the structure the algebra of the hedge (HA) [5–7], the authors studied the relational database model [5,6,8], and blurry object-oriented database model [2,3] based on the approaches of algebra of the hedge, in which semantic language quantified by quantitative semantic mapping of algebra of the hedge.

In this approach, the semantics of the language can be expressed in a neighborhood of intervals determined by the measure of fuzziness of language values of an attribute as a linguistic variable.

This article is based on approximate measures of the semantics of the two fuzzy data to define the blurry algebraic operations for objects blurry and classes blurry in FOOD model. These blurry algebraic operations are defined as bases for the treatmenting of blurry data in FOOD model.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 124–134, 2018.
https://doi.org/10.1007/978-3-319-77818-1_12

This article is organized as follows: Sect. 2 presents some basic concepts related to algebra of the hedge and FOOD as the basis for the following sections. Section 3 offers solutions to detect and manage redundant objects in FOOD. Section 4 defines blurry algebraic operations to classes blurred and objects blurred, and finally the conclusion.

2 The Basic Concepts

In this section, presents a general overview of the algebra of hedge linear full was proposed by Ho et al. And some related concepts on the for mapping quantification and how to determine neighboring semantic quantitative approach HA [5–9].

2.1 Hegde Algebra

Consider a complete hedge algebra (Comp-HA) $\underline{\mathbf{AX}}$ = (X, G, H, Φ, Σ, $\leq$), where G is a set of generators which are designed as primary terms denoted by c^- and c^+, and specific constants 0, W and 1 (zero, neutral and unit elements, respectively), H = $H^+ \cup H^-$ and two artificial hedges Σ, Φ, the meaning of which is, respectively, taking in the poset $\underline{X}$ the supremum (sup, for short) or infimum (inf, for short) of the set H(x) - the set generated from x by using operations in H. The word "complete" means that certain elements are added to usual hedge algebras in order for the operations Σ and Φ will be defined for all x $\in$ X. Set $Lim(X) = \underline{X} \setminus H(G)$, the set of the so-called limit elements of $\underline{\mathbf{AX}}$.

Definition 1. *A Comp-HAs* $\underline{\boldsymbol{AX}}$ *= (*$\underline{X}$*, G, H,* Σ*,* Φ*,* $\neq$*) is said to be a linear hedge algebra (Lin-HA, for short) if the sets* $G = \{0, c^-, W, c^+, 1\}$, $H^+ = h_1, \ldots, h_p$ *and* $H^- = h_{-1}, \ldots, h_{-q}$ *are linearly ordered with* $h_1 < \ldots < h_p$ *and* $h_{-1} < \ldots < h_{-q}$*, where* $p, q > 1$*. Note that* $H = H^- \cup H^+$.

Proposition 1. *Fuzziness measures fm and fuzziness measures of* $\mu(h)$, $\forall h \in H$*, the following statements hold:*

(1) $fm(hx) = \mu(h) fm(x)$, $\forall x \in X$.
(2) $fm(c^-) + fm(c^+) = 1$.
(3) $\sum_{-q \leq i \leq p, i \neq 0} fm(h_i c) = fm(c)$, *where* $c \in \{c^-, c^+\}$.
(4) $\sum_{-q \leq i \leq p, i \neq 0} fm(h_i x) = fm(x), x \in X$.
(5) $\sum\{\mu(h_i) : -q \leq i \leq -1\} = \alpha$ *and* $\sum\{\mu(h_i) : 1 \leq i \leq p\} = \beta$, *where* $\alpha, \beta > 0$ *and* $\alpha + \beta = 1$.

Definition 2 *(Sign function). Sgn:* $X \rightarrow \{-1, 0, 1\}$ *is a function which is defined recursively as follows, where h, h' ∈ H, and* $c \in \{c^-, c^+\}$*:*

(a) Sgn(c^-*) = −1, Sgn(c+) = +1,*
(b) Sgn(h'hx) = 0, if h'hx = hx otherwise:
Sgn(h'hx) = -Sgn(hx), if h'hx ∈ hx and h' is negative with h.
Sgn(h'hx) = +Sgn(hx), if h'hx ∈ hx and h' is positive with h.

Proposition 2. *With $\forall x \in X$, we have: $\forall h \in H$, if Sgn(hx)= +1 then $hx < x$, if Sgn(hx) = −1 then $hx < x$ and if Sgn(hx) = 0 then $hx = x$.*

From properties of fuzziness and sign function, semantically quantifying mapping of HA is defined as below:

Definition 3. *Let $\underline{\boldsymbol{AX}} = (\underline{X}, G, H, \Sigma, \Phi, \leq)$ be a free linear complete HA, fm(x) and $\mu(h)$ are, respectively, the fuzziness measures of linguistic and the hedge h satisfying properties in Proposition 2. Then, v is a induced mapping by fuzziness measure fm of the linguistic if it is determined as follows:*

(1) $v(W) = k = fm(c^-), v(c^-) = k - \alpha fm(c^-) = \beta fm(c^-), v(c^+) = k + \alpha fm(c^+)$.

(2) $v(h_j x) = v(x) + Sgn(h_j x)\{\sum_{i=Sgn(j)}^{j} \mu(h_i) fm(x) - \omega(h_j x)\mu(h_j) fm(x)\}$, *where* $\omega(h_j x) = \frac{1}{2}[1 + Sgn(h_j x) Sgn(h_p h_j x)(\beta - \alpha)] \in \{\alpha, \beta\}$, *for all j,* $-q \leq j \leq p$ *and* $j \neq 0$.

(3) $v(\Phi c^-) = 0, v(\Sigma c^-) = k = v(\Phi c^+), v(\Sigma c^+) = 1$, *and for all j,* $-q \leq j \leq p$ *and* $j \neq 0$, *we have* $v(\Phi h_j x) = v(x) + Sgn(h_j x)\{\sum_{i=sign(j)}^{j-1} \mu(h_i) fm(x)\}$ *and* $v(\Sigma h_j x) = v(x) + Sgn(h_j x)\{\sum_{i=sign(j)}^{j} \mu(h_i) fm(x)\}$.

2.2 Neighborhood Level

The authors in [8], [10–12] took the blurry intervals of length k as similar long between the elements. This means that the elements of which the representative values belong to the same interval blurred level k which is similar as k-level. However, to build the level k blurry intervals, representative values of the elements of x have length that k is always in the ends of the blurry level k. Therefore, intervals in the determination of k of district level, we expect these representative values as the interior points of the district level k.

We always assume that each set of H^- and H^+ contains at least two hedges. We reviewed by X_k to be a collection of all elements of length k. On the basis of blurred interval on level k and k + 1. The authors [8], [10] built a domain partition [0, 1] as the following:

(1) Similarity level 1: with k = 1, blurry intervals of level 1 includes $I(c^-)$ and $I(c^+)$. Blurry intervals of level 2 on the interval $I(c^+)$ is $I(h_{-q}c^+) \leq I(h_{-q+1}c^+) \leq \ldots \leq I(h_{-2}c^+) \leq I(h_{-1}c^+) \leq v_A(c^+) \leq I(h_1 c^+) \leq I(h_2 c^+) \leq \ldots \leq I(h_{p-1}c^+) \leq I(h_p c^+)$. When that, we build partitions on same level level 1 consists of the equivalence classes as follows: $S(0) = I(h_p c^-)$; $S(c^-) = I(c^-)$ $[I(h_{-q}c^-) \cup I(h_p c^-)]$; $S(W) = I(h_{-q}c^-) \cup I(h_{-q}c^+)$; $S(c^+) = I(c^+)$ $[I(h_{-q}c^+) \cup I(h_p c^+)]$ and $S(1) = I(h_p c^+)$.
As we can see that except for the two end-points $v_A(0) = 0$ and $v_A(1) = 1$, the representative values $v_A(c^-)$, $v_A(W)$ and $v_A(c^+)$ are inner points corresponding of similarity classes level 1 $S(c^-)$, S(W) and $S(c^+)$.

(2) Similarity level 2: k = 2, we build partitions similarity classes of level 2. Such, on a blurry interval level 2, $I(h_i c^+) = (v_A(\Phi h_i c^+), v_A(\Sigma h_i c^+)]$ with

the two blurring interval of adjacent is $I(h_{i-1}c^+)$ and $I(h_{i+1}c^+)$, We will have form equivalence classes the following: $S(h_ic^+) = I(h_ic^+)\,[I(h_ph_ic^+) \cup I(h_{-q}h_ic^+)]$, $S(\Phi h_ic^+) = I(h_{-q}h_{i-1}c^+) \cup I(h_{-q}h_ic^+)$ and $S(\Phi h_ic^+) = I(h_ph_ic^+) \cup I(h_ph_ic^+)$, with i so that $-q \leq i \leq p$ and $i \neq 0$.
By similarity, can the construction of partitions the same classes any level k. However, in reality the application according to [6], $k \leq 4$, that is, there is maximum of 4 hegdes consecutive impacts the element up primitive c^- and c^+. The value of clear and translucent called has similar in the level k if the value represented by them along in a similar class in the level k.

2.3 Fuzzy Object-Oriented Database

Real-world entity applications or abstract concepts are often fairly complex objects. These objects contain a certain set of information on objects and behavior on the basis of the information it. Object attribute and its value to determine information about the object. The case for the value of this: (1) clear values: usually values are the values of the primitive data type such as string or number, or is the set of primitive values; (2) blurry value: this blur value is complex, the language label is used to demonstrate the value of this type. For example, the value of the height attribute of an object is said to be height about 180 cm tall, or maybe a language value "high ability"; (3) Object: in this case usually attributes value can refer to an object another. The object that it refers to may be blurred; (4) Collection: usually this attribute value is the set of values or objects. The inaccuracies of this attribute is the set can blur, or a member of the set is the value of the blur or blur objects.

Thus, an object is blurred because of the lack of information or incorrect information is caused by the value of that attribute the information incorrect, unclear, which collectively fuzzy information.

In FOOD model, a class is defined as a set of properties including attributes and methods for determining objects of this class, each method is represented as an operation function on the object's attribute values. On the other hand, attribute values are imprecise (or fuzzy), so methods for determining this class's objects also become fuzzy and uncertain.

The class to be reviewed is blurred caused by the following: (1) a some of objects of a class are determined can be blurry. (2): the domain of an attribute that is blurred, so a translucent class is formed when class definition this. (3): the class is a translucent class when it is inherited from one or more superclass, in which at least one superclass is a translucent class.

CLASS *class name*
  INHERITES
    class name 1 WITH LEVEL OF level 1
    ...
    class name n WITH LEVEL OF level n
  ATTRIBUTES
    attribute 1: [FUZZY] DOMAIN dom 1: TYPE OF type 1
    ...

attribute n: [FUZZY] DOMAIN dom n: TYPE OF type n
METHODS
...
END

3 Evaluation of Duplicates in Fuzzy Objects

A basic task of algebraic operations is used to determine the semantic relationship between two objects and evaluate if they are duplicates. In this part, we will present methods of evaluation and handling of redundant blurry objects.

3.1 Approximation Level k

Based on the concept of neighborhood level k, the paper offers a definition of approximation level k of the object attributes. Approximation level k is defined as follows:

Definition 4. *Let fuzzy class C defined on the set of attributes Attr(C) =* $a_1, a_2, \ldots, a_n$ *and methods M,* $o_1, o_2 \in C$*. We say that* $o_1.a_i$ *is approximation level k* $o_2.a_i$ *and denoted by* $o_1.a_i \approx_k o_2.a_i$ *if* $o_1.a_i$ *and* $o_2.a_i$ *belong to the equivalence class* $FRN_k(fu)$*. Where* $FRN_k(fu)$ *is a partition interval of the equivalence classes level k.*

Example 1: Suppose that in FOOD, a class named Employees, with a fuzzy attribute *salary*, the values of attribute *salary* are corresponding to one of four objects of class Employees are o_1.salary = high; o_2.salary = 80; o_3.salary = 70; o_4.salary = 90. The neighborhood system is built as follows:

Consider the hedge algebra of the linguistic variable *salary*, where D_{salary} = [0, 100], the generators are {0, low, W, high, 1}, the set of hedges is little, possibly, more, very, $FD_{salary} = H_{salary}(\text{high}) \cup H_{salary}(\text{low})$.

Choose fm(high) = 0.60, fm(low) = 0.40, μ(possibly) = 0.15, μ(little) = 0.25, μ(more) = 0.25 and μ(very) = 0.35. [0, 100] is partitioned into five intervals similar *level 1* as follows: fm(very high) * 100 = 0.35*0.60*100 = 21. So S(1)*100 = (79, 100]. (fm(posibly high) + fm (more high))*100 = (0.25*0.60 + 0.15*0.60)*100 = 24, so S(high) = (55, 79]. (fm (little low) + fm (little high))*100 = (0.25*0.60 + 0.25*0.40) * 100 = 25, so S(W) = (30, 55]. (fm (posibly low) + fm (more low)) * 100 = (0.25*0.40 + 0.15*0.40)*100 = 16, so S (low) = (14, 30], and S(0)*100 = [0, 14].

Since then, we have the neighborhood *level 1* of the equivalence classes as follows: $FRN_1(0)$ = [0, 14], $FRN_1(\text{low})$ = (14, 30], $FRN_1(\text{W})$ = (30, 55], $FRN_1(\text{high})$ = (55, 79] and $FRN_1(1)$ = (79, 100].

So, we say that o_1.salary $\approx_1$ o_3.salary because o_1.salary = high $\in FRN_1(\text{high})$ and o_3.salary = 70 $\in FRN_1(\text{high})$; or o_2.salary $\approx_1$ o_4.salary because o_2.salary = 80 $\in FRN_1(1)$ and o_4.salary = 90 $\in FRN_1(1)$. With level k = 1.

3.2 Redundant Fuzzy Objects

In the precise object-oriented database, an object is considered to be redundant if and only if it is duplicated completely with another object. But in the FOOD model, due to the object is fuzzy, so to evaluate the redundancy of two fuzzy objects o_i and o_j, the paper offers the following definitions:

Definition 5. *Let fuzzy class C with the set of attributes* $a_1, a_2, \ldots, a_n$. *Let two objects* o_i *and* o_j *in the fuzzy class C, k is the partition level and* $i \neq j$ *considered to be redundant with respect to each other if* $\forall k = 1, 2, \ldots, n$, $\forall o_i.a_k \; \exists o_j.a_k$: $o_1.a_i \approx_k o_2.a_i$, *and otherwise. Use denoted* $o_1 \approx_k o_2$ *to say that* o_i *is redundant to* o_j *based on the partition level k, where* $k = k_1, k_2, \ldots, k_n$.

Example 2: Give a fuzzy class C with the set of attributes Attr(C) = {name, age}, and o_1(C) = <name: An, age: 18>; o_2(C) = <name: Binh, age: young>; o_3(C) = <name: Huong, age: 32>; o_4(C) = <name: Nhan, age: 34>.

Suppose k ={0, 1}. Meanwhile, the level of partition for attribute *name* is a k = 0 and k = 1 for the attribute *age*. That is, we only construct the partition level for fuzzy attribute of the object class.

Consider the hedge algebra of the linguistic variable *age*, where D_{age} = [0, 100], the generators are {0, young, W, old, 1}, the set of hedges is {little, possibly, more, very}, $FD_{age} = H_{age}(\text{young}) \cup H_{age}(\text{old})$.

Choose fm(young) = 0.4, fm(old) = 0.6, μ(possibly) = 0.25, μ(little) = 0.2, μ(more) = 0.15 and μ(very) = 0.4. [0, 100] is partitioned into five intervals similar level 1, and as the same way of calculation in Example 1 the intervals are as follows S(0) = [0, 16), S(young) = [16, 32), S(W) = [32, 52), S(old) = [52, 76), and S(1) = [76, 100].

Since then, we have the neighborhood level 1 of the equivalence classes as follows: $FRN_1(0)$ = [0, 16), FRN_1(young) = [16, 32), FRN_1(W) = [32, 52), FRN_1(old) = [52, 76) and $FRN_1(1)$ = [76, 100].

We have, $o_1.age$ = 18, $o_2.age$ = young $\in FRN_1$(young), and $o_3.age$ = 32, $o_4.age$ = 34 $\in FRN_1$(W). It is easy to see that $o_1 \approx_1 o_2$ and $o_3 \approx_1 o_4$, that is o_1 is redundant to o_2 and o_3 is redundant to o_4 with the partition level k= 1.

To remove the redundant fuzzy objects by the partition level k in class C, we combine the redundant objects together until there are no longer two fuzzy objects which are redundant to each other.

Let o_i and o_j are two redundant objects level k in the class C, to remove these redundancies, we will combine o_i and o_j into a new object o. There are three types of combination for fuzzy objects to meet different requirements in the object manipulations.

$$\begin{aligned} o = merge_{\cup_k}(o_i, o_j) &= < merge_{\cup_k}(o_i.a_1, o_j.a_1),\ merge_{\cup_k}(o_i.a_2, o_j.a_2), \\ &\quad \ldots, merge_{\cup_k}(o_i.a_n, o_j.a_n) > \\ o = merge_{-_k}(o_i, o_j) &= < merge_{-_k}(o_i.a_1, o_j.a_1),\ merge_{-_k}(o_i.a_2, o_j.a_2), \\ &\quad \ldots, merge_{-_k}(o_i.a_n, o_j.a_n) > \end{aligned}$$

$$o = merge_{\cap_k}(o_i, o_j) \ =< merge_{\cap_k}(o_i.a_1, o_j.a_1),\ merge_{\cap_k}(o_i.a_2, o_j.a_2), \ldots, merge_{\cap_k}(o_i.a_n, o_j.a_n) >$$

In this paper, the object's fuzzy attributes are considered as the linguistic attributes and represented by structure of hedge algebra. Since then, we can construct the partition of equivalence classes level k for the linguistic attributes. Determine the linguistic value of the equivalence class level k and thus determine [a, b] corresponding to the attribute values.

Thus, the combinations of attribute values of two objects are the intersection, composition and subtraction on the intervals.

- $merge_{\cup_k}(o_i.a_1, o_j.a_1)$: union two intervals $o_i.a_1$ = [a, b] and $o_j.a_1$ = [c, d].
- $merge_{\cap_k}(o_i.a_1, o_j.a_1)$: intersect two intervals $o_i.a_1$ = [a, b] and $o_j.a_1$ = [c, d].
- $merge_{-_k}(o_i.a_1, o_j.a_1)$: subtract two intervals $o_i.a_1$ = [a, b] and $o_j.a_1$ = [c, d].
  At this point, we have the following cases:
  1. if c $\in$ [a, b], and d $\notin$ [a, b] then result of subtraction is [a, c].
  2. if c $\notin$ [a, b], and d $\in$ [a, b] then result of subtraction is [d, b].
  3. if [c, d] $\subset$ [a, b] then result of subtraction is [a, c] $\cup$ [d, b].
  4. if [c, d] $\cap$ [a, b] = $\emptyset$ then result of subtraction is [a, b].

4 Fuzzy Algebraic Operations

We will present the blurry algebraic operations for blurry classes based on the semantic neighborhood of algebra of the hedge. This paper divides the blurry algebraic operations on FOOD model into two categories: algebraic operations for blurry classes and algebraic operations for blurry objects.

In order to define the algebraic operations for fuzzy objects and fuzzy classes, we first introduction some notations being used below. Let C is a class with attributes $\{a_1, a_2, \ldots, a_n\}$ and denoted by Attr(C), and Attr'(C) is the set of attributes which is obtained from the construction of the partition to determine the equivalence classes for the attribute in the Attr(C). Class C contains the set of (fuzzy) objects, denoted by C = $\{o_1, o_2, \ldots, o_n\}$, and o(C) is the object o of class C.

4.1 Algebraic Operations for Fuzzy Objects

The algebraic operation for fuzzy objects is eventual the fuzzy selection. A selection operation refers to such a procedure that the objects of the classes satisfying a given selection condition are selected. Let C is a fuzzy class, P_f is a fuzzy predicate and denoted by a selection condition and k is the partition level. The selection of P_f in C with the partition level k is defined as follows:

$$\sigma_{P_f}(C) = \{o(C) | o(C) \wedge P_f(o)\}$$

4.2 Algebraic Operations for Fuzzy Classes

4.2.1 Fuzzy Product

Fuzzy product of C_1 and C_2 is a new class C, which is composed of these general attributes of C_1 and C_2, as well as member attributes of C_1 and C_2. Generally, it is required that Attr'(C_1) ∩ Attr'(C_2) = ∅ in the fuzzy product. The objects of class C are generated from the combination of objects from class C_1 and C_2, in which the class C contains attributes Attr'(C_1) and Attr'(C_2).

$$\begin{aligned} C = C_1 \times_k C_2 \ &= \{o(C)|(\forall o_1)(\forall o_2)(o_1(C_1) \wedge o_2(C_2) \wedge o[\mathrm{Attr}'(\mathrm{C}_1)] \\ &= o_1[\mathrm{Attr}'(\mathrm{C}_1]) \wedge o[\mathrm{Attr}'(\mathrm{C}_2)] = o_2[\mathrm{Attr}'(\mathrm{C}_2)])\} \end{aligned}$$

4.2.2 Fuzzy Join

For two fuzzy classes C_1 and C_2 with Attr'(C_1) ∩ Attr'(C_2) ≠ ∅ and Attr'(C_1) ≠ Attr'(C_2). Then join between C_1 and C_2 will form a new class C, have Attr'(C) = Attr'(C_1)∩(Attr'(C_2) - (Attr'(C_1)∩Attr'(C_2))). The objects of class C are created by the composition of the objects from C_1 and C_2, whose semantics are equivalent on Attr'(C_1)∩Attr'(C_2) according to a given partition level, Then:

$$\begin{aligned} C \ &= C_1 \triangleright\triangleleft_k C_2 = \{o(C)|(\exists o_1)(\exists o_2)(o_1(C_1) \wedge o_2(C_2) \wedge o[Attr'(C_1) - (Attr'(C_1) \\ &\quad \cap Attr'(C_2))] = o_1[Attr'(C_1) - Attr'(C_1) \cap Attr'(C_2))] \wedge \\ &\quad o[Attr'(C_1) \cap Attr'(C_2)] = merge_{\cap_k}(o_1[Attr'(C_1) \cap Attr'(C_2)], \\ &\quad o_2[Attr'(C_1) \cap Attr'(C_2)]) \wedge o[Attr'(C_2) - (Attr'(C_1) \\ &\quad \cap Attr'(C_2))] = o_2[Attr'(C_2) - (Attr'(C_1) \cap Attr'(C_2))]\} \end{aligned}$$

4.2.3 Fuzzy Union

The fuzzy union between C_1 and C_2 requires Attr'(C_1) = Attr'(C_2), which implies that all the corresponding attributes in C_1 and C_2 must be completely similar. Let a new class C is the fuzzy union of C_1 and C_2, and the objects of the class C are composed of three kinds of objects: the first two kinds are the objects are such objects that directly come from one componed class (for example, C_1) and are not redundant to any objects in another component classes (for example, C_2). Final objects are the objects that are the resulted of merging the redundant objects from two component classes, with k is the partition level. We have:

$$\begin{aligned} C = C_1 \cup_k C_2 \ &= \{o(C)|(\forall o_2)(\exists o_1)(o_2(C_2) \wedge o_1(C_1) \wedge o = o_1) \\ &\quad \vee(\forall o_1)(\exists o_2)(o_1(C_1) \wedge o_2(C_2) \wedge o = o_2) \\ &\quad \vee(\exists o_2)(\exists o_1)(o_1(C_1) \wedge o_2(C_2) \wedge o = merge_{\cup_k}(o_1, o_2)\} \end{aligned}$$

4.2.4 Fuzzy Subtraction

The fuzzy subtraction of C_1 and C_2, also requires Attr'(C_1) = Attr'(C_2), which implies that all the corresponding attributes in C_1 and C_2 must be completely

similar. Let a new class C is the fuzzy subtraction of C_1 and C_2, and k is the partition level. Then we have

$$C = C_1 -_k C_2 = \{o(C) | (\forall o_2)(\exists o_1)(o_2(C_2) \wedge o_1(C_1) \wedge o = o_1) \vee (\exists o_1)(\exists o_2)(o_1(C_1) \wedge o_2(C_2) \wedge o = merge_{-_k}(o_1, o_2)\}$$

4.2.5 Fuzzy Intersection

The fuzzy intersetion of C_1 and C_2 is to combine the common objects of these two classes, which requires Attr'(C_1) = Attr'(C_2), which implies that all corresponding attributes in C_1 and C_2 must be completely similar. Let a new class C is the fuzzy intersection of C_1 and C_2, and k is the partition level. We have

$$C = C_1 \cap_k C_2 = \{o(C) | (\exists o_2)(\exists o_1)(o_1(C_1) \wedge o_2(C_2) \wedge = merge_{\cap_k}(o_1, o_2)\}$$

4.2.6 Fuzzy Projection

Let a class C' and S are a subset of the set of attributes of class C'. A new class C is formed from the projection of C' on S is to remove the attributes Attr(C') - S from C' and only retain the attribute S in C'. It is clear that S $\subset$Attr(C') and Attr(C) = S. Every object in C' becomes a new object, whose set of attributes only consists of attributes S and remove the attributes Attr(C') - S. Obviously, there may be redundancy in new objects. After removing the possible redundancies, the new objects constitute class C. The projection of C' on S is defined as follows:

$$C = \prod {}_S^k(C') = \{o(C) | (\forall o')(o'(C') \wedge o[S] = o'[S] \wedge o = merge_{\cup_k}(o[S])\}$$

4.3 Fuzzy Queries

The handling of queries in the object-oriented database refers to the method such that the objects that meet a certain condition are selected and distributed to the user according as formats the required. The format of the request includes the attributes that appear in the result and if the result is a group then the attributes will sort in an. A query can be viewed as containing two components that are query conditions and requested formats. In the interest of simple illustration, the formatting requirements will be ignored. An object-oriented query language (OQL) has regular structure like SQL query and is described as follows:

SELECT <list of attributes/methods> **FROM** <list of classes> **WHERE** <query conditions>

In which, <list of attributes/methods> lists the attributes (methods) will appear in the query results, this list has at least one attribute (method). The attributes (methods) in <list of attributes/methods> are select from the associated classes which are specified in the **FROM** statement <list of classes> contains the class names separated by commas: $class_1$, $class_2$,..., $class_n$, from which the attributes/methods are selected with the **SELECT** statement.

<Query conditions> is a logical expression, they always result in truth (1) or false (0), this is precise query conditions.

From above query structure, we can see that classical database suffer from a lacks of flexibility to query. The given query condition and the contents of the database are all crip. A query is flexible if the query condition is imprecise and uncertain information. For example, consider the following query *"show name all the students of possibly young age"*, and in this query *possibly young age* is fuzzy query condition.

Thus, a query in the fuzzy object-oriented database is structured as follows:

SELECT <list of attributes/methods> **FROM** <list of classes> **WHERE** <fuzzy query conditions>

The structure of this fuzzy OQL query is an extension of the structure OQL. Here, <fuzzy query conditions> is a fuzzy condition or combination of fuzzy conditions using the selection and association calculations.

As analyzed above, the domain of the fuzzy attributes of the classes and objects is very complex and can get the values such as numeric values, interval values or linguistic values. Then the identification of objects satisfying the query conditions is matching the attribute values of the objects and the fuzzy query conditions.

Consider a fuzzy class Student with the set of attributes Attr(Student) = name, age, and three objects: o_1(Student) = <name: an, age: young>; o_2(Student) = <name: binh, age: 34>; o_3(Student) = <name: department, age: middle-aged>

Assume have question as follow: Indicate the name the objects of class sinhvien have age is *young*.

Using the algebraic operations above, we can answer this question:

$$\prod^{1}_{student.name}(\sigma_{student.age=young})$$

And, OQL statement corresponding: **SELECT** Student.name **FROM** Student **WHERE** Student.age = young.

This query will return all objects of the class Student that satisfy the query conditions is young. Since the age query condition is young, we should only construct a partition level k = 1 for this attribute age. From example 2, we have the neighborhood level 1 of the equivalence classes for the attribute age as follows: $FRN_1(0)$ = [0, 16), FRN_1(young) = [16, 32), FRN_1(W) = [32, 52), FRN_1(old) = [52, 76) and $FRN_1(1)$ = [76, 100].

The received as a result: o_1(an, young), because o_1.young $\in FRN_1$(young).

5 Conclusion

To base oneself on the semantic quantity the algebra of the hedge, the paper proposes the definition of approximate the level of k of attribute values. Founded on that foundation, the paper proposed the definition of redundant objects and

three operations for combining objects to get rid of the redundancy. It also gave the definition of blurry algebraic operations for blurry classes and blurry objects. It also shows the methods for processing blurry queries in FOOD model. The next chapters present case studies. We will proceed to extend blurry queries on the blurry object-oriented database such as preference queries, keyword queries, ranking queries, and so on.

References

1. Cao, T.H., Nguyen, H.: Uncertain and fuzzy object bases: a data model and algebraic operations. Int. J. Uncertainty: Fuzziness Knowl.-Based Syst. **19**(2), 275–305 (2011)
2. De Caluwe, R., Van Gyseghem, N., Cross, V.: Basic notions and rationale of the integration of uncertainty management and object-oriented databases. In: Fuzzy and Uncertain Object-Oriented Databases, pp. 1–20 (1997)
3. Van Ban, D., Tuan, T.C., Van Thang, D.: Querying data with fuzzy information in object-oriented databases based on hedge algebraic semantics. In: Proceedings of the 4th International Conference on Knowledge and Systems Engineering, Danang, pp. 39–45 (2012)
4. Ho, N.C.: Quantifying hedge algebras and interpolation methods in approximate reasoning. In: Proceedings of the 5th International Conference on Fuzzy Information Processing, Beijing, pp. 105–112 (2003)
5. Ho, N.C.: Fuzzy set theory and soft computing technology. Fuzzy system, neural network and application. Publishing science and technology, 37–74
6. Nguyen, C.H., Wechler, W.: Hedge algebras: an algebraic approach to structures of sets of linguistic domains of linguistic truth variable. Fuzzy Sets Syst. **35**(3), 281–293 (1990)
7. Ho, N.C.: A topological completion of refined hedge algebras and a model of fuzziness of linguistic terms and hedges. Fuzzy Sets Syst. **158**, 436–451 (2007)
8. Yan, L., Ma, Z.M., Zhang, F.: Algebraic Operations in Fuzzy Object-oriented Databases. @springer Science + Business Media, Berlin (2012)
9. Ma, Z.M.: Advances in Fuzzy Object-Oriented Databases: Modeling and Applications. Idea Group Publishing, Hershey (2004)

Context-Adaptive Values-Based Games for the Young: Responsible Decision Making for a Sustainable World

Khushbu Tilvawala(✉), David Sundaram, and Michael Myers

Information Systems and Operations Management, University of Auckland, Auckland, New Zealand
{k.tilvawala,d.sundaram,m.myers}@auckland.ac.nz

Abstract. This research aims to design and implement a framework and system platform for creating values-based digital games for the young. Studies show that many digital games include violence and about half of the violent incidents have negative repercussions such as increased aggression in the "real" world. This influential nature of digital games calls for research on ways in which games can be leveraged. One way is to use digital games to foster positive values and sustainable practices. This is especially relevant in the context of young players since they are malleable. Digital games present opportunities for learning via decision-making. When players interact with games, they go through several decision-making cycles, each time, subconsciously reviewing the consequences of their actions to guide future decisions. This research will use a multi-methodological approach combining qualitative and design science research to explore and create a platform to design and implement context-adaptive values-based digital games.

Keywords: Context-adaptive · Values · Digital games · Young
Responsible decision-making · Sustainable

1 Introduction

Digital games, particularly those classified as serious games are a prevalent mode of learning. Researchers find that these games are becoming increasingly "influential" and "persuasive" in our real world [1–8]. Values and sustainable practices are complex yet essential concepts that are learnt experientially as they are context specific. The possibility of using digital games to teach these would be very valuable given their importance to every individual, albeit subjective. Given that our value systems and beliefs are primarily moulded when we are young [3, 9], it would be useful to explore the potential of digital games in teaching these concepts to the young (ages 0–8 defined as early childhood) [10]. Preliminary research in this area suggests that research on digital games and the early childhood context is limited when compared to other age groups [11, 12]. Research on values in digital games, particularly for specific contexts is also limited. Lastly, research on early childhood and values exists in education and psychology literatures, however, research on the intersection of all these topics is scarce. In particular, we have the following interrelated research gaps and opportunities:

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 135–144, 2018.
https://doi.org/10.1007/978-3-319-77818-1_13

1. There is a paucity of research on the design of values-based digital games, especially for specific contexts [13, 14] such as early childhood.
2. Research on the implementation, particularly, platforms for creating context specific (early childhood) values-based digital games is lacking.

The philosophical assumptions guiding this research are a combination of interpretive and critical research. A multi-methodological approach has been outlined combining qualitative with design science methods. This will involve an exploration and first-hand review of existing digital games for the young, specifically games focusing on values and/or sustainable practices. The exploration and review will then be combined with design science research to propose a platform for designing and implementing context-adaptive digital games that can foster values and sustainable practices in the young [15, 16]. We also plan to evaluate the proposed platform by using it to create context-adaptive, values-based games for the young. The motivation for this research is illustrated in Fig. 1 [3–5, 7, 17].

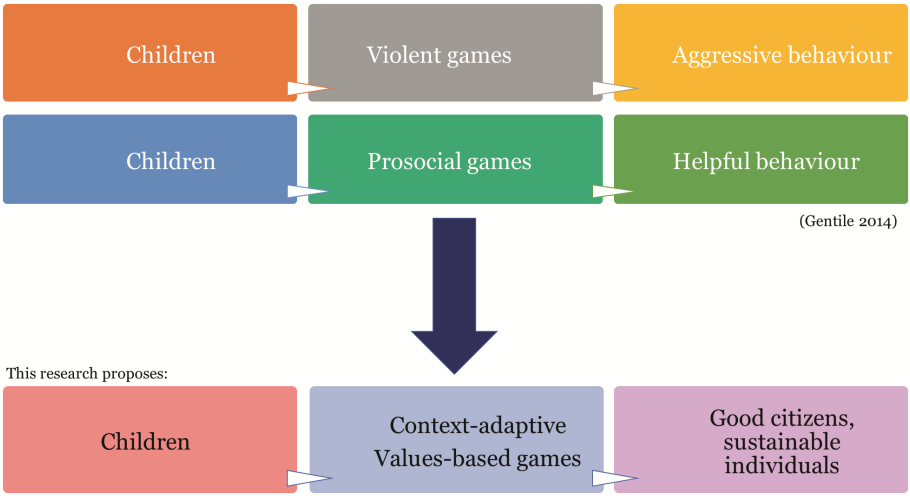


Fig. 1. Motivation for context-adaptive values-based games for the young

In the following sections, we discuss: values and sustainability in the context of this research; the context-adaptive nature of values; the role of decision-making in learning values and sustainable practices, and some preliminary findings on digital games and research in this area. We conclude with a summary of our preliminary research, challenges in this research, possible contributions and future research opportunities in this area.

2 Values and Sustainability

As a society, we have been passing down values and whatever we believe to be good practices for a sustainable life over the generations. The best way to teach these is through experience. Most of us can agree that this learning is therefore highly

unstructured, tacit, context-adaptive, subjective and complex. Learning these concepts also takes a lot of time. For this research, we use Rokeach's [18] view of a value as "a centrally held, enduring belief which guides actions and judgements across specific situations and beyond immediate goals to more ultimate end-states of existence". For example, the value "do not harm". While most of us can agree on this universal value, the context matters. If one "harms" in the process of defending themselves or their loved ones, the action may be justifiable despite the value.

When we talk about sustainability or being sustainable, we mean practices that enable the survival or preservation of existing systems or processes. The main types of sustainability are: *personal* – practices or beliefs that enable individuals thrive in the world; *social* – practices or beliefs regarding one's engagement with people in their surroundings, *economic* – beliefs regarding a country's survival and upkeep; *cultural* – practices or beliefs regarding the preservation of cultural aspects such as customs, language, dressing style; and *environmental* – refer to one's beliefs about maintaining and preserving the environment for the wellbeing of both the current and future generations. We consider the two concepts as being closely linked. In fact, being sustainable may also be considered a value. We therefore use the word "values" in a more holistic sense to include good sustainable practices.

The complex nature of values has also led to several controversies on the ideal time to teach these. In this research, we side with research that suggests that most individuals develop their value system, habits, and attitudes in their early years when they are most malleable [19]. In fact, many researchers find that a great part of learning values takes place during "play" [20–22], which is an essential component in a child's upbringing. In today's digital age, "play" is facilitated by technology in the form of digital games, making digital games a powerful means of fostering values, attitudes, and developing social and emotional learning in children [23]. In the next section, we discuss the potential of digital games in fostering values, particularly, the relevance of viewing this from a decision-making lens.

3 Responsible Decision-Making in Games

One way of understanding how digital games support learning, especially of complex concepts such as values, ethics and sustainable practices, is by viewing this process from a decision-making lens. This is because responsible decision-making is often viewed as a means to fostering good ethics and values [24]. This is because we are able to demonstrate possible choices for a scenario via decision-making. Further, decision-making may vary across contexts, in this case, decision-making in early childhood.

When one plays a game, the different components of the game and the holistic experience of playing the game gradually modify one's decision-making process. In fact, it may also subtly change one's mental model of decision-making. This is similar to the changes in our decision-making processes and models as we experience life in general [25]. For example, if a child is playing a game that involves different environments (a garden, a zoo, their early childhood centre) and they discover that they get rewarded if they decide to help someone in these different environments, they will always be looking

out for ways in which they can "help" to be rewarded. The opportunities to help can be provided via choices, where certain choices have positive consequences. The positive consequence may be someone reciprocating their good action later in the game or someone commending them on their positive actions.

The complex nature of values also calls for the need to acknowledge the context of the decision. For example, the value of "helping". When parents ask a child to pack their toys after them, they would not like to endorse is as the value of "helping" but rather the value of "cleaning up after them". Having context-adaptive games would therefore be appropriate in teaching values.

The decision-making process involved in the example above is typical in many other games (regardless of context) and demonstrates that as one progresses through the game, and as they play it more, they become better at it. This improvement can be tracked by the decisions and choices being made through the game. In majority of the games, these decisions have an impact on a game event later in the game. This may be compared to the five stage model of skill acquisition [26]. The decision-making process in the example above also demonstrates the interrelated nature of decisions in our life. Our decisions are often linked, where a certain decision will lead to a completely different pathway compared to a different decision. Langley's model of decision linkages [27] illustrates this interrelated nature of decisions, and may be a valuable tool when considering the design of values-based digital games. Overall, the "sustainable" aspects of decision making are inherent not only in the nature of decisions posed, but also in the actual process of decision-making and mental model formation.

The next section explores existing research and values-based digital games for the young. In particular, we assess specific decision-making aspects or their lack of in these games. The purpose of this is to explore any opportunities to leverage these existing games to "experience" and "learn" values and sustainable practices virtually.

4 Context-Adaptive Games for Inculcating Values in the Young

Digital games have progressed from teaching basic literacies to more complex, tacit skills and concepts such as ethics in business. In fact, in many instances, digital games foster certain implicit values that they were never explicitly designed to teach. While this research may contribute to other contexts, our focus is on the early childhood (the young) context [10] since this is a phase when most of us develop our value systems, ethics, behaviors, habits and attitudes; and a time when our mental models for decision-making are fostered.

We explored and reviewed a handful of digital games designed to teach values in the young (Table 1). Our initial findings show that there are some games that teach values and sustainable practices using responsible decision-making. However, there is little research on the analysis of these games [13, 14].

In terms of the design and implementation of these games, there are only a few platforms such as Common Sense Media and Gameful that may support the creation of values-based digital games. The platforms are however not specific to the young. The platforms also do not explicitly cite any application of decision-making theories, or

allow for creation of context-adaptive games. Therefore, overall, platforms and research artefacts that support the creation of context specific (early childhood), responsible decision-making values-based digital games are lacking or non-existent.

Table 1. Examples of values-based digital games for the young

Games	Values
Kokoro	A virtual global community where the goal of the players is to help a society work together by making ethical decisions throughout the game. Some values described in the game include solving conflicts, challenging prejudices and stereotypes, respect, and improving relations. The overarching goal of the game is to promote unity and peace in the world
Minimonos[a]	A virtual world where kids create their monkey avatars and play mini games. The game is set nature, encouraging values of looking after the environment and planet
E-Critter game	This game is aimed at teaching children financial responsibility and personal values to children. It also teaches the values of living within your means (being sustainable financially) and separating needs from wants. Decision-making is not explicitly mentioned but it is clearly at the heart of this game's mechanisms
Peek-a-Zoo	Teaches kids (ages 1–5) emotions and behaviors. Decision-making is not evident in this game, but could easily be embedded given its nature
Cool School: Where peace rules	Teaches kids how to resolve conflicts and reduce bullying. The players are empowered to make decisions and choices in terms of their locations and their solutions, and they're rewarded for making good choices
Little Green Island	A game about environmental conservation. Kids are presented with environmental issues and are required to solve these

[a]This game was closed in 2013.

The aim of this research is therefore to propose both theoretical and system artefacts for the design and implementation of context-adaptive, values-based digital games for the young. We plan to use theories on decision-making processes to architect the context-adaptive aspects of the game creation platform. The next section outlines examples of the types of games we aim to create using the platform.

5 Design of Context-Adaptive Values-Based Games

Some researchers have explored game mechanisms and components for values-based design of games [28–31]. This includes a generic framework for designing games with specific values [13]. An important consideration that researchers have discussed is the importance of having a "fun" element in serious games [32]. Last, but not least, designing for learning calls for a way to measure or validate the learning. Applying decision-making and making the decisions context specific not only guides the game design process, but also helps validate the learning by providing multiple versions of similar decisions across a period of time [26, 27]. With these in mind, the goal of this research

is to design and implement a platform for designing digital games for the young. Further, some researchers have acknowledged the need to assess values in digital games for the early childhood context and proposed a framework for identifying these values [21]. We plan to consider these in the evaluation of the game platform and games we design. We have come up with two scenarios that apply decision-making in fostering values.

5.1 Scenario 1: Context-Adaptive Values-Based Snakes and Ladders

The first scenario has been designed implemented non-digitally. It was created as a mock-up for a potential context-adaptive values-based digital game based on the popular board game Snakes and Ladders.

The game was used as an activity to encourage positive values in 5–6 year olds by taking them up the ladder for eliciting a positive value, and down a snake for a negative value. The activity was planned such that kids played on a live sized snakes and ladders board, and the respective values were read out to the kids when they landed on a snake or ladder. Table 2 outlines the list of values read out to the children. The game explicitly considers the context of the young players as well as the context of the decision/scenario. The context of the purpose of the game is also considered.

Table 2. Values-based Snakes and Ladders

Ladders	Snakes
1. You waited till mummy and daddy finished talking before you asked your question (Value: Respect/patience/discipline)	1. You broke a plate and lied that you didn't do it. (Value: Truth – kids need to know that it is safe to tell the truth and a mistake is okay)
2. You have learnt a new non-native cultural song. Well done! You respect and care for other cultures (Value: Unity)	2. You did not finish the food on your plate (Value: Being thankful, do not waste food)
3. You were brave to stand in front of the class and talk (Value: Self-confidence/courage)	3. You forgot to say your prayers before going to bed (Value: Thankfulness, appreciation, devotion)
4. You helped daddy water the plants (Value: Gentleness/caring for nature)	4. You pushed your brother/sister/friend (Value: Right-conduct)
5. You took your dog for a walk (Value: Responsibility)	5. Oops you forgot to brush your teeth (Value: Loving yourself/cleanliness)
6. You gave mummy a big hug and said thank you for the food (Value: Gratefulness)	6. You didn't sleep on time. You slept late and could not get up on time for school (Value: Discipline)
7. You said sorry to your brother/sister for breaking their favorite toy (Value: Right-conduct)	7. You did not listen to mum and watched TV for 2 h and didn't finish your homework. (Value: Respect/Discipline – Do not waste time)
8. I helped my friend get up when he/she fell (Value: Caring/helping)	8. You were not patient for your turn on the swing (Value: Understanding/Patience)

The snakes and ladders game was planned as a group, with teachers involved in facilitating various values-based activities outside the school curriculum. Careful consideration was made with the wordings of the negative values so as not to introduce the children to a negative value they didn't know of. Overall, the activity was enjoyed thoroughly by the young players and was used as a discussion topic in later sessions with the children.

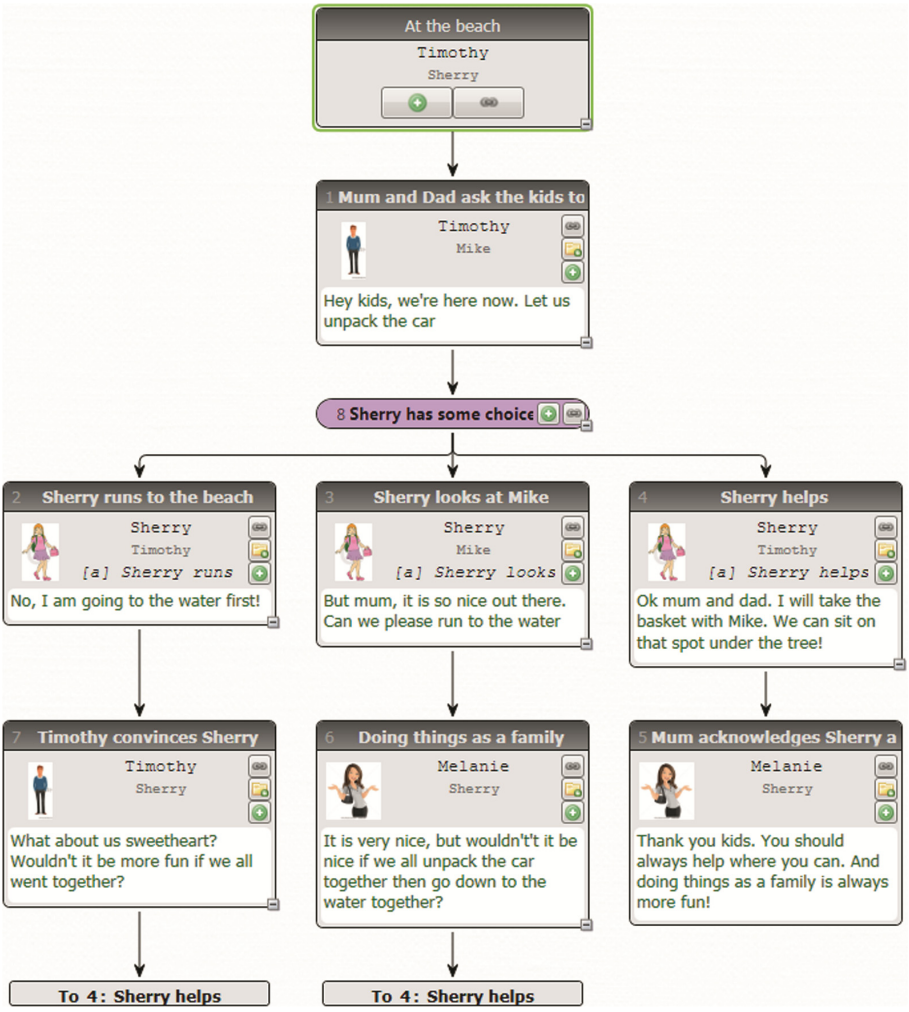


Fig. 2. Scenario 2 – Decision-driven Story of Helpfulness and Kindness created using ChatMapper

5.2 Scenario 2: Context-Adaptive, Decision-Driven Story of Helpfulness and Kindness

For the second scenario, we have used a storyline to pose various choices and invite the player to make decisions (Fig. 2). The player navigates through different pathways in the storyline to learn the underlying value behind the story. The storyline involves a decision map illustrating a child's potential response in a conversation with different members of the family. The main value in the storyline is the positive attitude of being helpful, kind, and doing things together as a family. The setting for the story is a family going to the beach. Figure 2 is a snapshot of the storyline, demonstrating some choices and pathways (decision map) in the storyline.

The goal of having this type of decision map is to come up with "before" and "after" scenarios around a specific value. The player's responses to each decision point can then be used to assess any changes because of the learning from their previous decisions.

6 Conclusion

This research applies decision-making theories to create and propose artefacts for the design and implementation of values-based digital games. A multi-methodological approach will be used to explore and create the artefacts.

As with any research, there are some challenges that we are trying to address as we progress through the research. These include designing serious games that are fun and engaging [32]. This is especially difficult with the plethora of digital games today, many of which are meaningless and mindless. Another challenge is to create a digital game that explicitly complements and promotes learning in the real world. While there is a benign relationship between real and virtual worlds [7, 29], we are aware that playing digital games may come at the cost of missing out on social interactions and play in the real world.

The cross-disciplinary nature of this research means that it will contribute to various fields of research that future researchers can expand on in their respective area(s) of expertise. These include, but are not limited to: digital games and responsible decision-making, digital games and early childhood, fun serious games, and values, ethics and sustainability in digital games. This research will contribute both to theoretical research on artefacts for designing values-based digital games, as well as system artefacts that may be used in practice. Future research may therefore build on the initial artefacts proposed in this research, as well as validate the proposed artefacts via experiments and implementations in different contexts. Our ultimate hope is that the artefacts will help bring about positive changes in individuals, families and organizations to foster a more sustainable, equitable and just society and world.

References

1. Chee, Y.S.: Embodiment, embeddedness, and experience: game-based learning and the construction of identity. Res. Pract. Technol. Enhanc. Learn. **2**(1), 3–30 (2007)
2. Prensky, M.: Digital game-based learning. Theor. Pract. Comput. Appl. Entertain. **1**(1), 21 (2003)

3. Gentile, D.: Violent video game effects on thinking and behaviour (2014). http://www.drdouglas.org/. Accessed 09 Nov 2015
4. Children Now. Talking with Kids™ about Violence (2015). http://www.childrennow.org/parenting-resources/violence/. Accessed 16 Nov 2015
5. APA. Technical violent games, Washington (2015)
6. Lieberman, D.A., Fisk, M.C., Biely, E.: Digital games for young children ages three to six: from research to design. Comput. Sch. **26**(4), 299–313 (2009)
7. Bogost, I.: Persuasive games: the expressive power of videogames. MIT Press, Cambridge (2006)
8. Anderson, C.A., et al.: Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: a meta-analytic review. Psychol. Bull. **136**(2), 151 (2010)
9. Tootell, H., Freeman, M., Freeman, A.: Generation alpha at the intersection of technology, play and motivation. In: 2014 47th Hawaii International Conference on System Sciences, pp. 82–90 (2014)
10. UNESCO: Early childhood care and education (2015). http://www.unesco.org/new/en/education/themes/strengthening-education-systems/early-childhood/
11. Hwang, G.-J., Wu, P.-H.: Advancements and trends in digital game-based learning research: a review of publications in selected journals from 2001 to 2010. Br. J. Educ. Technol. **43**(1), E6–E10 (2012)
12. Judge, S., Floyd, K., Jeffs, T.: Using mobile media devices and apps to promote young children's learning. In: Heider, K., Renck Jalongo, M. (eds.) Young Children and Families in the Information Age. EYC, vol. 10, pp. 117–131. Springer, Dordrecht (2015). https://doi.org/10.1007/978-94-017-9184-7_7
13. Flanagan, M., Nissenbaum, H.: Values at Play in Digital Games. The MIT Press, Cambridge (2014)
14. Martens, A., Diener, H., Malo, S.: Game-based learning with computers – learning, simulations, and games. In: Pan, Z., Cheok, A.D., Müller, W., El Rhalibi, A. (eds.) Transactions on Edutainment I. LNCS, vol. 5080, pp. 172–190. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69744-2_15
15. Hevner, A., March, S., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
16. Nunamaker, J.F., Chen, M., Purdin, T.: JMIS systems development in IS research. J. Manag. Inf. Syst. **7**(3), 89–106 (1991)
17. Mehallow, C.: Teaching sustainable values through serious gaming. http://www.triplepundit.com/2010/11/teaching-sustainable-values-serious-gaming/?doing_wp_cron=1415237815.1487360000610351562500
18. Rokeach, M.: The Nature of Human Values, vol. 438. Free press, Florence (1973)
19. Epper, R., Derryberry, A., Jackson, S.: Game-Based Learning (2012)
20. Farné, R.: Pedagogy of play. Topoi **24**(2), 169–181 (2005)
21. Verenikina, I., Harris, P.: Child's play: computer games, theories of play and children's development, pp. 99–106 (2003)
22. Verenikina, I., Kervin, L.: iPads, digital play and pre-schoolers. HeKupu **2**(5), 4–19 (2011)
23. Hromek, R., Roffey, S.: Promoting social and emotional learning with games: 'it's fun and we learn things'. Simul. Gaming **40**(5), 626–644 (2009)
24. Josephson Institute: Ethics (2017)
25. Craik, K.: The Nature of Explanation. Cambridge University Press, Cambridge (1943)
26. Dreyfus, S.E.: The five-stage model of adult skill acquisition. Bull. Sci. Technol. Soc. **24**(3), 177–181 (2004)

27. Langley, A., Mintzberg, H., Pitcher, P., Posada, E., Saint-Macary, J.: Opening up decision making: the view from the black stool. Organ. Sci. **6**(3), 260–279 (1995)
28. Abdullah, K.A., Ismail, Z.I.: Implementation of video games as a teaching aid for moral education subject in Malaysian kindergarten. In: 3rd Global Summit on Education, pp. 309–319, March 2015
29. Baranowski, T., Buday, R., Thompson, D.I., Baranowski, J.: Playing for real: video games and stories for health-related behavior change. Am. J. Prev. Med. **34**(1), 74–82.e10 (2008)
30. Belman, J., Nissenbaum, H., Flanagan, M., Diamond, J.: Grow-A-Game: a tool for values conscious design and analysis of digital games. In: Think Design PLay (2011)
31. Schrier, K., Gibson, D., IGI Global: Ethics and Game Design: Teaching Values through Play. Information Science Reference, New York (2010)
32. Iten, N., Petko, D.: Learning with serious games: is fun playing the game a predictor of learning success? Br. J. Educ. Technol. **47**(1), 151–163 (2016)

Applying and Deploying Cyber Physical System in Monitoring and Managing Operations Under Mines and Underground Works

Nguyen Thanh Tung[1(✉)], Vu Khanh Hoan[2], Le Van Thuan[2], and Phan Cong Vinh[2]

[1] International School, Vietnam National University Hanoi, Hanoi, Vietnam
tungnt@isvn.vn
[2] Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

Abstract. This paper presents results of a study on the current situation of monitoring underground gases in Vietnam, on the basis of which a model of cyber physical system will then be proposed in order to monitor and manage operations under mines and underground works. This system composes of: devices that measure Methane (CH4) [1] and Carbon monoxide (CO) [2] that are set up under mines in order to track these gases 24/7. The information received from these devices will be sent to on-ground stations and stored in the system database for collecting and managing information about gas detectors, warnings of fire and explosion, so timely warnings can be sent to workers under mines through the alarming system that is integrated in the detectors in case of unexpected incidents.

Keywords: Managing system of operations under mines and underground works
Methane detector · Carbon monoxide detector · Under-mine gas warning system

1 Introduction

In recent years, together with the general development of the whole country, the mine exploitation [3, 4] industry has gained an increasingly important role in Vietnam's economy. Over the past years, mineral exploitation has been accounted for up to 5.6% of Vietnam's GDP, largely contributing to the nation's renovation cause. However, mineral exploitation has also caused many bad effects on the environment including serious under-mine air pollution, health problems and even fatal accidents that underground workers are prone to. Therefore, enhancing work safety and productivity, preventing fire and explosion, and minimizing the waste of resources and energy are urgent requirements for the current mineral exploiting operation. That is why automatic devices that measure, monitor and alarm warning are more and more important both in reducing potential hazards and in timely rescue when an unexpected incident occurs. Moreover, they also support effective management and reduce the number of technical workers whose duties are simply to do measurements, keep records, do calculation and make reports.

At present in Vietnam, these conditions need to be improved in thousands of mines and underground works of all sizes.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 145–154, 2018.
https://doi.org/10.1007/978-3-319-77818-1_14

1.1 Developing a Cyber Physical System for Monitoring and Managing Operations Under Mines and Underground Works

This system comprises fixed and portable measuring devices, monitoring stations that receive data from wire and wireless connections, circuit switch that transmits data to monitoring software, and a database that stores the received information. These are made according to regulations on fire and explosion prevention under mines. The software installed in the central computer is highly practical, easy to access and use with such functions as collecting and displaying data on mine-route maps and updating current data of the map. It also allows setting up and adjusting the alarming levels, controlling, sending automatic commands to switch off and on the electric equipment under mines, making statistical calculations, storing information and making reports on the status of gases daily, weekly and monthly, supporting remote supervision via the internet.

The whole system including hardwares and softwares is designed with openness, and is adaptable to the adding or removing measuring devices and functions when needed.

Measuring devices are installed under mines to measure and give on the spot alarming signals so that workers can quickly move out of the dangerous areas. Other devices play an immediate role that transmit information, distribute and process data so that the system can be operated synchronously and continuously on a 24/7 basis without interrupt or mistakes.

The software installed at the server: manage and monitor the measured data, set up alarming levels and power supply cut levels, set up power supply cut. Displaying measured data on the mine map, send commands to cut power to devices under mines when the concentration of gases exceeds dangerous levels. The database located at server can store and manage all information received from detectors chronically.

1.2 New Technologies Applied in the Project

Figure 1 shows the system structure and the diagram of connections of a centrally and automatically methane monitoring system. The system consists of 2 parts: on-ground and underground.

On-ground part: via the software of the system, the server can set up arguments for the system such as system configuration, manage the detectors, set alarming levels, set power supply cut levels, collect data from fixed and portable detectors underground, display the data of mine route map, send automatic alarming signals when the concentration of gases exceeds the set levels, send partial or interchannel power supply cut command, make statistic reports, store and search for data, update the map according to current status of exploitation under mines, allowing remote supervising via the internet. Moreover, on the ground there are data storing devices and a backup computer that automatically switches on backup mode when the server encounters an incident to ensure the continuity of the system. The circuit-switcher on the ground has the functions of transmitting and distributing data to intranet devices and devices in other systems.

Under-mine part: consisting of input/output devices which are fixed and portable (wireless) detectors.

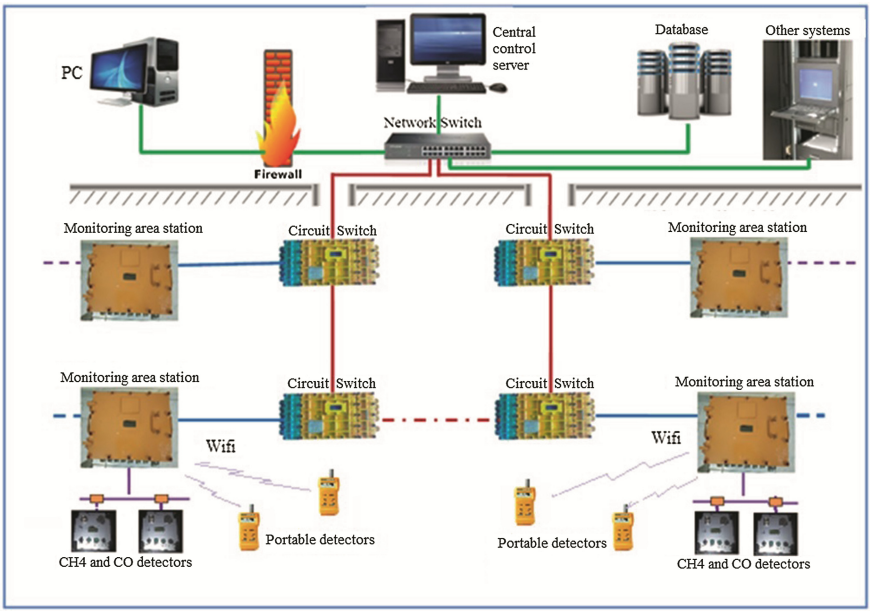


Fig. 1. The system diagram

+CH4 and CO detectors are fixed and have the functions of measuring the concentration of gases, displaying and send alarming signals when the concentration of gas goes beyond the set levels, automatically send power supply cut command to prevent risk of fire and explosion caused by high gas concentration under mines. Gas detectors transmit information/data to the centre through area control stations via RS485 line [5].

2 The Building of Database Structure

By practical investigation on the operating procedure, especially the storing and managing the information of gas detectors used in some underground gas tracking and monitoring systems, the researchers suggest model of storing information that needs monitoring in a system, including:

– Monitoring data information from gas detectors:

Data from gas detectors need to be stored in the database of the system including: The ordinal number of the name of detectors set by a person in charge, the ordinal number of data record from the main detector, the status of detector, alarming level 1, alarming level 2, the time of report, date of report submission, detailed description of the number of times of report and the standby mode.

This part provides detailed information about undermine gas detections at different times and continuously transmit to the system's central computerized controller.

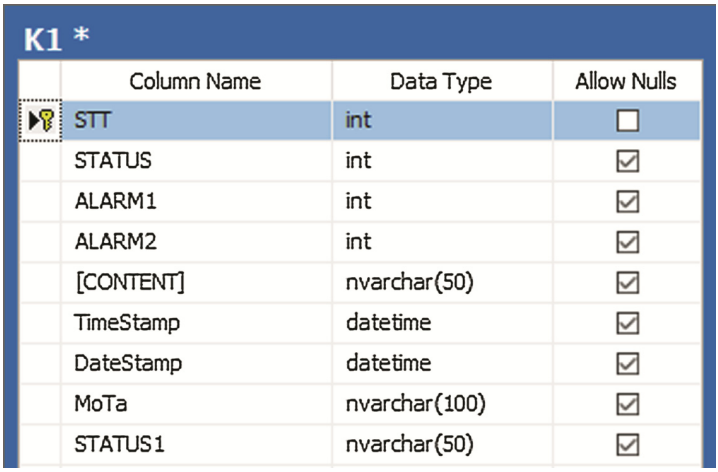
Function-related assumptions: each detector has it own storing table, with each detecting time is attached a unique code and automatically ascending in the database system.

– Managing the information of alarming data:

The alarming data lets us know for details about the times of alarming when there are systematic incidents like errors in starting detectors, power supply cut, lost of connection, detectors switching on and off. This part consists of the following information: alarming code which is the main key, status of switching on and off, timestamp of alarm, date of alarm, type of alarm indicating what happens to the system.

– Results and discussion:

From investigating and building the model of database structure to manage information, the researchers used MS SQL Server 2008 R2 [6, 7] to design the system. This is the most commonly used database managing software at present, with full functions that meet all technical requirements in managing information of the system. The researchers have set up all information table for every main key. Following are the results received on the MS SQL Server R2 (Fig. 2):


K1 *

	Column Name	Data Type	Allow Nulls
🔑	STT	int	☐
	STATUS	int	☑
	ALARM1	int	☑
	ALARM2	int	☑
	[CONTENT]	nvarchar(50)	☑
	TimeStamp	datetime	☑
	DateStamp	datetime	☑
	MoTa	nvarchar(100)	☑
	STATUS1	nvarchar(50)	☑

Fig. 2. Table of data about gas detector channels.

Building storing information table according to function-related assumptions above include: Alarm, Error, separate detectors K1, K2, K3, K4, K5, K6, K7, K8 and other information fields on each table are run on MS SQL Server 2008 R2. The product pilot run gave correct results, and meet the requirements of receiving and managing data information received from under mine gas detectors to the system database. The function-related assumptions are justified.

The researchers have built a general and unified structure model for managing information transmitted from fixed and portable detectors, contributing to the monitoring and managing the operation under mine and underground works at present (Fig. 3).

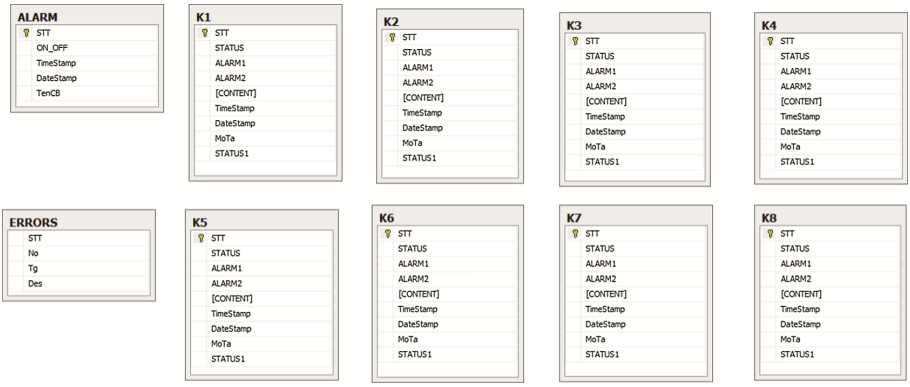


Fig. 3. Tables of data in the system

2.1 System Software

The environment tracking system used in mine exploitation is designed to measure and set alarm for gases such as CH4 and CO in the mines. The system will send an alarm signal when any one of the parameter exceeds the preset level of danger (sound and light), at the same time cut the power supply at the dangerous place. When the danger is resolved, the power supply can be turned on again from the centre or locally. This system can store, search and print out reports on environment parameters according to fixed templates.

Technical features:

- The system consists of: server computer, central controller, methane gas detector, CO detector, power supply and switch; repeater (used for large distance).
- The system can manage hundreds (or more) detectors.
- Time to update data ranges from 8–30 s (depending on the number of places that need monitoring).
- Can transmit data to the centre via typical cable, optical cable, wifi lines for portable hand devices.
- The system can send alarming signals by sound and light and automatically send command to cut power supply when there are gas related risks.
- Server displays the mine map, store and search for data in forms of tables or charts.
- The uses of this system:
  This software directly receives data of gases like CH4 and CO present at places under mines and transmit to centrally controlled computers at realtime, whose interface is simple and friendly so that person in charge can easily set up and use. This program can track and measure at any position of any detector under mine, helping the remote monitoring and collecting of data on the status of observation system, tracking multi-meter of gases, heat, and wind inside the mine climate according to real time. This program is written in Visual Basic language.

- Features of the program:
  - Setup details for sensor detectors like locations, types of detectors, detailed descriptions of detector, high and low level of gas warning.
  - Able to send measuring command to CO and CH4 detectors separately or all together in order to collect data and display on the screen.
  - Able to re-assign the detectors under mines using maps of mine routes.
  - Display concurrent parameters of CO and CH4 from all gas detectors in the system.
  - Able to track the value of gases received in graphs daily or a certain duration and print out for reports.
  - Display and print out the list of system warnings.
- Setting up sensor detectors:
  At this interface, the person in charge can use the following features:
  - Setup name for each channel.
  - Set up warning value 1 (alarm 1), warning value 2 (Alarm 2). Warning 1 is to remind and have lower value of warning compared to warning value 2. Warning value 2 is the official warning.
  - Set up measuring range for different channels.
  - Set up CH4 and CO gas detectors.

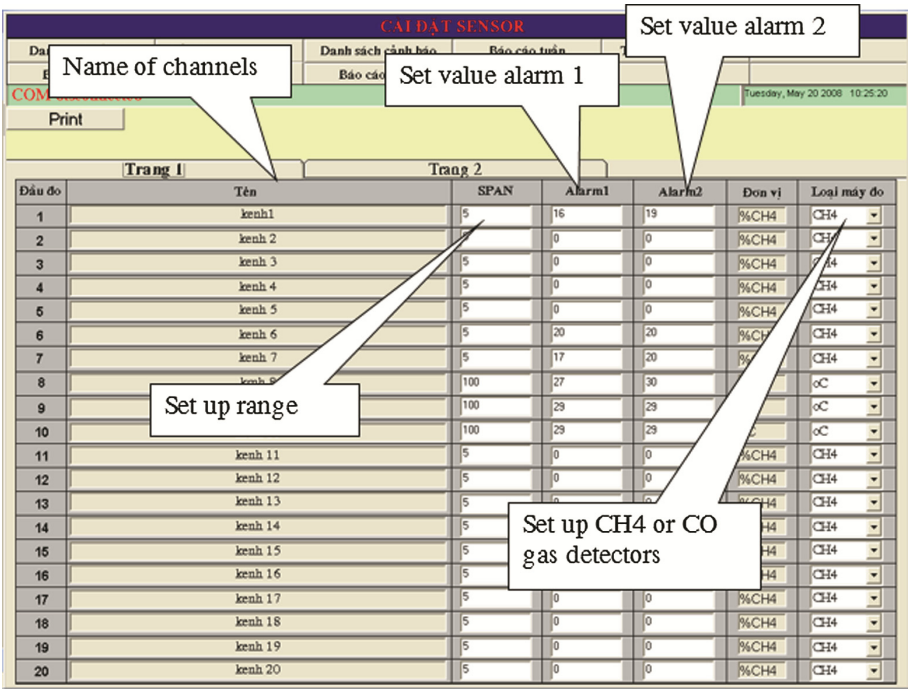


Fig. 4. Functions setup for sensor detectors

As shown in Fig. 4, we can set up names, range, alarming level 1, alarming level 2 and the types of detector. After that, during operation, if the value of received data exceeds the set up levels, the detectors will send alarming sound, on the server will send the same alarming sound.

- Tracking sensor detectors:
  Display the current status of measuring devices, set up for each channel. This interface is displayed on 2 pages. The first page displays channel 1 to channel 20, the second page displays channel 21 to channel 40. Features include:

– Display the current status of CO and CH4 measuring devices: whether it is on or off, the current measured value, alarming status, open-circuit, alarming level, status of sending measurement command of each devices.
– Display alarming levels for detectors: there are 2 alarming levels: alarming level 1 (alarm 1) and alarming level 2 (alarm 2).
– Display types of detector and measurement units.

The following Figure explains the current status of a channel in detail (Fig. 5).

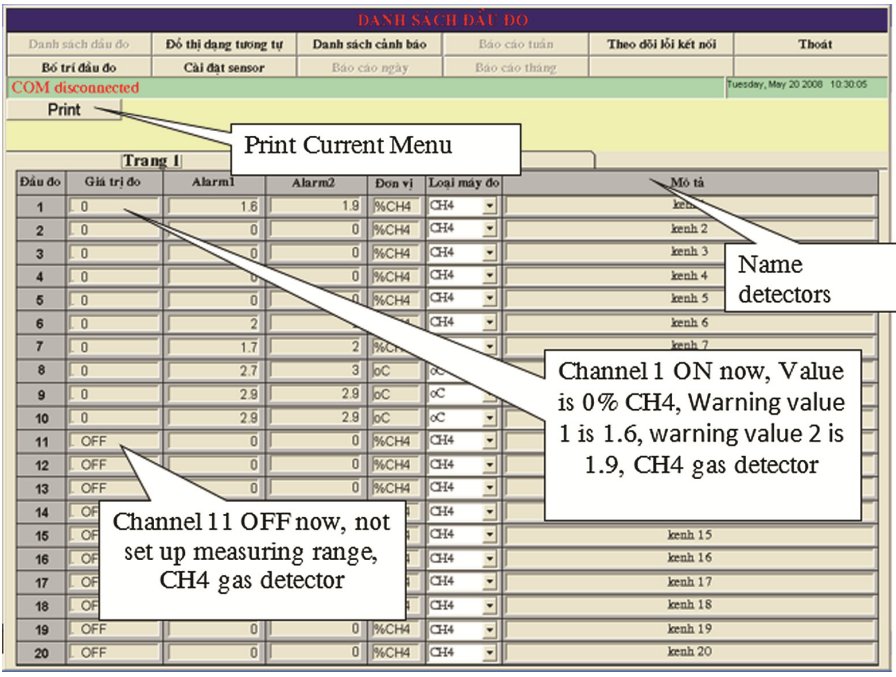


Fig. 5. Tracking all sensor detectors in the system

- Assigning gas detectors using map of mine routes:
  This part of the system helps the person in charge track measuring devices put in the map of mine route visually. The interface includes 10 pages, which users can add or remove channels of measuring displayed on each page. (there are 40 channels

maximum). Moreover, map of mine routes and the positions of CO and CH4 detectors can also be changed.

– Select the interface page: directly click the mouse on the page menu or click on previous/next page.
– Add or remove channel: Use mouse to tick the equivalent channel in the box "Add/remove channel" to add or remove the corresponding channels. The selected channel will hide from the map or reappear on the map accordingly. After that, symbols of channel of measuring can be altered on the maps (Fig. 6).

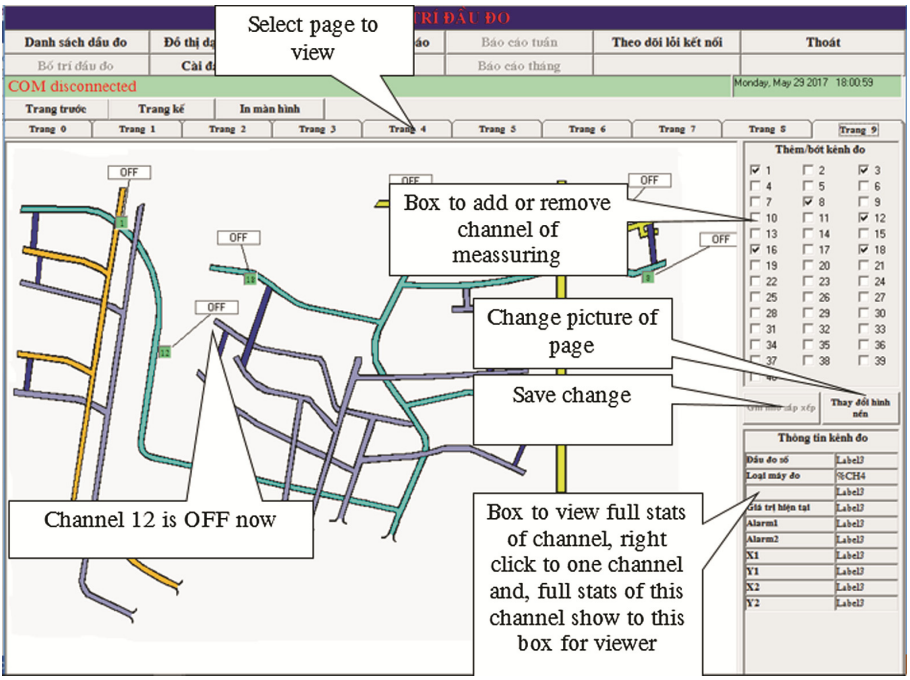


Fig. 6. Assigning multi-sensor gas detectors in the mine map

- Tracking graphic data (Fig. 7):
  – The concentration of CO and CH4 present in the air is shown in the form of continuous graph: digital value of measurements of each channel is connecting in the system help the person in charge easily monitor the parameters visually.
  – With the vertical axis showing the concentration of gases that need tracking and the horizontal line show values after every 3 h, 6 h, 12 h and 24 h.
  – Let users to track 6 different sensor detectors at the same time, with each detector assigned one color to distinguish between them.
  – Let users select dates to track data.
  – Send commands to selected detectors to continuously collect and display data of gases according to real time (Figs. 8 and 9).

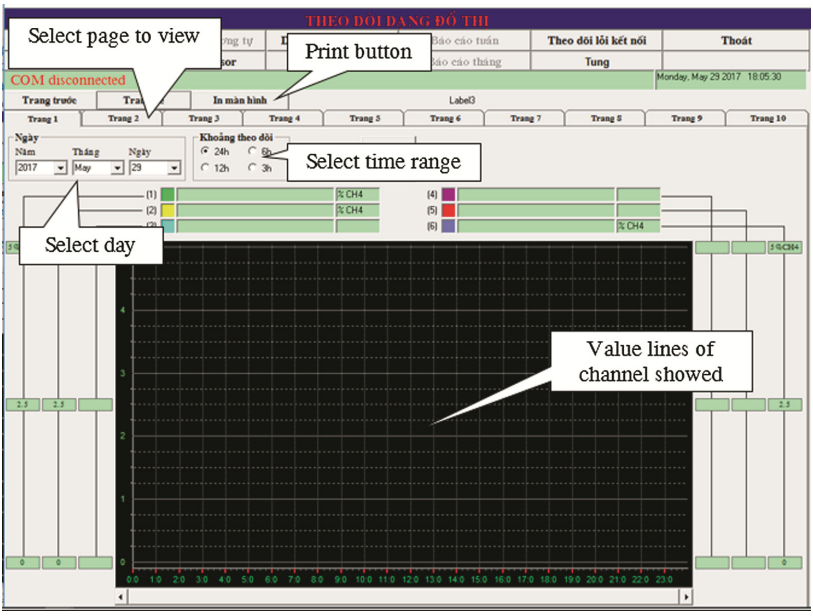


Fig. 7. Tracking data from sensor detectors in graphs. (Color figure online)

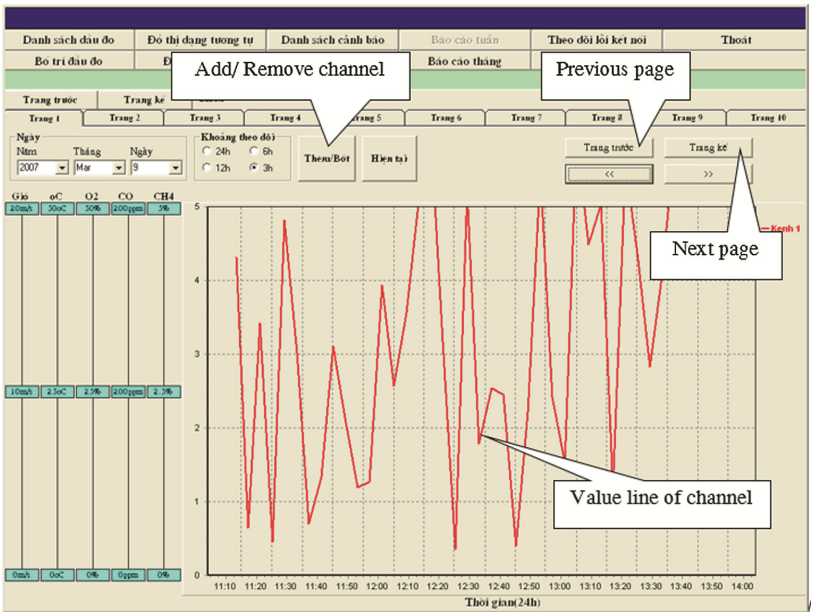


Fig. 8. Data of channel 1 in graph displayed in the program

STT	ON_OFF	TimeStamp	DateStamp	TenCB	Address
1	☑	10:49:41 AM	9/11/2008	HMI started	192.168.6.14
2	☑	10:50:07 AM	9/11/2008	COM disconne	192.168.6.3
3	☑	10:54:32 AM	9/11/2008	COM serial err	192.168.6.7
4	☐	10:54:32 AM	9/11/2008	COM disconne	192.168.6.1
5	☑	10:54:52 AM	9/11/2008	COM disconne	192.168.1.3
6	☐	10:55:48 AM	9/11/2008	COM disconne	192.168.1.16
7	☑	10:56:08 AM	9/11/2008	COM disconne	192.168.1.27
8	☐	10:56:12 AM	9/11/2008	COM disconne	192.168.6.2
9	☑	10:56:41 AM	9/11/2008	COM disconne	192.168.6.5
10	☐	10:58:06 AM	9/11/2008	COM disconne	192.168.1.1
11	☑	10:58:30 AM	9/11/2008	COM disconne	192.168.1.1
12	☑	1:16:46 PM	9/11/2008	HMI started	192.168.1.1
13	☑	5:09:53 PM	9/24/2008	HMI started	192.168.1.2
14	☑	5:10:14 PM	9/24/2008	COM disconne	192.168.1.2
15	☑	9:51:29 PM	9/24/2008	HMI started	192.168.3.8

Fig. 9. The data of ALARM table in the central database

3 Conclusion

The above research results suggest a model of database structure used in receiving and managing information and operation under mines and underground works automatically by applying information technology and data transmission. With this model of database structure, the received data from gas detectors will be sent to the central server to be stored and processed easily and conveniently, especially when the number of times receiving results are numerous and increasing in time.

References

1. Methane. https://en.wikipedia.org/wiki/Methane
2. Carbon monoxide. https://en.wikipedia.org/wiki/Carbon_monoxide, http://www.nhs.uk/Conditions/Carbon-monoxide-poisoning/Pages/Introduction.aspx
3. Under-mine gas warning system. http://journals.sagepub.com/doi/full/10.1155/2013/159273, https://www.ampcontrolgroup.com/gas-ventilation-monitoring-and-control/
4. The mine exploitation. https://en.wikipedia.org/wiki/Mining
5. RS 485. https://en.wikipedia.org/wiki/RS-485
6. MS SQL Server 2008 R2. https://en.wikipedia.org/wiki/Microsoft_SQL_Server.3
7. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)

The Method of Maintaining Data Consistency in Allocating Resources for the P2P Network Model

Ha Huy Cuong Nguyen[1(✉)], Hong Minh Nguyen[2], and Trung Son Doan[3]

[1] Department of Information Technology, Quangnam University, Tam Ky, Vietnam
nguyenhahuycuong@gmail.com
[2] People's Security University, Ho Chi Minh City, Vietnam
hongminhnguyen1982@gmail.com
[3] Faculty of Information Technology and Security, People' Security Academy, Hanoi, Vietnam

Abstract. Maintaining consistency of mutable data replication in a Peer-to-Peer (P2P) system is a fundamental challenge. That is the frequency problem of structural change in P2P system i.e. how often does the actual structure of the tree change. There are several models which exhibit better characteristics for updating structure. However, it is worth separating the changes of structure from effect of data. An novel approach for balanced consistency maintenance in structured tree in P2P systems supporting heterogeneous distributed platforms is presented in this article. Nodes of each object are organized into a tree structure for disseminating update processes. An analytical model to optimize is showed theoretically. Several simulations for experiment results will be carried out in the dynamic network conditions.

Keywords: P2P · Distributed system · Data consistency · Multi-copy

1 Introduction

P2P networking systems use an application level organization of the overlay network for flexibly sharing resources (e.g., files and multimedia documents) which are stored across a wide range of computer networks. Structured P2P systems have been effectively designed for many data applications [1]. Each computer of P2P system has both roles client and server at the same time. The advantages of P2P approach enable to improve data availability, fault tolerance, and scalability for dynamic content sharing. Because dynamic characteristics of network combined with diverse application require heterogeneously distributed resource hardware, it emanates challenges for P2P consistency management [2,3]. The communication and exchange of information is performed directly between the participating peers and the relationships between the nodes which are heterogeneous and distributed, in the network are not equal. Some messages will be lost

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 155–165, 2018.
https://doi.org/10.1007/978-3-319-77818-1_15

or attenuated when transmitted over the network. P2P networking systems differ from other Internet applications, hereby some user tends to share data from a large number of the more central virtual machines and Web Servers. Several well-known P2P networking systems P2P that are able to share files such as CAN, Chord, Gnutella, Free-net, Napster, Pastry, and Tapestry [3,13,14]. P2P networking systems, more importantly, allow the location of arbitrary data objects. Applying sequential consistency maintenance leads to prohibitively prolonging synchronization delays due to a large number –even unreliable– of peer. The "deadlock" may occur when a crashed heterogeneous nodes causes other nodes to wait forever. At the other extreme, eventual consistency are allowed to concurrently update their local copy. And it only requires that all heterogeneous nodes become identical after a long enough failure-free and update-free interval [2,4,5].

An important characteristic of P2P networking systems is their ability to provide a large storage, CPU power coupled with other resources while imposing a low cost for scalability, and for entry into and exit from the heterogeneously distributed platform [6]. The participation and exit of various nodes, as well as insertion and deletion of objects, are dynamic. The P2P networking systems exhibit a high level of self-organization and are able to operate efficiently despite the lack of any heterogeneous infrastructure [6]. This model requires that if a heterogeneous node wants to enjoy the services which is provided by other nodes, that node should provide services to other heterogeneous nodes.

In the past, both distributed computing and parallel computing have been commonly used for large-scale computation. Internet applications for global sharing like Napster, Gnutella and FreeNet have recently gained popularity based on P2P architecture. Several research projects aim at constructing other types of P2P applications. P2P systems can be characterized as distributed systems in which all nodes have identical capabilities and responsibilities and all communication is symmetric [2–6].

Several key challenges in P2P networking systems include: object storage mechanisms, object look-up in an efficient manner, retrieval in a scalable manner, dynamic reconfiguration as well as objects joining and leaving the network randomly, replication strategies to expedite object search, tradeoffs between object size latency and table sizes, anonymity, privacy, and security.

This article presents model structure based on tree broadcast, a generic decentralized P2P content location, and routing system for very large, self-configuring overlay networks of nodes which are connected via the Internet. The problems of routing and locating content in large scale P2P networking systems are considered in looking for efficient algorithms. Particularly, an algorithm Ary-Construction, constructing tree update to maintain consistency in structured tree of P2P networking systems, is proposed in this work.

The work is organized as follows: Sect. 2 summarizes briefly related works, Sect. 3 describes existing models, Sect. 4 proposes analytical model for resource allocation in heterogeneous distributed platforms, Sect. 5 describes algorithm in detail, Sect. 6 shows experiments and results and finally Sect. 7 presents conclusions and suggestions for future work.

2 Related Work

In [7], the problem of partitioning web-link graph for web ranking in P2P is formulated as a minimal cut-set with density balanced partitioning. The problem is proved to be an NP-Hard, by reducing to the minimum bi section problem [4,7]. In P2P based PageRank [5,7], each computational peer contains a local web-link graph and its PageRank in computed locally. P2P is a viable choice to address such limitation. To be able to compute the global ranking, a special node, so called word-node, is constructed to store the linkpage information of the other peers. In [8], toward authentication between familiar peer in P2P networking systems. A secure environment can only be achieved, when peers are sure, that they are communicating with partner desired.

There are two classes of P2P overlay networks: Structured and Unstructured. The technical meaning of structured P2P, strong consistency is provided by organizing heterogeneous nodes to an auxiliary structure on top of the overlay for update propagation [6]. Examples include the tree structure in SCOPE [9], the two-tiered structure in OceanStore, and a hybrid of tree and two-tiered structure in build a tree by recursively partitioning the identifier space and selecting a representative node as a tree node for each partition [4]. Only leaf nodes store object copies, all the intermediate nodes only store information of the tree structure in their sub-space. Nodes who may not be interested in the object are in the object's update dissemination tree, which adds unnecessary overhead of maintaining the tree from node failures. To the contrary, SCOPE, OceanStore constructs the tree structure dissemination by only involving heterogeneous nodes who are interested in the object which greatly reduce the overhead of maintenance and update propagation. The problem also efficiently builds the tree structure dissemination and binary tree decomposition to make it balanced and robust under the node churn.

In unstructured P2P systems, mainly two types of bounded consistency are provided rumor spreading and replica chain are used to ensure a certain probability of an update being received. The probability is tuned by adjusting the redundancy degree in propagating an update to balance the communication overhead with the consistency strictness. The message broadcast is an unstructured byte array that is delivered to all members of the group through the send method probabilistic bounded consistency. In previous work [10], propose method to be the best algorithm for constructing unstructured P2P graphs suitable for heterogeneous random selection. Here is a brief overview of other unstructured approaches. In extends Gnutella [14] by making both graph construction and query-resolution sensitive to node capacities. High-capacity nodes here have higher degrees, and are more likely to be traversed by random walks. While Swaplinks shares these two features with GIA, Swaplinks exhibits more accurate control over degree and probability of selection. Other examples of unstructured graph construction schemes include Araneola, an approach by Law and Siu, and Jianming Fu et al. None of these take node heterogeneity into account. In [9] mechanism can be used as a random node selection primitive, but as was the case with the previously mentioned schemes, does not take into account node

heterogeneity. In systems P2P a node sequentially contacts the cached ancestors to avoid conflict relocation decisions while in a node uses multiple side links in parallel to retrieve the lost packets. Previous works have proposed continuous models for consistency maintenance, which have been extended by a maintain consistency model in [1] for P2P applications. In that reference, the usage of core technique for maintaining consistency is a hybrid of push and pull methods, which are also used to provide application tailored cache consistency in [9]. Although each node can specify its consistency requirement, the model in [14] makes each node perform the strongest consistency maintenance from all its descendant nodes in the overlay replica hierarchy. Thus, the overhead of maintaining consistency at a node is not reduced even it only requires a weak consistency as long as one of its descendant nodes requires a strong consistency.

3 The Model Structure Tree Broadcasts Resource Allocation in Heterogeneous Distributed Platforms

The model structure tree broadcasts operation stores objects at a user defined number of diverse nodes. An operation reliably retrieves a copy of the requested object if one exists. The objects are retrieved from the live node, there is an issue how does look-up among the nodes storing the object. The mode P2P is measured here in terms of a scalar, application defined metric, such as the numbers of network hops or network delay. The model is completely decentralized, is self-configuring, it automatically adapts to the arrival, departure, and failure of nodes and it is scalable. The number of nodes traversed, as well as the number of messages exchanged while routing a client request is at most algorithms in the total number of nodes in the system.

P2P computing is carried out over application layer network where in all interactions among the processors are at a "peer" level, without any hierarchy among the processors. Thus, all processors are equal and play a symmetric role in the computation. P2P computing arose as a paradigm shift from client-server computing where the roles among the processors are essentially asymmetrical. P2P networks are typically self-organizing and may or may not have a regular structure to the network. No central directories (such as those used in domain name servers) for name resolution and object look-up are allowed.

An analytical model for adaptive update window protocol is presented in [12], where the window specifies the number of uncommitted updates in each replica node's buffer. The information of each node's update rate and propagation latency are required to optimize the window size in [12]. Such optimization is unrealistic for P2P systems due to their required global information. To the contrary, every BCoM node has a fair amount of consistency maintenance overhead because of the uniform buffer size and node degree in dDT. Moreover, BCoM provides incentives for nodes to contribute more bandwidth to update dissemination, as in proposed approach, it allows that faster nodes are closer to the root and they will receive updates sooner. The window size optimization

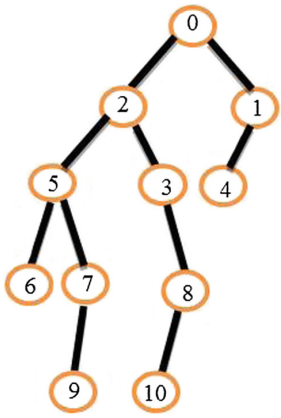


Fig. 1. A simple structure tree decomposition

model in BCoM only requires limited information that can be obtained in a fully distributed way (Fig. 1).

$$\sum_{i}^{N} E_{ij} = \sum_{i}^{N}(A_{ij} + \sum_{i=1}^{V} C_{\mathrm{ij}}) \tag{1}$$

E_{ijt} obeys the rules as follows:

$$\begin{aligned} &E \geq \sum_{i=1}^{N}\sum_{j=1}^{V_N} E_{\mathrm{ijt}} \\ &E_{\mathrm{ijt}} \geq C_{\mathrm{ij}} \geq 0\,(i = 1, \ldots, N; j = 1, \ldots, V_N) \end{aligned} \tag{2}$$

The resource allocation problem is how to control the resource allocation to VMs with the goal of minimizing the function F_t, giving the limited resources. The following formulation is obtained:

$$\begin{aligned} &F_t = min \sum_{i=1}^{N}\sum_{i=1}^{V_i} \frac{f_{\mathrm{ij}}(EN_{\mathrm{ijt}}, \sum_{x=1}^{V} EO_{\mathrm{ijt}}^{x}, D_{ijt})}{\Phi_{ij}} \times SP_{ij} \\ &\begin{cases} \sum_{i=1}^{N}\sum_{i=1}^{V_i} E_{ijt} \leq E \\ E_{\mathrm{ijt}} \geq C_{\mathrm{ij}} \;\; (i = 1, 2, \ldots, V; j = 1, 2, \ldots, N) \\ \sum_{i=1}^{V_i} EN_{\mathrm{ijt}} + \sum_{j=1}^{V_i} EO_{\mathrm{ijt}}^{i} \leq E_i \\ E_{it} \geq C_{ij} \;\; (i = 1, 2, \ldots, V; j = 1, 2, \ldots, N). \end{cases} \end{aligned} \tag{3}$$

This method can be used to avoid deadlock to solve optimal resource model provides $\boldsymbol{V}$ VM-out-of-$\boldsymbol{N}$ PM. The proposed algorithm is based on wait-for graphs (WFG) algorithm and presented in Sect. 5. The probability of obtaining a particular random graph $G = (V, L)$, where $|V| = n$ and $|L| = l$. With Eq. (4) as below:

$$P(G) = \rho^{l}(1-\rho)^{n(n-1)/2-l} \tag{4}$$

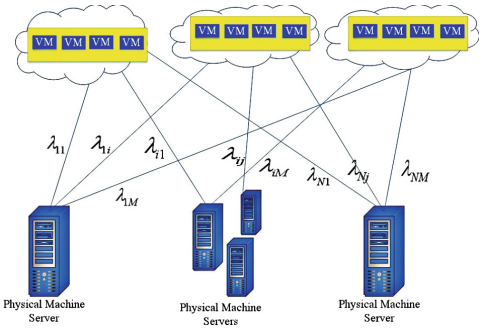


Fig. 2. A simple model $\boldsymbol{V}$ VM-out-of-$\boldsymbol{N}$ PM

4 Analytical Model for Resource Allocation in Heterogeneous Distributed Platforms

The frequent node churn in P2P systems forbids us to use any complicated optimization techniques that require several hours of computation (Fig. 2).

4.1 Queueing Model

Assuming the total number of heterogeneous nodes is N, the node degree is d, and there are $L(L = O(\log_d N))$ layers of internal nodes with an update buffer of size k_i. The leaf nodes are in layer L and do not have any buffer. The update arrivals are modeled by a Poisson process with an average arrival rate λ_i, since each update is issued by a heterogeneous node independently and identically at random. The latency of receiving an update from the parent and an acknowledgment from the child is denoted as the service time for an update propagation. The service time for one layer to its adjacent layer below is the longest parent-child service tim these two layers. μ_l denotes the service time for update propagation from layer l to layer $(l+1)$.

This queuing model explains that given a k_i, the effective buffer size $l * k_i$ is determined by l, which is the layer of the bottleneck node. The larger the effective buffer size, the lower the discard probability. When the bottleneck node is a leaf node $(l = L)$, buffer resources of dDT_i are fully used with an effective buffer size $L * k_i$. The discard probability of an update is computed based on the queuing model of dDT_i after being optimized by tree node migrations as shown. The queue becomes an $M/M/1$ queue with a buffer size $L * k_i$, an arrival rate λ and a service time μ_{L-1}.

4.2 Availability and Latency Computation

Define the update request intensity as ρ.

$$\rho = \frac{\lambda}{\mu_{L-1}} \tag{5}$$

Define the cost spread intensity as $cost_i$

$$\cos t_i = \log \frac{N}{d^0} + \log \frac{N}{d^1} + \ldots + \log \frac{N}{d^{i-1}} \tag{6}$$

$$\cos t_i = \log N + (\log N - b) + \ldots + (\log N - b * (i-1)) \tag{7}$$

$$\cos t_i = i * \log N - \frac{b * \rho * (i-1)}{2} \tag{8}$$

$$\cos t_i = \rho_1 * \cos t_1 + \rho_2 * \cos t + \ldots \rho_i * \cos t_L = \sum\nolimits_{i=1}^{i=L} \rho_i * \cos t_i \tag{9}$$

Define the probability of n updates in the queue as Π_n. Based on the queueing theory for $M/M/1$ finite queue, Π_n is represented as Eq. 9.

$$\Pi_n = \rho^n * \pi_0 \tag{10}$$

The discard probability is π_{L*k_i}, which indicates the buffer overflow. From $\sum_{n=0}^{L*k_i} \pi_n = 1$, we get $\pi_0 = \frac{1-\rho}{1-\rho^{L*k_i}}$. And the discard probability is computed in Eq. 10.

$$\pi_{L*k_i} = \frac{1-\rho}{1-\rho^{L*k_i}} \rho^{L*k_i} \tag{11}$$

The expected number of packets in the queue $E[N_{L*k_i}]$ is calculated in Eq. 11.

$$E[N_{L*k_i}] = \sum_{0 \le n \le L*k_i} n * \pi_n \tag{12}$$

Plug in the Eq. 12 for π_n, the final form of $E[N_{L*k_i}]$ is given in Eq. 13.

$$E[N_{L*k_i}] = \frac{(L*k_i+1) * \rho^{L*k_i+1}}{(\rho^{L*k_i+1}-1)} + \frac{\rho}{1-p} \tag{13}$$

5 Algorithm

In this article, an algorithm dDT construction is proposed for deadlock avoidance maintains property of n-vertex directed graph when the new is added in the graph using two-way search. The dDT construction algorithm uses the number of sub-tree nodes as the metric for insertions, instead of the tree depth used in traditional balanced tree algorithms. This is because a rejoining node with a sub-tree may increase the tree depth by more than one, which is beyond the one by one tree height increase handle by traditional balanced tree algorithms.

The method update dissemination in P2P systems has four requirements: (1) awareness of heterogeneous peer capacities, (2) scalability with a large number of peers, (3) a bounded delay for update delivery, and (4) robustness to the frequent node churn and different workload petterns. The constructs an update dissemination tree by considering each user's capacity and latency requirements to address (1) and (3), both of which are also handled by the tree node migration [1]. The major difference is that Pastry improves the performance to meet the

individual replica node's requirement, while the tree node migration in Pastry improves the overall system performance. Moreover, Pastry requires information on each user's latency requirement and capacity, which are infeasible to be implemented in P2P systems. To the contrary, Pastry node migration only involves local information, which is also performed on demand to support (2) without asking a replica node to specify requirements in advance.

The method ancestor cache has extra benefits by avoiding deadlock communication overhead to maintain end nodes on other sub trees in this article.

Algorithm 1. Algorithm dDT construction (p, q)

Input: node p receives node q's join request;
Output: parent of node q in dDT;

BEGIN
$i.locate.successor(key)$, where key $\neq$ i:
if key $\in$ (i, $successor$) then
return($successor$)
else if p does not have d children **then**;
$Sub_{no.}(p) += Sub_{no.}(q)$ **return** p **else**
find a child f of p s.t. f has the smallest $Sub_{no.}$
$Sub_{no.}(f) += Sub_{no.}(q)$
return dDT Construction(f,q)

END.

In many systems, a heterogeneous environment is preferable to one that is homogeneous. However, it provides better performance particular systems and workload. Even if the workload itself is more suitable to a heterogeneous distributed platforms, the systems rescheduling algorithm should exploit heterogeneity well to benefit from it. Time-bound consistency: TTL guided push and/or pull methods are used to indicate a valid period for a replica copy. When the period expires, the replica node checks the validity of the replica copy with the source to serve the following read requests.

6 Experiments and Results

Nakashima [4] propose a construction of an updated tree statically only from replica nodes. It connects to the tree by computation of the balance of the number of subtree nodes. When new updates are received, the root propagates to all nodes below, then other new updates arrive in the root that will be discarded (Fig. 3).

Yi [6] propose BCoM for construction of an updated tree statically from replica nodes which are ordered by their arrival times and balance of the number

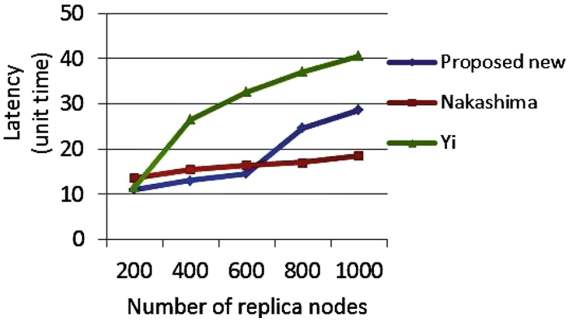


Fig. 3. Comparison the optimal time of proposed algorithms with Nakashima and Yi algorithms

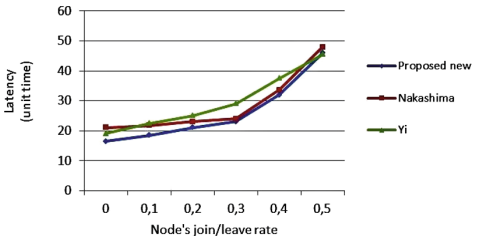


Fig. 4. Comparison the optimal time of proposed algorithms with Nakashima and Yi algorithms

of subtree nodes. Each replica node (except leaf nodes) is given a buffer of size k for buffering updates.

The root receives an update, then sequentially sends it to all its children. Such operation is repeated until the leaf nodes received.

The sliding window size k is critical for balancing the consistency strictness, object availability, and latency. So BCoM proposed a class that each node requires a maximum and sequences of generated updates, balanced among the parameter above.

The comparative analysis of experimental result can be seen in many times, apter task execution, although there were individual time improved proposed algorithm response time was not significantly less than an optimal time algorithm. In most cases, improved algorithm is better than the optimal time algorithm, thus validated the correctness and effectiveness. The process of rescheduling parallel tasks determines the order of task execution and the processor to which each task is assigned (Fig. 4).

7 Conclusion and Future Works

The article provides a solution based on algorithm for allocating effectively resources. The method of optimizing the resource usage is exhibited in the for-

mulae 2. The algorithm proposes a novel way to build and maintain the update tree and then perform the optimal replication based on the update rates. Moreover, it is attributed a the effective method to update propagation. The resultant experiments show that node's join/leave rate, the update rate, and the number of replica nodes increase. Therefore, the proposal is more stable and effective than the work of Nakashima and Yi about the latency and the ratio of the successful update.

Additionally, maintaining the total number of nodes in each sub-tree is simpler and more efficient in term of time than maintaining the depth of each sub-tree. Internal nodes need to wait until an insertion completion. In that case, the updated tree depth can be collected layer by layer from leaf nodes back to the root. This makes the real-time maintenance of the tree depth difficult and unnecessary when tree nodes are frequently joining and leaving. However, internal nodes can immediately upadate the total number of sub-tree nodes after forwarding a new node to a child.

References

1. Rowstron, A., Druschel, P.: Pastry: scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In: Guerraoui, R. (ed.) Middleware 2001. LNCS, vol. 2218, pp. 329–350. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45518-3_18
2. Bermbach, D., Kuhlenkamp, J.: Consistency in distributed storage systems. In: Gramoli, V., Guerraoui, R. (eds.) NETYS 2013. LNCS, vol. 7853, pp. 175–189. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40148-0_13
3. Kshemkalyani, A.D., Singhal, M.: Distributed Computing Principles, Algorithms, and Systems. Cambridge University Press, Cambridge (2008)
4. Nakashima, T., Fujita, S.: Tree-based consistency maintenance scheme for peer-to-peer file sharing systems. In: 2013 First International Symposium on IEEE Computing and Networking (CANDAR), pp. 187–193 (2013)
5. Shen, H., Liu, G., Chandler, H.: Swarm intelligence based file replication and consistency maintenance in structured P2P file sharing systems. IEEE Trans. Comput. **64**, 2953–2967 (2015)
6. Li, Z., Xie, G., Li, Z.: Efficient and scalable consistency maintenance for heterogeneous peer-to-peer systems. IEEE Trans. Parallel Distrib. Syst. **19**, 1695–1708 (2008)
7. Nassermostofi, F.: Toward authentication between familiar Peers in P2P networking systems. In: Proceeding of the 9th GI Conference Autonomous Systems, pp. 88–103 (2016)
8. Pang, X., Wang, C., Zhang, Y.: A new P2P identity authentication method based on zero-knowledge under hybrid P2P network. TELKOMNIKA Indones. J. Electr. Eng. **11**(10), 6187–6192 (2013)
9. Chen, X., Ren, S., Wang, H.: SCOPE: scalable consistency maintenance in structured P2P systems. In: 24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, pp. 1502–1513 (2005)
10. Vishnumurthy, V., Francis, P.: On heterogeneous overlay construction and random node selection in unstructured P2P networks. In: Proceedings of the IEEE INFOCOM (2006)

11. Adami, D., Gabbrielli, A., Giordano, S., Pagano, M., Portaluri, G.: A fuzzy logic approach for resources allocation in cloud data center. In: Proceedings 2015 IEEE Globecom Workshops (GC Wkshps), pp. 1–6 (2015)
12. Zhao, B.Y., et al.: Tapestry: a resilient global-scale overlay for service deployment. IEEE J. Sel. Areas Commun. **22**, 41–53 (2004)
13. www.napster.com/
14. www.gnutella.com/

Fragmentation in Distributed Database Design Based on KR Rough Clustering Technique

Van Nghia Luong[1(✉)], Van Son Le[2], and Van Ban Doan[3]

[1] Faculty of Information Technology,
Pham Van Dong University, Quang Ngai, Vietnam
nghia.itq@gmail.com, lvnghia@pdu.edu.vn
[2] Da Nang University of Education, Da Nang, Vietnam
levansupham2004@yahoo.com
[3] Institute of Information Technology,
Vietnamese Academy of Science and Technology, Hanoi, Vietnam
dvban@ioit.ac.vn

Abstract. Knowledge mining according to rough set approach is an effective method for large datasets containing many different types of data. Rough clustering, as in rough set theory, using lower approximation and upper approximation, allows objects to belong to multiple clusters in a dataset. KR Rough Clustering Technique (K-Means Rough) we propose in this paper follows k-Means primitive clustering algorithm improvement approach by combining distance, similarity with upper approximation and lower approximation. In particular, appropriate focuses will be calculated to determine whether an object will be assigned to lower approximation or upper approximation of each cluster.

Keywords: Rough set theory · Vertical fragmentation · Rough cluster
Cluster focus

1 Rationale

Rough clustering algorithms use distance measure to construct a similar matrix and each pair of objects in this matrix is assigned to the current cluster or new cluster depending on one or both objects in the pair currently being distributed [3]. With this approach, a large number of clusters will be created. It may be uncertain to ensure whether lower approximations of the clusters have the most effective overlay area of the dataset [4].

Clustering technique according to rough set theory supports clustering in two directions:

- Improve such classic clustering algorithms as k-Means, k-Medoids into rough_ k-Means (k-Means Rough), rough_k-Medoids (k-Medoids Rough), by combining distance, similarity with upper approximation and lower approximation [10].
- Support to identify the minimum number of clusters, based on the number of initial suggestion clusters provided by the user. Clusters will be clustered if approximations on the intersection clusters are non-empty [11].

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 166–172, 2018.
https://doi.org/10.1007/978-3-319-77818-1_16

This article is organized as follows: Sect. 2 presents some related concepts of rough clustering technique. Proposed KR algorithm for vertical fragmentation in distributed data based on rough clustering technique is presented in Sect. 3. Section 4 in turn presents the experimental setup on KR and compares the experimental results with primitive k-Means. Section 5 is the conclusion.

2 Some Related Concepts

2.1 Data Discretization and Attribute Selection, Attribute Extraction According to Rough Set Approach

In the field of knowledge mining, the problem is how to process mixed data with continuous values. Many algorithms are used to discrete data such as logical reasoning methods, NAIVE algorithm, etc. However, there is no optimal algorithm. An algorithm is selected depending on the type of data to be processed. Authors in [2] outline some data discretization methods based on rough set and logical reasoning.

Attribute Selection, Attribute Extraction based on rough set [4]: Databases in practice often have many attributes. Attributes required for KPDL problem being processed are not all. Selecting the appropriate attributes for KPDL method is necessary.

2.2 Information System, Indistinguishable Relation

Definition 1. Information system [2] is a pair $SI = (U, A)$, in which $U = \{t_1, t_2, .., t_n\}$ is a finite set of objects, A is a non-empty finite set of attributes and $a : U \rightarrow V_a$ with all $a \in A$. Set V_a is called the value set of attribute a.

Definition 2. With any information system $SI = \{U, A\}$ and a non-empty set of attributes $B \subseteq A$, an *information function* B is defined as follows [2]:

$InfB = \{(a, a(x)) : a \in B\}$ with all $x \in A$.

In special case $B = A$, then set $\{InfA(x) : \mathrm{x} \in A\}$ is called *information set* A, abbreviated as $INF(A)$.

One of basic characteristics of rough set theory is to store and process data that is ambiguous, indistinguishable [3]. In information system as defined above, there can also be indistinguishable objects.

Definition 3. An indistinguishable relation, denoted by $IND_A(B)$, is defined as: $IND_A(B) = \{(x, x') \in U^2 | \forall a \in B : a(x) = a(x')\}$, in which:

- B: an attribute set of objects, $B \subseteq A$.
- x, x': any two objects belonging to U.

Then $IND_A(B)$ is an *equivalence relation* B [3].

When two objects x, x', that $(x, x') \in IND_A(B)$, then two objects x, x' is called *indistinguishable* by attributes in B. When considering a definite information system, symbol A is often omitted, and we will abbreviate it as *IND(B)* instead of $IND_A(B)$. Equivalence class containing x of *indistinguishable relation* on B is denoted by $[x]_B$.

2.3 Reference-Specific Vector and Similarity

Definition 4. Reference-specific vector VA_j of attribute A_j corresponding to reference of transactions $(q_1, q_2, .., q_m)$ is determined [12] as follows:

VA_j =

q_1	q_2	...	q_m
M_{1j}	M_{2j}	...	M_{mj}

Definition 5. Similarity measure [12] of two attributes A_k, A_l, with two reference-specific vectors corresponding to set of transactions $Q = (q_1, q_2, .., q_m)$ of:

$VA_k = (M_{1k}, M_{2k}, .., M_{mk})$ and $VA_l = (M_{1l}, M_{2l}, .., M_{ml})$, is determined by *cosine measure* as follows:

$$s(A_k, A_l) = \frac{VA_k * VA_l}{\|VA_k\| * \|VA_l\|} = \frac{\sum_{i=1}^{m} M_{ik} * M_{il}}{\sqrt{\sum_{i=1}^{m} M_{ik}^2} * \sqrt{\sum_{i=1}^{m} M_{il}^2}} \quad (1)$$

3 Proposed Vertical Fragmentation Algorithm Based on KR Rough Clustering

3.1 KR Rough Clustering Algorithm

The most common rough clustering technique [2] is derived from primitive *k-Means* clustering. The goal is to randomly generate k clusters from n objects. Assume that objects are represented by m-dimensional vectors.

Each cluster is also represented by a m-dimensional vector, which is the *focus* or *vector* for that cluster. The process starts by randomly selecting k focuses of k *clusters*. Objects are assigned to one of k clusters based on the minimum value of the distance $d\ (v,\ x)$ between the object vectors $v = \{v_1, \ldots, v_j, \ldots, v_m\}$ and cluster vectors $x = \{x_1, \ldots, x_j, \ldots, x_m\}$ with $1 \le j \le m$. Distance $d(v, x)$ given: $d(v, x) = |v - x|$, is usually the Euclidean standard [5].

The process stops when the focuses of the cluster are stable, i.e. the *focus vectors* in the previous iteration coincide with the new *cluster focus* in the current iteration. Combining rough set into *k-Means* clustering [6] requires the addition of concepts of *lower approximation* and *upper approximation*. In particular, appropriate focuses will be calculated to determine whether an object will be assigned to lower approximation or upper approximation of each cluster. *KR rough clustering algorithm uses three basic properties*:

(1). Each object belongs only to one *lower approximation.*
(2). If the object belongs to a *lower approximation*, it also belongs to a corresponding *upper approximation*.
(3). An object belongs to at least two *upper approximations* if it does not belong to any *lower approximation*.

Describe the KR rough clustering improvement algorithm in the following steps:

Step 1: Calculate the cluster focuses according to primitive *k-Means*, with modifications including *lower approximation* and *upper approximation* [9].
Step 2: Determine whether an *object* is assigned to a *lower approximation* or *upper approximation* of a *cluster*.
Step 3: Determine the distance to the previous focus.

For each object vector v, distance $d(v, x_j)$ between v and the cluster focus x_j, there are two options to identify members of an object [10]:

Option 1. Determine the nearest focus [6] by the formula:

$$d_{\min} = d(v, x_i) = \min_{1 \le j \le k} d(v, x_j) \tag{2}$$

Option 2. Check the distance with the nearest cluster focus and other focuses: $T = \{t : d(v, x_i) - d(v, x_j) <= Th_i, i \neq j\}$ [11].

- If $T \neq \varnothing$ then v belongs to *upper approximation* of two or more clusters.
- If $T = \varnothing$ then v belongs to *lower approximation* of only one cluster.

3.2 Proposed KR Rough Clustering Algorithm

KR algorithm

Input: - *D*: Set of n objects to be clustered;
- *k*: Number of clusters;
- Threshold Th_i ;

Output: Set of clusters of *D;*

Algorithm
Begin
1. Initialize randomly k focuses of the derived objects $x=\{x_1,\ldots x_k\}$;
2. **Repeat**
3. Assign objects v to the upper and lower approximations of the clusters; /* *Cluster* */
4. Calculate the distance $d(v, x_i)$, $d(v,x_j)$ between objects v with the cluster focus x_i , x_j ; /* $1 \le i,j \le k$ */
5. **If** $(d(v, x_i) - d(v, x_j) <= Th_i)$ **Then** object vector v will not belong to any lower approximations /* by nature 3*/ ;
6. **Else** $d(v, x_i)$ is minimal;
7. Update focus x_i with new focus;
8. **If** the cluster focus coincides with the previous iteration **Then** stop;
9. **Else** go back to Step 2;
10. **Until** <*Cluster focuses do not change*>

End.

3.3 Evaluation of KR Rough Clustering Algorithm

- KR rough clustering solution *is similar to KO* [12], which is capable of grouping objects in different clusters. In addition, KR also generates more clusters than number of clusters needed to describe the data depending on the measurement distance. This causes the opportunity for an object to be high when *clustering* in the same cluster [1].
- However, *KR r*ough clustering solution proceeds with a large set of data, making the solution more complex, degree of overlap among clusters to increase, so calculating the focus is slower than primitive *k-Means.*
- KR algorithm complexity is $O(t^*n^*k)$, in which t is number of iterations, n is number of objects to be clustered, and k is number of clusters. However *t, k* are usually very small compared to *n* when the dataset is large enough and contains many objects. Therefore, the complexity is usually calculated as $O(n)$. This complexity is more optimal than vertical clustering algorithm according to attribute affinity such as BEA algorithm [7] of $O(n^2)$.

4 Experimental Results of KR Rough Clustering Algorithm

We compared experimental results of vertical fragmentation according to KR rough clustering and primitive k-Means by total time cost and memory cost. Dataset installed [8] consists of 20 objects as (Table 1):

Table 1. Dataset D consists of 20 instance

@NAME = Instance 1	@NAME = Instance 6	@NAME = Instance 11	@NAME = Instance 16
5.1 3.5 1.4 0.2	4.4 2.9 1.4 0.2	5 3 2 1	20 50 52 21
@NAME = Instance 2	@NAME = Instance 7	@NAME = Instance 12	@NAME = Instance 17
4.9 3.0 1.4 0.2	4.9 3.1 1.4 0.2	15 13 12 11	10 15 52 21
@NAME = Instance 3	@NAME = Instance 8	@NAME = Instance 13	@NAME = Instance 18
4.7 3.2 1.3 0.2	5.4 3.7 1.5 0.2	30 60 52 51	21 25 25 22
@NAME = Instance 4	@NAME = Instance 9	@NAME = Instance 14	@NAME = Instance 19
4.6 3.4 1.7 0.2	4.8 3.7 1.5 0.2	50 40 42 41	11 15 35 42
@NAME = Instance 5	@NAME = Instance 10	@NAME = Instance 15	@NAME = Instance 20
5.0 3.6 1.4 0.2	4.8 3 1.4 0.1	30 50 42 31	11 25 45 45

With k-Means algorithm:

- *Experiment with* $(k = 6)$*, result as (Fig. 1):*

```
========== KMEANS - SPMF 2.09 - STATS ============
 Distance function: euclidian
 Total time ~: 8192 ms
 SSE (Sum of Squared Errors) (lower is better) : 64.80000000000001
 Max memory:0.6792984008789062 mb
 Iteration count: 4
===================================
```

Fig. 1. Clustering result by k-Means algorithm (k = 6)

With KR rough clustering algorithm:

After similar experiment with number of clusters ($k = 6$), experimental results of *vertical fragmentation according to KR rough clustering* (similar to KO [12]), Fig. 2:

```
======
Improved cluster KR =============
 Distance function: euclidian
 Total time ~: 16 ms
 SSE (Sum of Squared Errors) (lower is better) : 248.76096491228066
 Max memory:1.2878875732421875 mb
 Iteration count: 8
=======================================
```

Fig. 2. Clustering results by KR with (k = 6)

Based on above two experimental results [8], the pager compiles a comparison table between two algorithms as primitive *k-Means* and proposed *KR* algorithm according to 3 tests, corresponding to number of clusters k selected ($k = 6$; $k = 13$; $k = 15$) as (Table 2).

Table 2. Comparison of KR and k-Means clustering results

Algorithm	Number cluster k	Total time (ms)	Sum of squared errors (Min)	Max memory usage (Mb)	Frequent itemsets count
k-Means	$k = 6$	8192	64.8000	0.6793	4
	$k = 13$	2623	455.4550	1.3000	3
	$k = 15$	1689	751.6216	1.6000	3
KR improved	$k = 6$	16	248.7609	1.2878	8
	$k = 13$	15	548.8960	1.2879	8
	$k = 15$	15	548.8960	1.2879	8

5 Conclusion

In this paper, we have proposed an improvement in the vertical fragmentation problem in distributed data based on *k-Means* rough clustering technique by combining *distance and similarity with upper and lower approximations.* In particular, calculate

appropriate focuses to determine whether an object will be assigned to lower approximation or upper approximation of each cluster [11].

Experimental results using KR rough clustering technique show:

- With a small number of clusters k ($k = 6$), *k-Means* algorithm has large total time, satisfactory error average cost and memory cost. Meanwhile, KR rough clustering algorithm optimizes all three criteria.
- When increasing number of *clusters* k ($k = 13$, $k = 15$), *KR* algorithm clearly expresses optimizations on all three criteria in comparison with *k-Means* algorithm. However, error average cost of KR is high as both *upper* and *lower approximations* are to be considered during the process of updating the new focus.

Complexity KR is usually calculated as $O(n)$. This complexity is more optimal than k-Means clustering algorithm [9] as $O(t^*n^*k)$ in which t is number of iterations, k is number of clusters, and n is number of objects on the set D to be clustered.

References

1. Darabant, A.S., Darabant, L.: Clustering methods in data fragmentation. Rom. J. Inf. Sci. Technol. **14**(1), 81–97 (2011)
2. Bean, C.L., Kambhampati, C., Rajasekharan, S.: A rough set solution to a fuzzy set problem. In: Proceeding of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), World Congress in Computational Intelligence, Honolulu, Hawaii (2002)
3. Bean, C.L., Kambhampati, C.: Automonous clustering using rough set theory. Int. J. Autom. Comput. **5**(1), 90–102 (2008). ISSN 1476-8186
4. Darabant, A.S.: A new approach in fragmentation of distributed object oriented databases using clustering techniques. Studia Univ. Babes-Bolyai, Informatica **3**(2), 91–106 (2005)
5. Sinwar, D., Kaushik, R.: Study of Euclidean and Manhattan distance metrics using simple k-means clustering. IJRASET **2**(V), 270–274 (2014). ISSN 2321-9653
6. Park, H.-S., Lee, J.-S., Jun, C.-H.: A k-means-like algorithm for k-medoids clustering and its performance. Department of Industrial and Management Engineering, POSTECH, South Korea (2016)
7. Hui, M., Schewe, K.D., Kirchberg, M.: A heuristic approach to vertical fragmentation incorporating query information. In: Seventh International Baltic Conference on Databases and Information Systems, pp. 69–76. IEEE (2006)
8. http://www.philippe-fournier-viger.com/spmf/
9. Jun, S., Kobayashi, S.: Large-scale k-means clustering with user-centric privacy-preservation. Knowl. Inf. Syst. **25**(2), 253–279 (2010). ©Springer
10. Shalini, S., Singh, N., Chauhan, C.: K-means v/s K-medoids: a comparative study. In: National Conference on Recent Trends in Engineering and Technology, May 2011
11. Okuzaki, T., Hirano, S., Kobashi, S., Hata, Y., Takahashi, Y.: A rough set based clustering method by knowledge combination. IEICE Trans. Inf. Syst. **E85-D**(12), 1898–1908 (2002)
12. Luong, V.N., Nguyen, H.H.C., Le, V.S.: An improvement on fragmentation in distribution database design based on knowledge-oriented clustering techniques. Int. J. Comput. Sci. Inf. Secur. (IJCSIS)-USA **13**(5), 13–17 (2015). ©IJCSIS publication 2015 Pennsylvania

Nature of Computation and Communication

Architectural Framework for Context Awareness and Health Conscious Applications on Mobile Devices

Huynh Trong Duc[1], Phan Cong Vinh[2(✉)], and Nguyen Dang Binh[1]

[1] University of Science - Hue University, Hue, Vietnam
huynhduc68@gmail.com, ndbinh@hueuni.edu.vn
[2] Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
pcvinh@ntt.edu.vn

Abstract. In recent years context-aware applications have emerged on smart devices and have become a new development trend. Context-aware applications allow for users to adapt to an ever-changing environment and also enhance interaction between users and their mobile devices. In the field of health care, along with the vigorous development of science and technology and especially the Internet, wireless mobile devices have changed the delivery of health care services from "brick & mortar medical facilities" to an "anywhere, anytime" service model.

This paper proposes a framework for the application of contextual healthcare on mobile devices with the idea of providing health care "anywhere – anytime". The aim is to improve health care service by enabling them to become easily accessible with the result being a higher quality of life for all users up to and including the saving of lives.

Keywords: Context aware health care · Mobile device · Support system

1 Introduction

The rapid development of the mobile industry has resulted in smartphones becoming center stage in the digital arena. Recently, manufacturers are focusing on developing smartphone tracking in healthcare applications. Most notably with wearable tech to monitor heart rate, blood pressure, body temperature, etc. which will provide vital information to the user. However, these devices are not secure and such information could be accessed by other parties making security a concern.

The context-aware health care application model that we propose will collect user information and build a "health profile" for each individual. Each user will have their unique profile specific to his or her own condition. It will be based on user information recorded in a medical context (via a measuring device) and placed in a medical database. The application will provide users with useful updated health information coupled with medical advice directly related to the user's needs. And at the same time it will caution a person about symptoms which may lead to health complications or send immediate signals to relatives or doctors in an emergency.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 175–183, 2018.
https://doi.org/10.1007/978-3-319-77818-1_17

The rest of paper is organized as follows. Section 2 shows an overview of related research. In Sect. 3, modeling context aware sensibility for mobile systems and theoretical basis for model building are introduced. In Sect. 4, we propose a framework for context-aware healthcare applications for the mobile platform. Finally, Sect. 5 draws conclusions and comments on further research work.

2 Related Research

2.1 Contextual and Context Aware

Dey [2] defines context as "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the applications and the users themselves".

Schmidt et al. [1] defines C context space with a definition of a combination of context parameters, ontology domain elements, and service descriptions. C = {U, P, L, T, D}, where U is the set of user & role factors, P is action & process, L is location, T is time, D is device, I is the information object available, S is the service available or described. A specific context is a point in the context space.

Dey and Abowd [3] define context awareness as "A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task". This definition tells us how to determine whether an application is context-aware or not. According to Dey, context is an attribute of a system that uses context to provide information or services related to the user, in which the relevance depends on the user's task.

2.2 Contextual Aware Applications in Health Care

The application of context aware technology in the field of health care has a lot of scientists interested in the research as follows:

In [4], the authors used context aware mobile agents to aid critical decision making in emergency medical applications.

In [7], the paper presents research in the design and implementation of a framework for cost cutting electronic healthcare delivery services for rural and suburban communities. This is achieved through the development of a semantic web services framework that would be deployed providing wireless mobile healthcare delivery services and health management services for rural African communities.

In [10], the study focuses on the design of the framework that supports an active repository in developing a component. To test the proposed Component-Based Development (CBD) intelligent framework, the healthcare monitoring system has been presented.

In [9], the system is based on context-awareness and case-based logic. It collects patients' newest context and provides a particular patient's details to the nurse 24 h call system allowing the nurses to provide urgent care.

In [5], the author has proposed a system for using mobile phones to track patients remotely, using the advantages of serial ports available in mobile phones to make a common interface with patient monitors.

In [6], a intelligent, context-aware monitoring home health care system was developed. In this system, sensors are used to collect data from a patient and then sent to a center for monitoring, analysis and to identify crucial patient details while the patient remains at home. The identification of an abnormal patient situation can activate a local device or commence an interaction directly with the person or issue an emergency message.

In [8], it is a proposed and subsequently implemented policy based architecture that will allow autonomous and continuous monitoring of patient thereby providing continuous necessary medical information to hospital personnel by utilizing software agents and wireless sensing technologies.

Most of the health care systems that have been proposed are for patients, although supervision and care are often limited to specific areas, such as in hospitals, nursing homes, etc. So we propose a healthcare model that will provide support for everyone, without the limitations of time and distance.

3 Modeling Context Aware Sensibility for Mobile Systems

The context aware application system can be described in the following expression:

F: $P \rightarrow Q$

With:

F: set of rules.
Q: set of recommendations or capabilities, for example: in healthcare applications Q is a set of the user's health states; in the decision support application, Q is a set of user actions.

$P = L \times T \times E \times U \times V$: context space of the system.

L: spatial context variables include: location, related objects (people, widgets nearby).
T: time context variable.
E: environmental context variables such as temperature, humidity, weather, etc.
U: user context variables include: profile information related to the user.
V: application context variables also calls the context variable or target context variable containing context information that has a major impact on the application. For e.g. in health care applications, V is information about the health of the user, including: body temperature, blood pressure, heart rate, etc. With the decision support application, V is the user behavior.

The process of arguments in the application is to identify a mapping from the set P to the set Q. Let p be a context in P($p \in P$), q be a possibility in Q($q \in Q$), then:

$$p \mapsto Q$$

It means that for a value p ∈ P we will have the corresponding value q ∈ Q.

So to solve the problem we have to determine the set Q and the relation holds P and Q.

To determine the set P we must define the context variables L, T, E, U, & V. These variables are collected through sensors and user information. Sensor sources and context information are often not uniform, so there needs to be a context retrieval department (Fig. 1). Another problem is in the presentation of contextual knowledge within the system and in applications where we use formal logic, descriptive logic, ontology or a combination of forms.

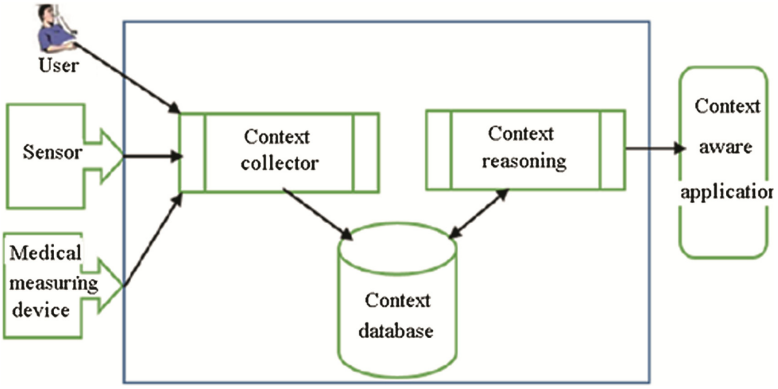


Fig. 1. Context - aware application model

Determining correlations P and Q use reason based on inference rules and RDF, OWL, SWRL or build a hybrid language based on these languages.

3.1 Logical Description Knowledge Base

Descriptive logic (DL) is designed as an extension of the semantic frame and semantic network and these types were not provided semantic based on formal logic. The complex concepts in the description logic are constructed using the Attribute Language (AL) or the extended languages of AL, called the "descriptive language" which will help us to formulate new concepts.

e.g.: Suppose we have elementary concepts such as "person" and "Male":
Man ' Person ∩ male.
Woman ' Person ∩ ¬male.
Father ' Man ∩ ∃hasChild.Person.
With: C ∩ D is intersection of concept C and D
¬D is negative of concept D.

The knowledge based system represented by the descriptive logic contains two main components (Fig. 2):

- TBox (Terminological Box) presents the terminology of an application domain. TBox defines the concepts and it specifies concept hierarchies which portray how atomic concepts and atomic roles are interrelated.
- ABox (Assertion Box) is an assertion component that describes facts associated with concepts and roles inside the knowledge base. ABox contains assertions on named individuals in terms of vocabulary. ABox has extensional knowledge about the domain of interest called membership assertion.

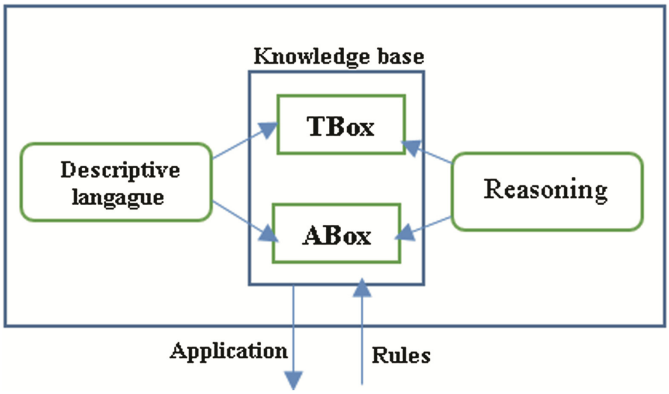


Fig. 2. General structure of the description logic system

Ex: Doctor(Peter), Patient(Alan), Careof(Peter, Alan), HasFever(Alan, high), Father(Jol, Alan), RelativeOf(Peter, Alan).

$$\text{Doctor(Peter)} \cap \text{Careof(Peter, Alan)} \rightarrow \text{DoctorOf(Peter, Alan)}$$

$$\text{Father(Jol, Alan)} \cap \text{Careof(Jol, Alan)} \cap \text{Patient(Alan)} \rightarrow \text{RelativeOf(Jol, Alan).}$$

3.2 Ontology

In computer science and informational science, an ontology is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain of discourse. Thus, it's a practical application of philosophical ontology, with a taxonomy.

An ontology compartmentalizes the variables needed for some set of computations and establishes the relationships between them.

Common components of ontologies include:

Individuals: Instances or objects (the basic or "ground level" objects).

Classes: Sets, collections, concepts, classes in programming, types of objects, or such kinds of things.

Attributes: Aspects, properties, features, characteristics, or parameters that objects (and classes) can have.

Relationships: Ways in which classes and individuals can be related to one another.

For example: The Ontology medical knowledge base has:
Individuals: doctors, patients, relatives, etc.
Concept: sick, symptom, situation, treatment.
Attributes: gender, body temperature, heart rate.
Relationship: Identified by, symptomatic, treatment.

3.3 Multi-agent System

Agent is a complete computational system or program placed in a certain environment that is capable of operating autonomously and with flexibility in that particular environment to achieve its intended purpose.

A multi-agent is a set of agents that work together in a system; each of which can have different purposes but the entire agent system is oriented toward the same goal through various interactions.

When used in multi-agent systems, ontology is a semantic structure that is referenced in the communication process between agents, which enables agents with different knowledge domains to understand each other in the exchange process. By understanding the messages it receives.

4 Framework for Context-Aware Healthcare Applications on Mobile System

4.1 Architecture Framework

The main component of the context-aware healthcare support system is its portability, which integrates the following devices (Fig. 3):

- **User interface:** Allows entering user context related information. Exporting information and giving supportive health care for users when necessary or when the user requests it.
- **Environmental sensor:** Provides contextual information about environment such as temperature, pressure and weather information.
- **Medical measuring devices:** provides medical measurements of the user such as heart rate, blood pressure, blood glucose, etc.
- **Locater:** Used to locate the user to provide timely help in cases of emergency.

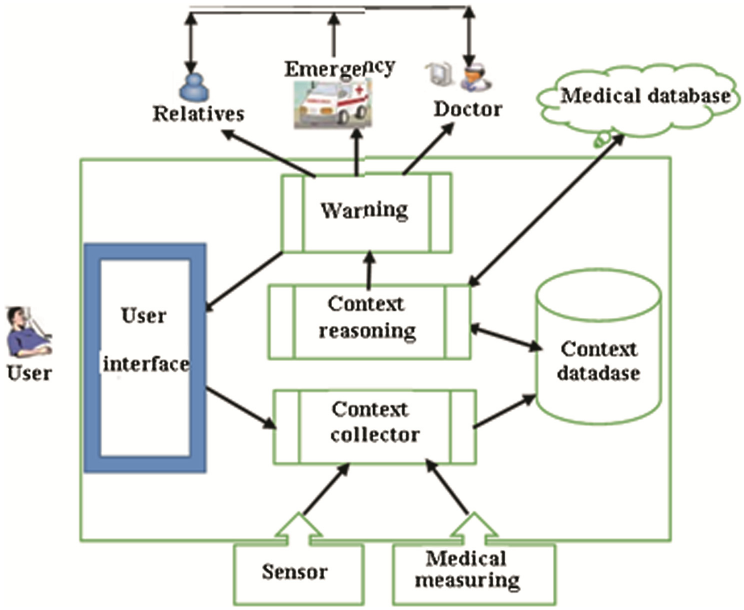


Fig. 3. Context-aware health care system on mobile devices

Components of portability:

- **Context collecting:** collecting context and identifying user context information.
- **Context databases:** organizing and storing user context databases.
- **Medical database:** Organizing and storing medical databases in the form of laws.
- **Contextual Argument:** Based on the contextual database and the health database, the argument provides user health information for warning.
- **Warnings:** Health information for when users request it or support for users health when needed. In case of emergency sends information and emergency signals to doctors, relatives, or for the emergency switchboard for timely critically needed solutions.

Participants in the system include:

- **User:** A person who needs health care support.
- **Relatives:** Those who support health care for users when necessary.
- **Doctor:** He directly supports the care, treatment and health advice for users when necessary.
- **Emergency call center personnel:** The medical facility where the emergency services are registered by the user, will provide emergency medical assistance in case of emergency.

4.2 Mechanism of Operation

(a) Service registration: A user who subscribes to a healthcare service will be provided with a wearable device (users can use a compatible mobile device) and a code that allows an installation program and connects to the system.

When the device is activated, the user logs into the system by the code provided, then enters their relevant information and connects with the support services:

Information provided includes: personal information and health claims.

Registration for support services includes: consultations, emergency services and relative's phone numbers.

The context collector accumulates environmental and health contexts, including health metrics. On this basis the system will compile a health profile for each user.

(b) Operating system: After a t cycle, if there is a change in the environmental information from the sensor and the measurement from the medical measuring device, the context collector will update the context database.

Based on contextual database information, contextual history and medical databases, the contextual argument will infer and then provide the user's health result for warning.

Based on the results obtained from the set the alert will send information according to each case as follows (Fig. 4):

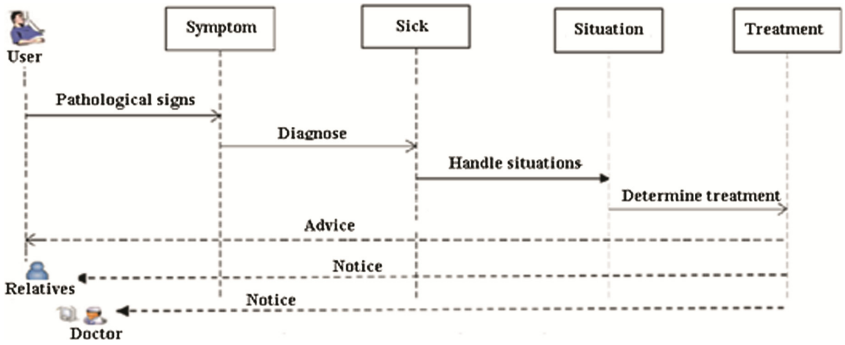


Fig. 4. Process reasoning in the system

Normal: Displays user health information, advice on health care for users such as exercise, sleep, diet and more.

Problem: Display health information to the user and send information to the doctor and family to follow up treatment support and prompt the user to comply with treatment schedules and advise on medical issues to support treatment.

Emergency: Send information to emergency services, doctors and relatives for timely assistance.

5 Conclusion

Through the study of the context sensibility in mobile platforms we have proposed a model of context aware support for mobile users and a theoretical basis for modeling. This model will be the basis for the production of health care support devices in the future contributing to improving the quality of life for people in any area and in real time.

In further development research we will finalize the detailed components throughout the model. Specifically, research develops solutions for context collectors and context filters and studies the theoretical basis for storing contextual databases and building language arguments on the context database. Research also studies the organization of medical databases supporting the arguments on the system.

References

1. Schmidt, A., Beigl, M., Gellersen, H.-W.: There is more to context than location. University of Karlsruhe (1999)
2. Dey, A.K.: Providing architectural support for building context aware applications. Georgia Institute of Technology (2000)
3. Chen, G., Kotz, D.: A survey of context-aware mobile computing research. Technical report TR2000-381, Department of Computer Science, Dartmouth College (2000)
4. Burstein, F., Zaslavsky, A., Arora, N.: Context-aware mobile agents for decision-making support in healthcare emergency applications. This Project was Funded by Monash University Grants (2005)
5. Figueredo, M.V.M., Dias, J.S.: Mobile telemedicine system for home care and patient monitoring. In: 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, vol. 2, pp. 3387–3390 (2004)
6. Alessandro, C., Loques, O., Leite, J.C.B.: Intelligent context-aware monitoring in home care. In: 2nd International Conference on E-Medical Systems, EMEDISYS 2008, Tunisia (2008)
7. Oladosu, J.B., Ojo, S.O., Adigun, M.O.: Framework for a context-aware mobile e-health service discóvery infrastructure for rural suburban healthcare. J. Theor. Appl. Inf. Technol. **6**(1), 081–091 (2009)
8. Vassis, D., Belsis, P., Skourlas, C., Pantziou, G.: A pervasive architectural framework for providing remote medical. In: 1st International Conference on Pervasive Technologies Related to Assistive Environments, Greece, p. 23 (2008)
9. Wei, Q.: A context-aware system to cómmunicate urgency cues to health care workers. Master of Telematics - Cómmunication Networks and Networked Services. Norwegian University of Science and Technology (2014)
10. Gelogo, Y.E., Kim, H.-K., Chung, Y.-K.: A design of CBD intelligent framework for U-healthcare application. Int. J. Softw. Eng. Appl. **9**(6), 247–256 (2015)

Holistic Personas and the Five-Dimensional Framework to Assist Practitioners in Designing Context-Aware Accounting Information System e-Learning Applications

Hien Minh Thi Tran(✉), Farshid Anvari, and Deborah Richards

Macquarie University, Sydney, Australia
hientran@minh-hien.com, farshid.anvari@acm.org,
deborah.richards@mq.edu.au

Abstract. Despite a growing demand for e-learning applications, there is a lack of research on eliciting requirements for design of context-aware Accounting Information Systems (AIS) e-learning applications (CAAISeLA) that incorporate the multiple-perspectives of adult learning theories in a workplace environment. This paper describes how Holistic Persona, an archetypical learner and user of an AIS application, can assist practitioners to elicit requirements for design of context-aware AIS e-learning applications for employees. It presents empirical results of twenty employees at a research-intensive Australian university who participated in evaluating the effectiveness of various e-learning resources. The results demonstrate that Holistic Personas and the Five-Dimensional Requirements Elicitation Framework (5DREF) can assist practitioners in designing context-aware AIS e-learning applications.

Keywords: Context-aware e-learning application · Adult learning theory Knowledge · Accounting information system · Holistic personas

1 Introduction

Consideration of diverse perspectives of users' activities balanced with multiple theories of learning in a workplace environment leads to more successful e-learning applications. In order to gather information about users' requirements at the business process level, each group of users should participate to contribute their components of the requirements. The main usability goal of e-learning applications is ease of its operation so that the learners can concentrate on learning new materials. User-Centered Design (UCD) methodology considers the goals of the users as a primary requirement of software applications and this can be better achieved if educators take into account each learner's characteristics and expertise [7]. However, it is often not possible for practitioners to meet with all users of diverse groups. Persona, an archetypical user of an application, is widely used by software engineers to design applications that are focused on the needs of users. For example, Madsen et al. [15] used personas, which they created, to assist with designing a personalized professional development website. Anvari and Tran [4]

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 184–194, 2018.
https://doi.org/10.1007/978-3-319-77818-1_18

have proposed Holistic Persona, persona that incorporate demographics, personalities, intelligence, knowledge and cognitive processes, designed to more closely resemble the end users. Holistic Persona have been used in the design of teaching material in computing [3].

Context-aware applications have been deployed extensively [24]. The goal of context-aware applications is to make interaction with computers easier [11]. However, most applications address learners' context such as location, time, surrounding resources, learners' learning styles etc. Our extensive literature search was unable to identify any study that reports how users of AIS select e-learning resources to learn in the context of workplace environments and how practitioners consider the multiple perspectives and multiple characteristics of users towards learning AIS. This paper describes how Holistic Personas can assist practitioners to elicit requirements for designing e-learning resources of contest-aware AIS e-learning applications for employees. The findings of our empirical study will be valuable for academics and professionals who are interested in designing context-aware e-learning applications for employees. The rest of the paper is organized as follows: (2) background and related work; (3) methodology: the 5DREF and the design of e-learning resources using Holistic Personas; (4) results; (5) discussion; (6) contributions and limitations; (7) conclusion.

2 Background and Related Work

There are many context-awareness definitions. For example Li et al. [14] defined five context dimensions as who (user), what (object), how (activities), where (location), and when (time). Verbert et al. [24] presented the components of various context definitions: location, time, physical conditions, computing, resource, user, activity, and social. Abowd and Mynatt [1] identified the five W's (Who, When, What, Why, Where) as the minimum necessary information to understand context. Dey, Abowd and Salber [11, p. 106] stated that context-aware applications consider "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves*". Synthesizing context-awareness definitions from the literature [1, 11, 14, 24] and reflecting on experience in designing AIS teaching materials, we have identified that in addition to *why, who, when, what and how, evaluation* is needed to characterize the interaction between a user and a CAAISeLA. **Why:** The 'Why' addresses context-aware features such as learning style, teaching methods and strategies. Different learning theories require different sets of tools to support practitioners to develop learning resources. For example in a direct instructional approach, practitioners break a new body of knowledge into small steps with clear objectives to provide learners with the opportunity to practice with feedback on each step learnt [16]. According to Constructivist Learning theory [20] learning is an active process of creating meaning from experiences based on the learner's current or past knowledge [10] and constructivist learning occurs when the '*learner actively builds a mental model of the system she is to learn*' [6, p. 74]. 'e-*Learning, or better computer supported learning, focuses on the individual's acquisition (or rather construction) of new knowledge and*

the technological means to support this construction process' [21, p. 204]. Socio-constructivist theory states that interaction between learners and their peers is a necessary part of the learning process [25]. Hence practitioners of Socio-constructivist theory provide forums to promote social interaction and learning. Practitioners of scaffolding instruction teach students concepts that are just above their knowledge level by encouraging them to do an activity to improve beyond their current skills level [12, p. 138]. Hence some practitioners include quizzes to provide scaffolding instructions or provide e-learning materials that are less guided such as scenario writings to encourage learners '*to produce some outputs that are not contained in or presented in the learning materials*' [6, p. 79]. **Who:** In the context of AIS e-learning, the 'Who' is diverse and each group of users has their specified learning objectives [22]. When users are not available for eliciting requirements, personas can be used in designing e-learning applications [15, 17, 19]. For example, Maier and Thalmann [17] created three personas, that represented three distinctive informal learner types that resulted in three service areas: '*individuation, interaction and information*' [17 p. 59]. Panke et al. [19] incorporated personas that offer information on how to integrate digital media into teaching. Anvari and Tran [4] and Anvari et al. [3] used Holistic Personas for design of learning applications. **When:** Users of AIS are adults. Adults are motivated to learn by internal factors [32]. Therefore to design AIS context-aware applications, practitioners should consider when learners are motivated to learn. Deci and Ryan's [8] Self-determination theory states that a person's motivation depends on the fulfilment of the needs for competence, relatedness and autonomy. **What:** In a work environment, learning materials need to promote learners' new knowledge that '*can be transferred back to job and utilization of new skills to enhance organizational performance*' [26, p. 194]. Bloom's Revised Taxonomy incorporates the Knowledge Dimension and the Cognitive Process Dimension. The Knowledge Dimension considers what learners need to learn. The 'What' to learn for a CAAISeLA contains *Factual, Conceptual, Procedural* and *Meta-cognitive knowledge* [2] of AIS. **How:** The 'How' considers learners' view of the Cognitive Process. According to Bloom's taxonomy there are six levels: *Remember, Understand, Apply, Analyze, Evaluate* and *Create* [2]. **Evaluation:** An e-learning application should have tools to provide timely feedback to learners so that they can self-evaluate their performances and for educators to monitor learners' activities and progresses [23].

3 Methodology

This section discusses the application of the 5DREF and Holistic Personas for the design of context-aware AIS e-learning resources and the procedure for the study.

3.1 The Five-Dimensional Requirements Elicitation Framework

To guide the design process we used the 5DREF that consists of five dimensions [23]. In a context of AIS e-learning applications, the 5DREF provides a framework for practitioners to focus on (1) why (need) educators teach in certain ways, (2) who (users) are involved in the learning process, (3) what (resources) and (4) how (activity, social)

learners learn, and (5) evaluate the teaching and learning process. Each dimension is discussed below.

(1) **Curriculum Development Dimension:** To address 'Why' certain learning resources are used by practitioners while they teach, we examine the *Curriculum Development* factor of the *Change Management* dimension [23]. E-learning resource requirements are elicited mainly from educators. For example, applying Piaget's [20] Constructivist Learning Theory and Deci and Ryan's [8] Self-Determination Theory, forums are used in teaching to allow learners to interact and construct ideas with peers and quizzes are used to scaffold learners.

(2) **User (or Learner) Characteristics Dimension:** The context-awareness dimension, referred as the *learner* dimension, addresses the 'who', the *User Characteristics*. To assist the practitioner who lacked participation of learners during the design process, we authored two Holistic Personas, Megan and Kim, to represent AIS users. The Holistic Personas, Megan and Kim, can be obtained from the authors.

(3) **Knowledge Dimension:** The Knowledge dimension addresses 'what' to teach. To be context-aware Online AIS needs to take into account learners' existing knowledge and experiences and allow learners to choose their own target levels of knowledge to be gained at the conclusion of sessions and select contents that are relevant to them. The Holistic Personas provided for prior knowledge of AIS which was taken into account for the design of e-learning materials. For example *factual* and *conceptual knowledge* materials are included for learners who had no prior knowledge of the subject matter and procedural knowledge materials for advanced learners to reflect and to construct meta-cognitive knowledge.

(4) **Cognitive Process Dimension:** The Cognitive Process dimension considers how learners learn. Cognitive Process dimension has six levels: *Remember, Understand, Apply, Analyze, Evaluate* and *Create* [2]. E-learning resource requirements are elicited from the Holistic Personas as well as educators. The *Cognitive Process* dimensions assist the design of e-learning resources that are context-aware to diverse learner groups who engage with the learning material using different levels of the cognitive processes. For example to encourage learners to learn at the higher rung of the cognitive process dimension, real-life worked-example scenarios [9] were provided.

(5) **Evaluation Dimension:** The Evaluation dimension helps practitioners to monitor the effectiveness of the teaching and learning processes. This study discusses educators as well as learners' reflective evaluation of the Online AIS.

3.2 The Design of Online AIS Learning Resources Using Holistic Personas

We developed an online AIS course, which allows the practitioner to enroll participants, through log files, follow their learning progress and interact with them. This section presents the design of context-aware e-learning resources that teaches Chart of Accounts (COA). The study was conducted between November 2015 and January 2016 and between May and July 2016. In this study we targeted the needs of AIS users who would

attend the study: administrative staff at middle rank and academic staff at researcher and lecturer levels. At the university where this research was conducted, the majority of the users of AIS (about 72%) were female. Hence for this study Holistic Personas, Megan and Kim, represented the users of AIS. Their different perspectives towards AIS influenced the design of the Online AIS study.

The AIS e-learning resources: In this study the effectiveness of the e-learning resources designed based on the Detailed Documentation method (*guided-construction*) and the Holistic Persona and Scenario method (non-guided construction) was studied. The knowledge, cognitive-process and curriculum development dimensions of the 5DREF provided guidelines for the design of e-learning resources. Considering Self-Determination Theory as a teaching and learning strategy, online AIS included various types of e-learning resources e.g. textual documents, visual posters, videos, quizzes, reflective exercises and forums for learners of various learning styles to self-direct learning. The Detailed Documentation method was designed using the direct instruction and scaffolding strategy. 'The learner's activities in the context of instructional dialogues can be referred to as "guided-construction"' [14, p. 82]. Three e-learning resources were textual materials with some visual images. The level of complexity was increased from easy to medium. Learning resources included quizzes and forum. The quizzes scaffolded learners at the cognitive process levels of '*Understand*' and '*Apply*'. The forum was provided for learners to post their work at the cognitive process level '*Analyze, Evaluate* and *Create*' [2] as well as to read and to raise questions. In short, the Detailed Documentation method provided learning resources with step-by-step instructions. The Holistic Persona and Scenario method was designed using Constructivist Learning with scaffolding strategy and Self-determination theory as a teaching strategy. Participants could construct their knowledge by understanding the posters, participate in the quizzes, reflect on questions raised in the forum and interact with peers via forum.

3.3 Procedure

Between November 2015 and July 2016, university employees were invited by the lead author via email to take part in the study. Participation was voluntarily and required completion of a consent form; participation in a pre-study questionnaire about demographics, their pre-knowledge of the university AIS and their ID numbers for enrolment. Participants were randomly assigned into Group A or Group B. Ethics approval was granted by Macquarie University's Human Research Ethics Committee, effective 13/10/2015 (ethics reference number 5201500782).

4 Results

The objective of this paper is to report on how users of AIS select e-learning resources and how practitioners consider the multiple perspectives and multiple characteristics of users towards learning AIS. We present only the results of demographics and e-learning materials that were accessed by participants.

Participants' demographics: Twenty-eight people from all faculties participated in the pre-study survey. Twenty-five people enrolled in the Online AIS. Five people never signed into the Online AIS hence they were excluded from the analysis.

The sample population for group A is 10 participants, for Group B is 10 participants and for the whole study is 20 participants. There were 10 Admin Officers, 2 Finance Officers, 3 Head Department Professionals, 1 Business Manager, 2 IT Officers, 1 Accountant and 1 Postgraduate Student.

Participants' access to each of the learning resources: Each participant accessed at least one e-learning resource. Figure 1 provides a graphic comparison of e-learning resource accessed for both groups. The average number of e-learning resources which participants accessed for Group A was 8.5 and for Group B was 8.10.

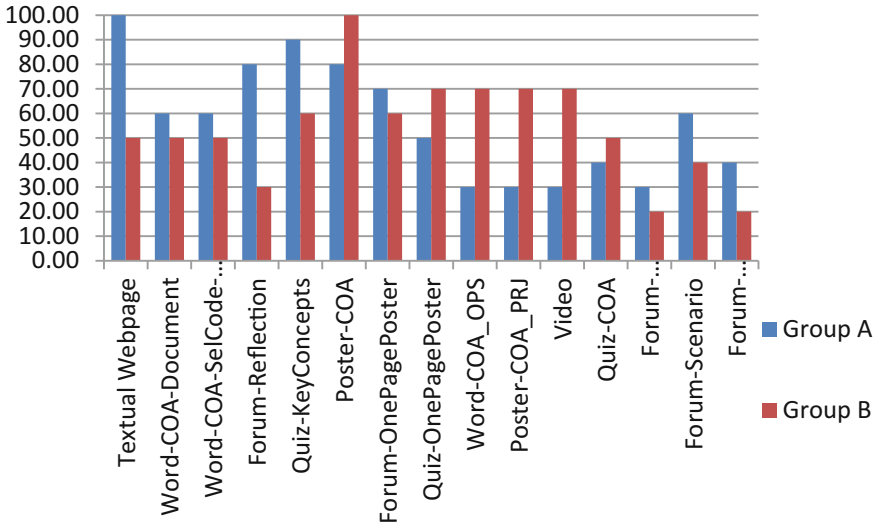


Fig. 1. Percentage of participants accessed learning resources

5 Discussion

The design of CAAISeLA for this study was guided by the five dimensions of the 5DREF [23]: who (the learners), why (the learning theories), what (the knowledge), how (the cognitive process) and finally an evaluation of the study. In this section we reflect on the study, its design and the results obtained.

(1) **The Learners:** The participants were staff at an Australian university. Participants received no financial gains. The highest number of participants for this study was Admin Officers (50% of the sample population) which indicates that Admin Officers think AIS is relevant and they need to learn COA. Participants' selections of e-learning resources indicate that CAAISeLA needs to provide different e-learning resources for different methods of teaching (Fig. 1). For example 90% for Group A

versus 60% for Group B took the core quiz. In the sample population 10% posted in the forum and 30% expressed about their experiences. Others only viewed the forums but did not post. These findings can assist practitioners to design context-aware eLearning resources for AIS users.

(2) **Learning theories (Curriculum development):** The results indicate that most participants in Group A (90%) quickly participated in the core quiz activity that gave them access to the Holistic Persona and Scenario method (Group B materials). Participants in Group B were slower in completing the course content and many Group B participants did not participate in the core quiz and hence did not get access to the Detailed Documentation method (Group A materials) which they would have had access if they finished the quizzes. Hence the participants in Group A who received '*guided-construction*' [6, p. 82] in the form of detailed documents that included step-by-step scaffolding instruction, '*to prompt students to attain the additional skills needed to reach this zone, teachers encourage them to learn by doing an activity*' [12, p. 138], were ready to test their knowledge quicker than Group B who initially received a description of Holistic Persona Megan, a poster and five different learning materials without any direct instruction. The results show (Fig. 1) that Group B accessed videos, a poster of COA for Projects ledgers (Poster-COA_PRJ) and a Word document of COA for Operations ledgers (Word-COA_OPS) much more than Group A (70% each for Group B versus 30% each for Group A). This finding indicates that users' choice of e-learning resources and activities depend on the methodologies used to design course contents. Hence CAAISeLA must have tools that are suitable for various practitioners so that they can design relevant and context-aware e-learning resources for learners. For example, Holistic Persona and Scenario writing would be more suitable to those who are highly self-directed in their learning and have grasp of the fundamentals of the subject because '*learners interpret concepts and principles in terms of the 'schemata' that they have already developed*' [5, p. 22]. It is of interest to note that during the study period, participants were reminded to complete the study. Many Group B participants asked for more time to do the studies and they were active during the study period, but the majority of them did not make any postings. This finding indicates that CAAISeLA should have tools to support the design of e-learning materials that incorporate the multiple-perspectives of learning theories so that practitioners can design context-awareness e-learning resources to meet the needs of the learners.

(3) **Knowledge:** A higher number of participants in Group A completed the post-study survey compared with Group B (90% for Group A versus 50% for Group B). The results indicate that participants found e-learning materials relevant and thus completed the study and gained more knowledge of the university's AIS. Hence e-learning resources that are aware of users' knowledge and their positions in an organization should incorporate input parameters from the learners and educators. For example, certain e-learning resources can be presented when the learner passes a certain level of the quiz test or holds a certain position in the organisation e.g. Accountant or IT Officer.

(4) **Cognitive Process:** The Cognitive Process explains how participants construct new knowledge from the learning resources provided. Quizzes promote conceptual understanding and, applying and analyzing. Eleven participants (64.71%) displayed their participation at the Cognitive Process level '*Analyze*' as they actively participated in the quizzes other than the core quiz and/or posted on the forum. Two participants (10% of the sample population) evaluated and created. The results also indicate that Group B participants might have been overwhelmed with the large amount of learning resources available and hence they were experiencing cognitive overload. Hence the Online AIS can be more effective if it is implemented targeting the mid-level of the Cognitive Process, '*Apply*'. To get learners to the higher rung of the Cognitive Process: *Analyze, Evaluate* and *Create* [2], educators need to have strategies to scaffold learners. For example the e-learning resources are made context-aware according to learners' cognitive processes and display resources accordingly.

(5) **Evaluation:** The results from this study reveal that most participants only viewed the forums and did not post their reflection. This finding is consistent with Nielson's 90–9–1 rule [18] that 90% of users read or observe; 9% of users contribute from time to time and 1% of users account for most contributions. This study has demonstrated that the Holistic Personas provided context-aware design guidelines and helped to predict the behavior of the participants. A noticeable feedback from the participants who completed post-study surveys was that they found quizzes built their confidence in AIS; most participants prefer the Detailed Documentation method and few (30%) posted reflection on the forums which indicates that their choice of learning strategy is guided-construction.

6 Implications and Limitations of Our Findings

Following are some recommendations derived from this study: (1) CAAISeLA needs to provide context-aware e-learning resources to facilitate self-directed learning for various learner groups. (2) CAAISeLA needs to allow learners to select content that are relevant to them. (3) Practitioners can apply UCD methodologies that take into account each learner's choice of learning strategy while designing context-aware AIS e-learning applications. (4) Although Socio-constructivist theory states that interaction between learners and their peers is a necessary part of the learning process [25], the results of this study indicate that learners of AIS wanted to be active but were not ready to interact with others. Hence forums are for facilitators to communicate, to provide model questions and answers and to guide learners. (5) When access to users of AIS is not available, Holistic Personas can assist practitioners to focus on users' needs in the design processes [3]. (6) The 5DREF can assist practitioners in providing context-aware e-learning resources to learners and educators. (7) The context-aware information provided by CAAISeLA can provide e-learning resources that are relevant to educators' teaching methods.

Limitation of our study: Our sample is small and the participants were from one university in Australia; hence care should be taken when applying the results to other scenarios. Further research is needed with larger sample sizes and diverse participants.

7 Conclusion and Future Research

We have demonstrated how Holistic Personas [4] can assist practitioners to design context-aware online AIS learning resources for employees and have provided examples of how Holistic Personas can assist practitioners to design context-aware online AIS learning resources. We have also demonstrated that the multiple-perspectives provided by incorporating multiple learning theories guided by the 5DREF are necessary in designing context-aware online AIS learning resources for diverse users. The novel contribution of this paper is the application of the Holistic Personas to the design of AIS and the empirical findings that extend the multiple-perspectives of the Adult Learning theory, the Self-Determination theory and the Constructivist Learning theory. The empirical findings are in line with the theory of andragogy [13], adults are self-directed learners and would actively learn online if learning resources are relevant and intrinsically motivate them. Context aware AIS e-learning resources need to include quizzes, videos, visual posters, textual documents and guided constructivist learning activities presented according to learners' demonstrated knowledge and preferences. In the context of AIS, reflective learning and scenario writings were too advanced for most professionals and as expected they visited forums only to read [18]. This study has provided empirical support for direct instructions [16] and guided construction. For a future study we plan to include strategies to encourage participants to learn at the higher rung of the cognitive process dimension. In sum this paper extends the literature in the application of personas in general and Holistic Personas in particular, and using the 5DREF to guide the multiple perspectives of e-learning theories and UCD in the design of context-aware e-learning resources for diversified groups of AIS users in working environments. We plan to study the effects of personality traits and intelligence on learners' choice of e-learning resources in future.

Acknowledgements. We thank staff at Macquarie University for their participation in the study. We appreciate the comments from the two anonymous reviewers.

References

1. Abowd, G.D., Mynatt, E.D.: Charting past, present, and future research in ubiquitous computing. ACM Trans. Comput.-Hum. Interact. **7**(1), 29–58 (2000)
2. Anderson, L.W., Krathwohl, D.R.: A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged edn. Longman, New York (2001)
3. Anvari, F., Richards, D., Hitchens, M., Babar, M.A.: Effectiveness of persona with personality traits on conceptual design. In: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering (ICSE), pp. 263–272 (2015)

4. Anvari, F., Tran, H.M.T.: Persona ontology for user centred design professionals. In: Proceedings of the ICIME 4th International Conference on Information Management and Evaluation, Ho Chi Minh City, Vietnam, pp. 35–44 (2013)
5. Biggs, J., Tang, C.: Teaching for Quality Learning at University. McGraw-Hill International, New York City (2011)
6. Chi, M.T.H.: Active-constructive-interactive: a conceptual framework for differentiating learning activities. Top. Cogn. Sci. **1**(1), 73–105 (2009)
7. De Marsico, M., Kimani, S., Mirabella, V., Norman, K.L., Catarci, T.: A proposal toward the development of accessible e-learning content by human involvement. Univ. Access Inf. Soc. **5**(2), 150–169 (2006)
8. Deci, E.L., Ryan, R.: Overview of self-determination theory: an organismic dialectical perspective. Handb. Self-Determination Res. 3–33 (2002)
9. DeLotell, P.J., Millam, L.A., Reinhardt, M.M.: The Use of deep learning strategies in online business courses to impact student retention. Am. J. Bus. Educ. (AJBE) **3**, 12 (2011)
10. Denton, D.: Enhancing instruction through constructivism, cooperative learning, and cloud computing. TechTrends **56**(4), 34–41 (2012)
11. Dey, A.K., Abowd, G.D., Salber, D.: A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Hum.-Comput. Interact. **16**(2), 97–166 (2001)
12. Jaramillo, J.A.: Vygotsky's sociocultural theory and contributions to the development of constructivist curricula. Education **117**(1), 133 (1996)
13. Knowles, M.S., Holton III, E.F., Swanson, R.A.: The Adult Learner. Routledge, Abingdon (2012)
14. Li, L., Zheng, Y., Ogata, H., Yano, Y.: A framework of ubiquitous learning environment. In: The Fourth International Conference on Computer and Information Technology, CIT 2004, pp. 345–350. IEEE (2004)
15. Madsen, A., McKagan, S.B., Sayre, E.C., Martinuk, M., Bell, A.: Personas as a powerful methodology to design targeted professional development resources (2014). arXiv preprint arXiv:1408.1125
16. Magliaro, S.G., Lockee, B.B., Burton, J.K.: Direct instruction revisited: a key model for instructional technology. Educ. Technol. Res. Dev. **53**(4), 41–55 (2005)
17. Maier, R., Thalmann, S.: Using personas for designing knowledge and learning services: results of an ethnographically informed study. Int. J. Technol. Enhanced Learn. **2**(1–2), 58–74 (2010)
18. Nielsen, J.: Participation Inequality: Lurkers vs. Contributors in Internet Communities. Jakob Nielsen's Alertbox (2006)
19. Panke, S., Gaiser, B., Werner, B.: Evaluation as impetus for innovations in e-learning—applying personas to the design of community functions. MERLOT J. Online Learn. Teach. **3**(2), 179–190 (2007)
20. Piaget, J.: The Construction of Reality in the Child. Routledge, London (1954)
21. Schmidt, A.: Bridging the gap between knowledge management and e-learning with context-aware corporate learning. In: Althoff, K.-D., Dengel, A., Bergmann, R., Nick, M., Roth-Berghofer, T. (eds.) WM 2005. LNCS (LNAI), vol. 3782, pp. 203–213. Springer, Heidelberg (2005). https://doi.org/10.1007/11590019_23
22. Tran, H.M.T., Anvari, F.: How reflective professionals design and evaluate financial information management systems courses. In: Proceedings of the 4th International Conference on Information Systems Management and Evaluation ICIME 2013, Ho Chi Minh City, Vietnam, pp. 263–271 (2013)

23. Tran, H.M.T., Anvari, F.: A five-dimensional requirements elicitation framework for e-learning systems. In: Proceedings of the ICCRD 8th International Conference on Computer Research and Development, Nha Trang, Vietnam (2016)
24. Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., Duval, E.: Context-aware recommender systems for learning: a survey and future challenges. IEEE Trans. Learn. Technol. **5**(4), 318–335 (2012)
25. Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press, Cambridge (1980)
26. Wang, M.: Integrating organizational, social, and individual perspectives in Web 2.0-based workplace e-learning. Inf. Syst. Front. **13**(2), 191–205 (2011)

Abnormal Behavior Detection Based on Smartphone Sensors

Dang-Nhac Lu[3(✉)], Thuy-Binh Tran[2], Duc-Nhan Nguyen[2], Thi-Hau Nguyen[1], and Ha-Nam Nguyen[1]

[1] University of Engineering and Technology, Vietnam National University in Hanoi, Hanoi, Vietnam
{nguyenhau,namnh}@vnu.edu.vn
[2] Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
tran.binh95@gmail.com, ducnhan@ptit.edu.vn
[3] Academy of Journalism and Communication, Hanoi, Vietnam
nhacld.di11@vnu.edu.vn

Abstract. There are a lot of applications were developed to take advance of smartphone sensors for utilizing the personal services such as health-care, walk-counting, routing etc. Users behavior analysis is attracted a lot of researches interested with various approaches. We proposed a novel framework to detect the abnormal driving behavior using smartphone sensors. It named Abnormal Behavior Detection System (ABDS). The system keep track the driver activities during he's trip based on smartphone sensors. The Practice Swarm Optimization (PSO) algorithm is used to automatically select suitable features extracted from sensors data. The oriented accelerometer is used to detect activity. The abnormal behavior is collected and labeled then detection by Artificial Neural Network (ANN). The implementation shown the promising results in case of seven activities (stop, moving, acceleration, deceleration, turn left, turn right and U-turn) with 86.71% accuracy.

Keywords: Activity recognition · Behavior recognition · Detecting behavior PSO algorithm

1 Introduction

The mobile phone is indispensable device in modern life and there is a lot of applications using sensor signals [1] for human activity analysis or behavior recognition. Recently, researchers concerned much about personal data analysis based on collected sensors data with various approaches [2]. Behaviors can be defined as sequences of primitive activities and it can be repeated time by time. Sometime it can be called complex activities. Driving behavior system is meaningful with traffic participant and defined by any activity analysis technical based on set of values such as: distances, gaps [3], time, angle [4] and velocity [5]. Hence, it could be help for recognizing the harmful activities when they are moving with their smartphone.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 195–204, 2018.
https://doi.org/10.1007/978-3-319-77818-1_19

The accelerometer is the most commonly sensor for reading motion signals and utilize in various application [6–8]. The smartphone sensors signal is time series data, which can be easy to collect but difficult to analysis. The quality of analysis depended on devices quality, environment conditional and sensitive applying model. In some cases, feature extraction technical is applied for human activity recognition, driving behavior problem [9, 10].

Vavouranakis et al. [11] proposed method to recognize driving behavior by smartphone sensor. However, the device is fixed in car and reoriented data by sensor fusion method, mobile coordinate system be reoriented by its coordinate system. The windowing technical utilize with 5 s on accelerometer data. Then, abnormal behavior are predicted by thresholds of x and y-axis accelerometer values. The 12 distinguish events about six safe and six unsafe behaviors deployed by their method.

Li et al. [4] developed system for detected dangerous driving behavior. It gathers accelerometer signal with ground truth position on taxi. The yaw angles is estimated by transformation matrix converter between vehicles coordinate and smartphone coordinate system and helping them to detect behavior on manual dangerous driving behavior set which define by accelerometer value threshold and 90% accuracy was received by their experiment.

Xu et al. [5] detected human behavior rules base on accelerometer with Fourier transform take 1 s to 8 s point for analysis and calculated velocity from accelerometer.

The one of challenges with driving behavior recognition using accelerometer is signal quality in difference devices. Following, analysis technical and models used for abnormal behavior recognition, which are complex and difficult. Hence, our problems is detecting user activity while they are moving, then the sequence of activity corresponding to their trip is background for system predicts and announce abnormal behavior.

The ABDS is divided into four parts: firstly, it collects label data, analysis and extract suitable features subset by PSO. Secondly, it predicts seven activity by some classification algorithms as Random Forest (RB), Naïve Bayes (NB), k-Nearest Neighbor (KNN) and Support Vector Machine (SVM). Finally, it predicts abnormal driving behavior by ANN. The experiment on collected data set with various driving types = {walking, bicycle, motorbike, bus, car} obtained the promising results with safe and unsafe behaviors compare to traditional methods.

2 The Proposed Framework

Our proposed method, named ABDS framework, is presented in Fig. 1. The system is consisting of four steps: firstly, the label data collected from each volunteer and features selection aim to predefine input value for predict activities. The raw data is converted into a set of features and then PSO are applied to select the suitable subset feature for driving activities recognition. Secondly, data with the best features subset is used for classifying and online recognition activity on smartphone. Finally, the abnormal driving behaviors are detected by ANN model which is trained in previous step.

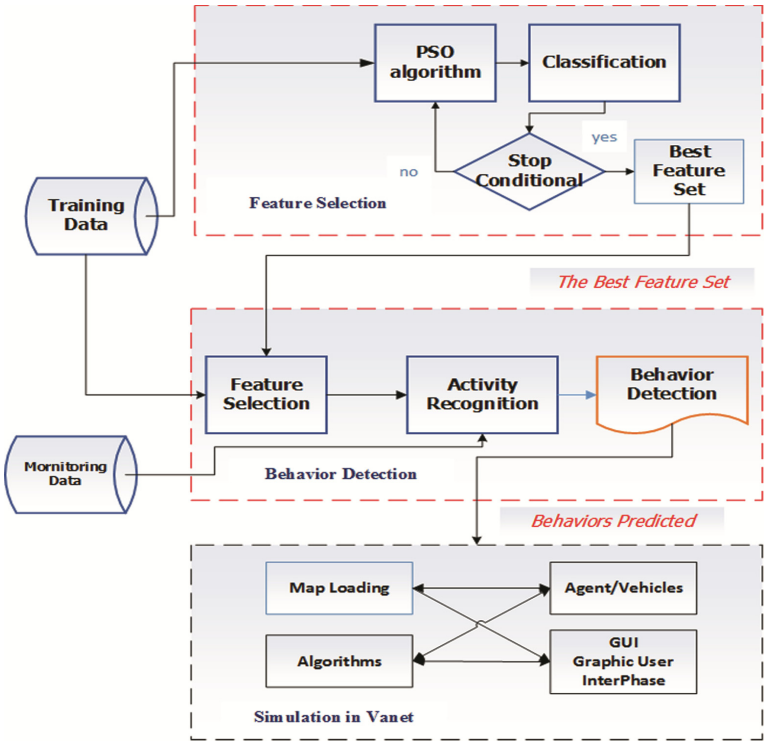


Fig. 1. The proposed framework (ABDS)

2.1 The Feature Selection

The Feature Set

In the phase 1, signal data label obtain from smartphone on their pocket, handbag, or in hands, etc. while they are moving. Hence, the orientation and signal value is frequent changing. Our paper suggest an approach to preparing data for classifier by collected and transform data base one windowing technical and feature extraction. The set of features is constructed by data sequences that were extracted and calculated from time-based, power-based and frequency-based domain data. It is a base for dynamic and suitable selected features system with complex activity. An approach to solve noisy accelerometer data is reoriented. We use accelerometer, gyroscope and magnetometer sensors to transform accelerometer data from the smartphone coordinate system to the Earth coordinate system [12]. Considering that, a (x, y, z) is data point in coordinate system then earth coordinate system of a′(x′, y′, z′) by R matrix and it is computed by function below:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{1}$$

Following, directly analyze with amount raw sensor data is needed a lot of time or memory space. There are any features was extracted by time series technical analysis but, it depends on goal, environment, condition of problems and no method is enough good for all issues. We propose several types of features from instance in three domain above. The windowing technical is usually choice in time series data then it also apply to calculate value of features from root mean square, sample correlation coefficient, cross-correlation, vertical and horizontal accelerometer energy of window signal, energy of M coefficient Fourier values, Signal magnitude area, average Energy of X, Y, Z axis, and the entropy of signal. These total 21 features apply for detection in system and is shown in Table 1.

Table 1. The suggestion feature set for system

The features	Variable name
Root mean square (RMS)	X_{rms}
Correlation coefficient	$CorreCo_{xy}$, $CorreCo_{xz}$, $CorreCo_{yz}$
Cross-correlation	$Cross_{xy}$, $Cross_{xz}$, $Cross_{yz}$
Signal magnitude area	*SMA,*
The accelerometer energy of windows	*Ev, Eh*
Statistical value: mean, variance, standard deviation	*sM, sV, sSD*
The energy of M coefficient Fourier	*EM*
Average energy of X, Y, Z axis	E_x, E_y, E_z
The entropy	H, H_x, H_y, H_z

The Training Model

There are two approach to classify data as offline and online training. The offline method in advance model, which compute on personal computer or server then client to send input and parameters of model. The online method implements computation, recognition on smartphone. The offline method has more propitious conditional and resources. However, it depends on linking and services. Nowadays, hardware and devices quality is more improved. So that, we use online training method on smartphone and assess by some appropriate algorithm such as RF, NB, SVM, KNN, which applied in researches in this field and have shown appropriate accuracy. The WEKA tool has used and integrated in ABDS for classification and recognition. Experiment in this paper indicated that, RF is appropriate and higher accuracy. This algorithm will be applied to optimize feature set.

Optimization Feature Set Using PSO Algorithm

In fact, any approaching chose several kind of feature suitable with data and problem in their field. The selected features is usually via experiment and no method agree to all problem. Hence, we suggest using PSO algorithm to select suitable features for improving prediction accuracy base on wrapper method. The PSO was introduced by Eberhart and Kennedy [13]. In PSO, each potential solution is corresponding to particle and assumption that, $x = [x_1, x_2, \ldots, x_D]$ with $X_i = (X_1, X_2, \ldots, X_D)$ are features at particle i^{th} and the x_k^i is particle position; the v_k^i is particle velocity; the p_k^i is the best individual

particle position; p_k^g is the best swarm position; c_1, c_2 are cognitive and social parameters; r_1, r_2 is random numbers between 0 and 1. The PSO algorithm is express below (Fig. 2):

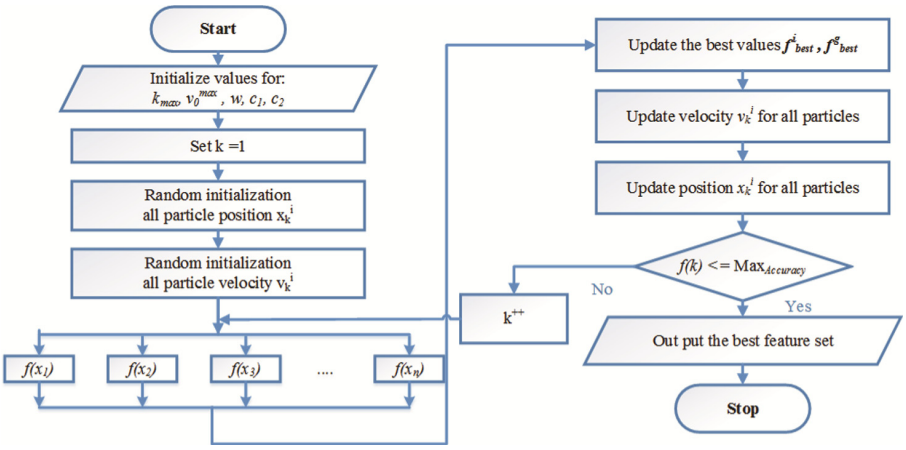


Fig. 2. The feature selection using PSO algorithm

Where

$$x_{k+1}^i = x_k^i + v_{k+1}^i \tag{2}$$

And:

$$v_k^i = v_k^i + c_1 r_1 (p_k^i - x_k^i) + c_2 r_2 (g_k^i - x_k^i) \tag{3}$$

When phase one finished, the best of subset feature is chosen and utilizing to predict activity from online accelerometer signal sensors data.

2.2 Abnormal Behavior Analysis

The definition context safety or not depend on realities issue and opinions. The repeating activities usually reflect abnormal driving behaviors. It is also belonging personality of user habit. Hence, ABDS use k-series activity aim to predict abnormal driving behavior via accelerometer sensors signal. The system is also monitoring and predicting behavior at status of vehicle on real time. When systems has detected current driving activity a_c then combined with $k - 1$ previous activity to instance consist k linear activity as $(a_{c-k-1}, \ldots, a_{c-1}, a_c)$. In fact, abnormal behaviors have realized after one or several complex activity, which are abnormal and repeating. Hence, series activity in instance is basic to predict them. The problems affect to prediction accuracy is value of k and personal habit.

The abnormal behavior training dataset is build up from series activity while they are driving on sleeping, drunk and frequency swing with high velocity. The clustering technical with k-means clustering algorithm is used to set of k linear activity into three class and the label of abnormal behavior instance number i is shown by $S_i(S_{i_1}, S_{i_2}, \ldots, S_{i_k}, l_i)$. It indicated that, abnormal behavior has built up from sequence basic activity. The ANN algorithm is applied by some research in human activity and behavior recognition system. The neurons will be received inputs from other neurons. The value of each input is determined by a weight associated with them. The sum of input weights computed and value output is according to its transfer function. The neurons in a layer do not interconnect with each other, but interconnect with neurons in other layers. Neural networks can have one or more layers between input and output layers. The typical ANN is consisting of input layer, hidden layers and output layer that is expressed in Fig. 3.

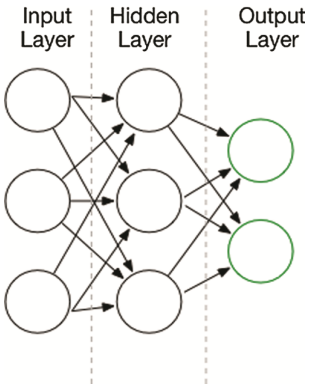


Fig. 3. Artificial Neural Network structure

The series monitoring activity is used to predict abnormal behavior in Fig. 4.

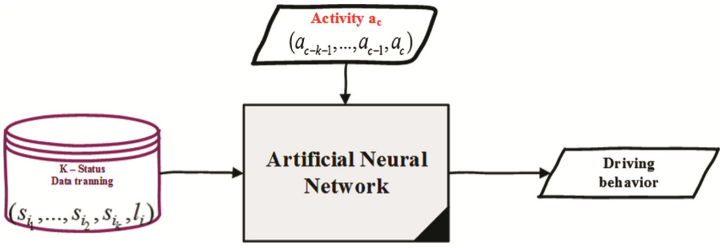


Fig. 4. Abnormal Behavior Detection

2.3 Monitoring Application

The training data set collects by any difference subjects with advanced supposed conditional. However, each user might have different in habits. Therefore, the prediction accuracy may be fall down when the system used by another or new users. Okeyo et al. [15] developed the idea to incrementally update the training data set by utilizing real-time feedbacks from users. As the system provides the activity, behaviors prediction, user can confirm the right of the results. The newly instance data labeled, which correcting from users is then added to the training data set. It is really meaningful with abnormal behavior, which happening and depend on user habit, complexity when user is moving. Specially, characteristic behavior of difference users will be recognize and update. The idea is express in Fig. 1 and some the interfaces of ABDS system is shown in Fig. 5.

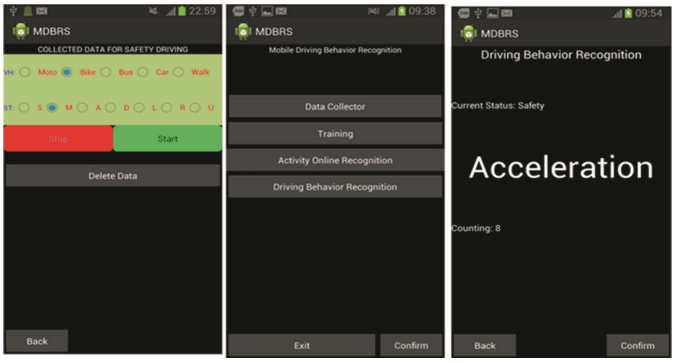


Fig. 5. The interfaces of ABDS

3 Experiment and Results

3.1 Experiment Environment

We implemented ABDS on the Android Operating System from 4.0 to 5.0 platform. The dataset with activity labels collected by 20 subjects when they are driving by walking, bicycle, motorbike, bus and car. They freely carry a Samsung galaxy S4, Quad-core 1.6 GHz Cortex-A15 processor, 2 GB of Ram, 2600 mAh battery, Android 4.2.2 Jelly Bean OS. The sevens activity recognition are {stop, moving, acceleration, deceleration, turn left, turn right and U-turn}. The Weka tool is used for deploying on our framework to predict the vehicle status. The classification was used such as Random Forest, KNN, Naive Bayes, SVM. In each case, the default setting is used by setting parameter of algorithms. We also used 10-fold cross validation for evaluating the accuracy of each classification algorithm.

3.2 The Collected and Processing Data

In this paper experiment, signal data is collected from three types of sensors as acceleration, gyroscope and magnetic sensor signal with 50 Hz frequency. These sensor returns x, y, and z coordinates values at point. The raw data stream is first cut out one seconds at the starting point, and one seconds at the end point, cause these periods time are usually redundant. Then, split into a window by 4 s size and the overlapping time is 50% of window size. We collected at least 200 samples for each activity from subjects. The training data set for behavior recognition also improved by users during their trips. It contained meaningful habit characteristic of users.

We have chosen the walking and motorbike for collecting abnormal driving behavior. With the abnormal driving behavior on training dataset, we use the k-means clustering algorithm with k = 3 base on series activity of users, reflecting to three label of abnormal driving behavior such as sleeping driving, drunk and frequency swing by high velocity. After that, the training dataset with normal and abnormal labeled is used for recognizing driving behavior.

3.3 The Abnormal Behavior Analysis Results

In this experiment, we deploy on about 3500 sample; the system will build up model for subjects to predict activity and abnormal behaviors base on two choosing feature method. The result when ABDS deploy with traditional on 21 features and using PSO to select features, it shows in Fig. 6.

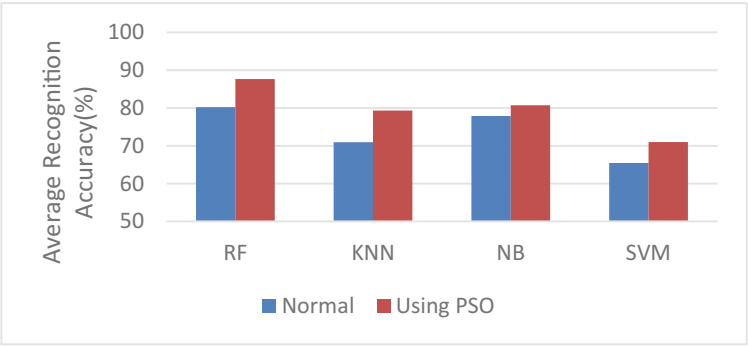


Fig. 6. Activity recognition accuracy of algorithms using PSO and traditional

The Fig. 6 is shown the results between two method select features, the average of recognition accuracy when user uses PSO is always higher. With RF is max as 87.66% and SVM is min with 71%. It indicates that, the RF algorithm is suitable in ABDS system.

The Fig. 7 is shown the average activity recognition accuracy by RF on normal method using 21 features and using PSO algorithm for optimization with the best feature subset, which selected by PSO. It indicated that, the accuracy of the abnormal activity are lower than normal, specially abnormal turn left and turn right activity is lowest.

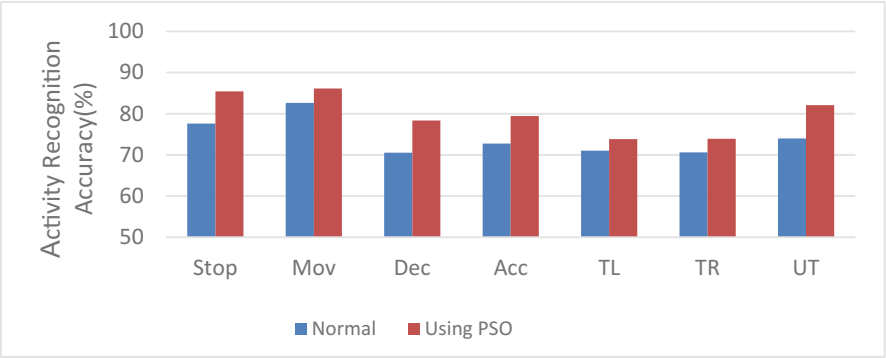


Fig. 7. The activity recognition by Random Forest algorithm

Thereafter, we are collected a behavior training dataset, which contain 300 abnormal behavior instances aim to detect current abnormal behavior base on activity. Through experiments, we chose k = 6 with 6 linear activity are built for each instance. The parameters value of ANN is default setting with behavior detecting accuracy is up to 86.71%.

4 Conclusion and Future Works

In this paper, we proposed a flexible framework to predict current driving behavior base on smartphone sensor, when user moving, dynamic changing position and direction. Besides, our proposed framework also using PSO to select suitable features. Following, ABDS detect vehicle activity and this is basic for recognition driving behavior using ANN. It utilizes and simulates on transportation dartboard using Vanet simulator. In the experiments, ABDS can achieve on average 86.71% accuracy for predicted driving behavior. Furthermore, Random Forest classifier is a promising one for our framework. In the future, we are planning to further improve the current framework to increase accuracy and integrated with any solutions in traffic simulation system.

Acknowledgments. This work has been supported by Vietnam National University, Hanoi (VNU) under project No. QG.17.39.

References

1. Reyes-Ortiz, J.-L., Oneto, L., Sama, A., Parra, X., Anguita, D.: Transition-aware human activity recognition using smartphones. Neurocomputing **171**, 754–767 (2016)
2. Gomes, J.B.R., Krishnaswamy, S., Gaber, M.M., Sousa, P.A., Menasalvas, E.: MARS: a personalised mobile activity recognition system. In: 2012 IEEE 13th International Conference on Mobile Data Management (MDM), pp. 316–319. IEEE, July 2012. https://doi.org/10.1109/MDM.2012.33

3. Kusuma, A., Liu, R., Montgomery, F.: Gap acceptance behavior in motorway weaving sections. In: Proceedings of the Eastern Asia Society for Transportation Studies, vol. 9 (2013)
4. Li, F., Zhang, H., Che, H., Qiu, X.: Dangerous driving behavior detection using smartphone sensors, pp. 1902–1907 (2016)
5. Xu, H., Zhang, L., Zhai, W.: Detection of human movement behavior rules using three-axis acceleration sensor. In: Jin, D., Lin, S. (eds.) Advances in Multimedia, Software Engineering and Computing Vol. 1. Advances in Intelligent and Soft Computing, vol. 128, pp. 647–652. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25989-0_103
6. Jain, A., Kanhangad, V.: Exploring orientation and accelerometer sensor data for personal authentication in smartphones using touchscreen gestures. Pattern Recogn. Lett. **68**, 351–360 (2015)
7. Kalra, N., Bansal, D.: Analyzing driver behavior using smartphone sensors: a survey. Int. J. Electron. Electr. Eng. **7**(7), 697–702 (2014)
8. Ferrer, S., Ruiz, T.: Travel behavior characterization using raw accelerometer data collected from smartphones. Procedia - Soc. Behav. Sci. **160**, 140–149 (2014)
9. Bayat, A., Pomplun, M., Tran, D.A.: A study on human activity recognition using accelerometer data from smartphones. Procedia Comput. Sci. **34**, 450–457 (2014)
10. Yu, J., Chen, Z., Zhu, Y., Chen, Y., Kong, L., Li, M.: Fine-grained abnormal driving behaviors detection and identification with smartphones. IEEE Trans. Mob. Comput. **16**(8), 2198–2212 (2017)
11. Vavouranakis, P., Panagiotakis, S., Mastorakis, G., Mavromoustakis, C.X., Batalla, J.M.: Recognizing driving behaviour using smartphones. In: Batalla, J.M., Mastorakis, G., Mavromoustakis, C.X., Pallis, E. (eds.) Beyond the Internet of Things. IT, pp. 269–299. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50758-3_11
12. Lu, D.-N., Nguyen, T.-T., Ngo, T.-T., Nguyen, T.-H., Nguyen, H.-N.: Mobile online activity recognition system based on smartphone sensors. In: Akagi, M., Nguyen, T.-T., Vu, D.-T., Phung, T.-N., Huynh, V.-N. (eds.) ICTA 2016. AISC, vol. 538, pp. 357–366. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-49073-1_39
13. Eberhart, R., Kennedy, J.: A new optimizer using particle swarm theory. In: 1995 Proceedings of the Sixth International Symposium on Micro Machine and Human Science, MHS 1995, pp. 39–43 (1995)
14. Liu, Z., Wu, M., Zhu, K., Zhang, L.: SenSafe: a smartphone-based traffic safety framework by sensing vehicle and pedestrian behaviors. Mob. Inf. Syst. **2016**, 1–13 (2016)
15. Okeyo, G., Chen, L., Wang, H., Sterritt, R.: Dynamic sensor data segmentation for real-time knowledge-driven activity recognition. Pervasive Mob. Comput. **10**, 155–172 (2014)

An Effective of Data Organizing Method Combines with Naïve Bayes for Vietnamese Document Retrieval

Khanh Linh Bui[1], Thi Ngoc Tu Nguyen[1], Thi Thu Ha Nguyen[1(✉)], and Thanh Tinh Dao[2]

[1] Vietnam Electric Power University, 235 Hoang Quoc Viet, Tuliem, Hanoi, Vietnam
{linhbk,hantt}@epu.edu.vn
[2] Le Quy Don Technical University, 100 Hoang Quoc Viet, Tuliem, Hanoi, Vietnam
tinhdt@mta.edu.vn

Abstract. Data is uploaded to Internet daily that make more and more difficult to mine it. Currently, the available of data mining tools still cannot discover knowledge from data that need semantic with difference dimensions. In this paper we present a method to search the related documents based on clustering that grouped by content. In this, the features are assigned weight by supporting. Experimental results show that the proposed method is really effective, high accuracy and the response results are quickly.

Keywords: Support · Data mining · Text retrieval · Information retrieval Clustering · Document retrieval

1 Introduction

The development of Internet brings an explosive amount of information on the web. Sometimes, it makes users feel quite hard to read and search information that they need. Therefore, data mining is hot and related field as information retrieval, information extraction, data clustering are concerned [1, 2].

Information retrieval is a sub field of data mining that aims to store and allows quick access a large amount of information. The text is often considered as documents, books, articles, etc. However, this is not an easy task, because the booklets in the information systems often have to deal with tens of thousands or tens of millions of documents. So, the search engine can not process more quickly if we don't use any technique to reduce time for processing and enhance accuracy of system [1, 4].

There are several proposed approaches previously mentioned organizing data and feature reduction that have been able to effective search engines [5, 6]. However, it is very difficult to determine feature and how to reduce it. In this paper, we present a method to search effectively by reducing the feature and enriching the semantics of features by using support measure that improve from association rule. It is really better than two – dimensional tf - idf before.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 205–213, 2018.
https://doi.org/10.1007/978-3-319-77818-1_20

The rest of the paper is organized as follows: In Sect. 2, we will introduce some related works. In Sect. 3 is the presentation of our method for data organizing, methodology of Vietnamese document retrieval will be presented in Sect. 4. Experiments and results will show in Sect. 5. And finally, Sect. 6 is a conclusion and future works.

2 Related Works

The earliest studies on the task of information retrieval are described through keywords. It is the simplest approach by matching the words that are entered as a search query and the documents in the data warehouse [3, 10]. To increase the effectiveness of search engines, there are several studies suggested to organizing data task, index documents in the warehouse or ranking data [1]. The other studies also added matching problem that can enhance accuracy between query and data [3]. They also concerned how to select features and reduce it to speed up search engines [5, 6].

The problem of organizing data, the number of studies often uses clustering or classification based on machine learning methods as HAC, SVM, neural network or decision tree. After clustered or classified, documents is organized in clusters with similar kinds of semantic or content [7, 8].

To enhance accuracy of the search engines, some researches focus on relevant feedback. They proved the effectiveness of the search engine when receive feedback from the users [9, 10].

Feature reduction is a solution to speed up the search engine. Some studies showed that, the full features often make system slower. Therefore, to speed up effectively, feature vectors are needed to reduce. However, the selection of useful features and remove unneeded features is a difficult problem [5, 6].

3 Organize Documents in the Warehouse

3.1 Feature Selection

Feature selection is one of the key topics in machine learning and other related fields. Real-life datasets are often characterized by a large number of irrelevant or redundant features that may significantly hamper model accuracy and learning speed if they are not properly excluded. Feature selection involves finding a subset of features to improve prediction accuracy or decrease the size of the structure without significantly decreasing prediction accuracy of the classifier built using only the selected features.

To overcome the disadvantages of large feature vectors we selected by using a word segmentation tool for separating word and selecting only national words. A national word set is define is a set of words that are include verb, noun and adjective.

3.2 Organize Document Based on Clustering

Clustering algorithms group a set of documents into subsets or clusters. The algorithms' goal is to create clusters that are coherent internally, but clearly different from each other.

In other words, documents within a cluster should be as similar as possible; and documents in one cluster should be as dissimilar as possible from documents in other clusters. We use HAC algorithm and the similarity score to cluster documents.

4 The Methodology of Effective Document Retrieval

4.1 Calculating Score of Features Based on Support

In the clustering process (Sect. 3.2), there are n clusters made. It is called *C* and presented as below

$$C = \{C_1, C_2, \ldots, C_n\} \tag{1}$$

In each cluster *C*, we have a set of documents *D*.

$$D = \{d_1 \ldots d_m\} \tag{2}$$

Suppose that, in each cluster *C*, if we consider a document is a transaction, frequency of national word is considered an item, we have a table like this:

After that, we calculate score of term. We use the improving support (in the association rule) to assign value to terms. With each term in Table 1, support of it with each C is calculated as

Table 1. Transactions and item set

TID	Term
d_1	$t_{11}, t_{12},\ldots$
d_2	$t_{21}, t_{22}\ldots$
……	
d_k	$t_{k1}, t_{k2},\ldots$

$$\text{supp}(t_i \rightarrow C_j) = \frac{n(t_i)}{N} \tag{3}$$

In which:

- $n(t_i)$: number of document in cluster C_j that includes t_i
- N_{C_j}: number document in each cluster C_j.

Finally, we built a relationship of national words and topics. In this, each national word has a score to topics. We can set a threshold to adjust amount of national words in each topic. It called feature reduction (Fig. 1).

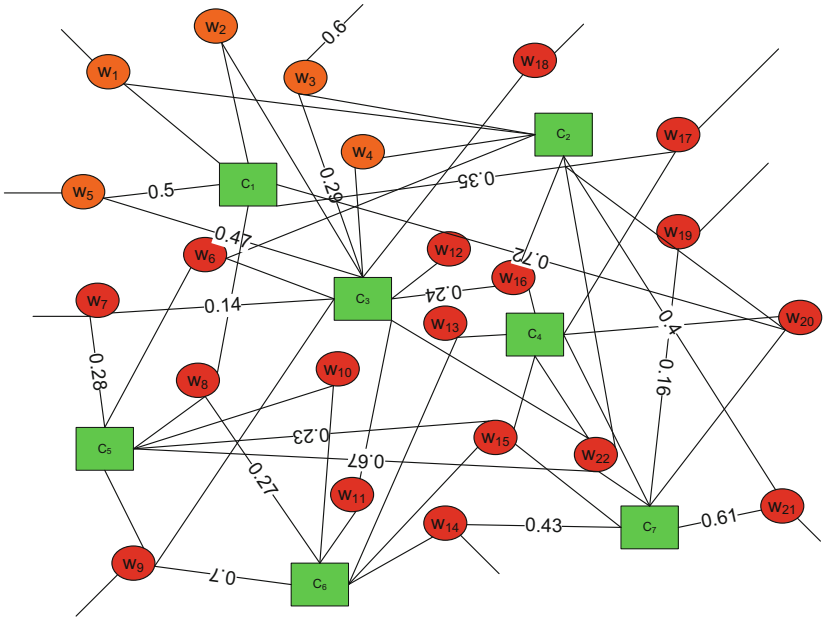


Fig. 1. Relationship between national words and topics

4.2 Calculating Similarity Between Query and Clusters

In the entered query Q, we perform to extract national words (Fig. 2).

$$Q = \{q_1, q_2, \ldots, q_k\} \tag{4}$$

Then, we calculate total of national words in the query Q with each cluster C.

$$\text{total\_supp}(Q_{C_i}) = \sum_{j=1}^{k} \text{supp}(\text{w}_j) \tag{5}$$

In which:

– *Supp(w_j)* is the support of the term w_j with cluster C.

The highest of total support is the cluster that is the most similar with the query (Fig. 3). The algorithm likes below:

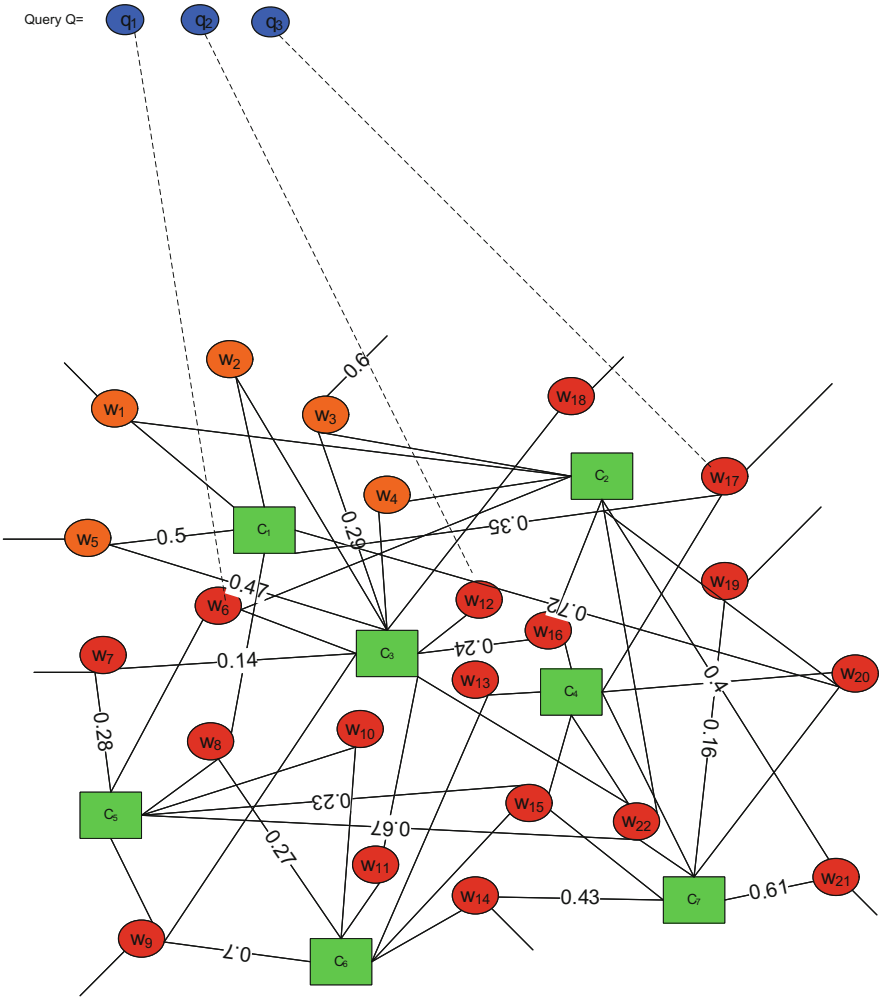


Fig. 2. Similarity between query Q and topics

REBDO Algorithm

Input: query Q, national word dictionary V
Output: d_i
Initialization
$L=Ø;\ M=Ø;\ i=0;$
Begin
1. *Segment real words in query Q*
   For i←1 to ***length(Q)***
   If Q[i] ∈V ***then***
   $L \leftarrow Q[i];$
2. *Calculate sum of supporting in each cluster*
   For each cluster C_k
   $M \leftarrow Total\_Supp(L(i))$
3. *Ranking clusters*
   For i:=1 to ***count(M) do***
   $K \leftarrow M[i];$
4. *Output results*
   Sorting (K)'
   Output d_i ∈ cluster in K;

End;

Fig. 3. REBDO algorithm

4.3 Document Retrieval

In machine learning, Naïve Bayes belongs to classifier methods that use probability with features is independence (conditional independence). Naïve Bayes is applied widely in data classification. Assume that, there are two classes: C = {R, $\overline{R}$). In which, R is set of document that relate with query $\overline{R}$ is collected by unrelated document to query. So that, the information retrieval problem becomes to determine which documents in cluster are related to the query. Similarity of document d_j and query q is denoted by Bayes rule as

$$sim(\overrightarrow{d_j}|q) = \frac{\Pr(d_j|R) \times \Pr(R)}{P(d_j|\overline{R}) \times \Pr(\overline{R})} \tag{6}$$

In which:

– Pr(R): Probability of documents set that related to query.

- $\Pr(\overline{R})$: Probability of documents set that unrelated to query.
- $\Pr(d_j|R)$: Probability d_j with set of documents R that related to query.
- $P(d_j|\overline{R})$: Probability d_j with set of documents $\overline{R}$ that related to query.

Smoothing (6) by logarithm, obtain (7)

$$sim(\vec{d_j}|q) = \log \frac{\Pr(d_j|R)}{P(d_j|\overline{R})} + \log \frac{\Pr(R)}{\Pr(\overline{R})} \tag{7}$$

Assume that, term in query q is independence

$$q = \{t_1, t_2, \ldots, t_k\} \tag{8}$$

Then

$$\Pr(\vec{d_j}, R) = \prod \Pr(t_i|R) \tag{9}$$

$$\Pr(\vec{d_j}, \overline{R}) = \prod \Pr(t_i|\overline{R}) \tag{10}$$

Assume that set of training data is enough larger $R << \overline{R}$

$$sim(\vec{d_j}|q) \approx \log \frac{\prod \Pr(t_i|R)}{\prod \Pr(t_i|\overline{R})} \tag{11}$$

5 Experimental

5.1 Corpus

There is no standard corpus for Vietnamese text summarization now. Therefore, we built corpus by manual. Documents in corpus are downloaded from websites' news as: http://thongtincongnghe.com, http://echip.com, http://vnexpress.net, http://vietnamnet.vn, http://tin247.com. There are over 300 documents in it. Table 2 presented some documents in corpus and number sentences in each document.

Table 2. Corpus

Document	Source	Sentences	File name
Ứng dụng Twitter trong lớp học	thongtincongnghe.com	28	18-10.txt
Hacker "sờ tới" website chính phủ Malaysia	Vietnamnet.vn	15	11-5.txt
Yahoo ra mắt công cụ tìm kiếm app cho Android	Ngoisao.net	12	12-9.txt
TQ phủ nhận điều tra chống độc quyền Microsoft	Tin247.com	21	13-8.txt
Cấu hình tối thiểu để nâng cấp lên Mac OS X Lion	Sohoa.vnexpress	18	16-3.txt
Chọn hệ điều hành của bạn	pcworld.com	69	21-10.txt
Linux ở khắp mọi nơi	Vietbao.vn	71	22-1.txt
Màn hình cảm ứng: Đằng sau những cú chạm	Pcworld	86	25-4.txt
Phanh phui bí mật thế giới ngầm hacker Việt Nam	Echip.com	137	33-4.txt
Người dùng di động quan tâm giá cả hơn sáng tạo công nghệ	baomoi.com	39	33-7.txt

All file downloaded from website will be saved in corpus by *.txt and preprocessed.

5.2 Word Segmentation

We build a dictionary of national words and used VnTagger tool that downloaded from vlsp website to segment words. VnTagger is published on internet via address: http://vlsp.hpda.vn:8080/demo/?page=home [11].

5.3 Evaluation

At the present, Vietnamese does not have any standard assessment method, we use recall measure for evaluation. Recall is the fraction of the documents that are relevant to the query that are successfully retrieved (Table 3)

Table 3. Some topics for retrieving

Topics	Number of relevant documents	Recall
Business	52	0.573333
Education	54	0.68
Football	46	0.453
Travel	42	0.514
Information	78	0.526
Technical	36	0.511

$$recall = \frac{|\{relevantdocument\} \cap \{retrieveddocuments\}|}{|\{relevantdocuments\}|} \quad (12)$$

6 Conclusion

The task of information retrieval based on content been concerned by researchers and scholars when the current systems still search by keyword or phrase. In this paper, we propose an effective method for information retrieval based on content and added objectives are fast and accurate. With the results of experimental show that, our method really effectively to reduce complex computing and time for processing when performing with Vietnamese text.

References

1. Bhattacharyya, P., Datta, J.: Ranking in information retrieval, 16 April 2010
2. Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., Fraternali, P., Quarteroni, S.: Web Information Retrieval. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39314-3. ISBN 978-3-642-39314-3
3. Buscher, G., Dengel, A., van Elst, L.: Query expansion using gaze-based feedback on the subdocument level. In: Proceedings of the 31th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM Press, Singapore, pp. 387–394 (2008)
4. Zadeh, M.V.: Improving the performance of text Information Retrieval (IR) System, Ph.D thesis. Porto University (2012)
5. Lau, J.H., Newman, D., Karimi, S., Baldwin, T.: Best topic word selection for topic labelling, Coling 2010, Posters, pp. 605–613 (2010)
6. Moens, M.-F., Vulić, I.: Monolingual and cross-lingual probabilistic topic models and their applications in information retrieval. In: Serdyukov, P., Braslavski, P., Kuznetsov, S.O., Kamps, J., Rüger, S., Agichtein, E., Segalovich, I., Yilmaz, E. (eds.) ECIR 2013. LNCS, vol. 7814, pp. 874–877. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36973-5_106
7. Park, H., et al.: Agglomerative hierarchical clustering for information retrieval using latent semantic index. In: 2015 IEEE International Conference Smart City/Socialcom/Sustaincom (SmartCity), 19–21 December 2015
8. Kalyanasundaram, C., Ahire, S., Jain, G., Jain, S.: Text clustering for information retrieval system using supplementary information. Int. J. Comput. Sci. Inf. Technol. **6**(2), 1613–1615 (2015)
9. Kuhn, L., Eickhoff, C.: Implicit negative feedback in clinical information retrieval. In: Medical Information Retrieval Workshop (MedIR), Pisa, Italy, 21 July 2016
10. Rocchio, J.J.: Relevance feedback in information retrieval (1971)
11. http://vlsp.hpda.vn:8080/demo/?page=home

An Effective Time Varying Delay Estimator Applied to Surface Electromyographic Signals

Gia Thien Luu[1(✉)], Abdelbassit Boualem[2], Philippe Ravier[2], and Olivier Buttelli[2]

[1] Department of Physics, Posts and Telecommunications Institute of Technology, Ho Chi Minh City, Vietnam
lgthien@ptithcm.edu.vn
[2] University of Orleans, Orleans, France
{philippe.ravier,olivier.buttelli}@univ-orleans.fr

Abstract. Muscle Fiber Conduction Velocity (MFCV) can be calculated from the time delay between the surface electromyographic (sEMG) signals recorded by electrodes aligned with the fiber direction. In order to take into account the non-stationarity during the dynamic contraction (the most daily life situation) of the data, the developed methods have to consider that the MFCV changes over the time, which induces time varying delays and the data is non-stationary (change of Power Spectral Density (PSD)). In the present paper, the problem of time varying delay (TVD) estimation is considered using a parametric method. First, the polynomial model of TVD has been proposed. Then, the TVD model parameters are estimated by using a maximum likelihood estimation (MLE) strategy solved by a stochastic optimization technique, called simulated annealing (SA). The Monte-Carlo simulation results show that the estimation of both the model parameters and the TVD function is unbiased and that the variance obtained is close to the Crammer-Rao Lower Bound (CRLB). We also compared the performance of the proposed method with non-parametric approaches. The results indicate that the proposed method outperform the non-parametric one.

Keywords: Electromyography · Time-varying delay
Muscle Fiber Conduction Velocity · Non-stationarity
Maximum likelihood estimation · Simulated annealing (SA) · CRLB

1 Introduction

Muscle Fibers Conduction Velocity (MFCV) is an interesting physiological indicator, *e.g.* for monitoring neuro-muscular degenerative diseases [1] and also for the assessment of pain in the case of fibromyalgia [7]. This indicator was also widely used for fundamental studies on motor control, *i.e* Motor Unit (MU)

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 214–225, 2018.
https://doi.org/10.1007/978-3-319-77818-1_21

recruitment modality based on force levels; study of fatigue [13,19] whose applications include both the medical field that the physiology of the exercise and ergonomics. MFCV can be estimated from intramuscular or surface electromyography recordings [14]. In this work, we are only interested in surface EMG signals (sEMG).

In the case of dynamics contraction (the most daily situation), Farina et *al.* [6] adapted the maximum likelihood estimator to short analysis intervals. However the resulting approach does not provide an instantaneous delay estimation [17]. In [9,10,17], the time frequency and time scale approaches were developed for the problem of time varying estimation for each instant. In such cases, the TVD model choice is not critical since the investigated methods are independent of TVD model. However, the obtained performance suffers from high noise levels and the performance of these methods are depends on the PSD shapes of the sEMG signal.

Optimal TVD estimators can be derived with Maximum-Likelihood estimation (MLE) method. However, in such case, we have to estimate N parameters, the time of calculation is too much. In [12], we proposed a polynomial model for the TVD and adapted the maximum likelihood estimation for the short time to estimate the model parameters. A deterministic optimization was used into the sliding window to solve the optimization problem. However, this technique does not guarantee the global optimization. To solve this problem, in present paper, we proposed an stochastic optimization called stimulated annealing technique combined with the MLE method.

The paper is organized as follows. In Sect. 2, the models of signals and TVD will be defined. In Sect. 3, the MLE method two channels will be derived. Section 4 presents the simulation results with synthetic sEMG compared with the CRLB which was derived in [11]. In Sect. 5, we conclude the paper.

2 Model of sEMG Synthetics Signals

In this section, we first present an analytical model for two-channel sEMG acquired signals, and then a generating model of synthetics EMG signals. This model is helpful for statistical performance studies.

2.1 Signal Model

Considering the sEMG signal s(n) propagating between channel 1 and channel 2, a simple analytical model of two observed signals $x_1(n)$ and $x_2(n)$ in a discrete time domain, without shape differences, can be given respectively by

$$\begin{aligned} x_1(n) &= s(n) + w_1(n) \\ x_2(n) &= s(n - \theta(n)) + w_2(n) \end{aligned} \tag{1}$$

where $\theta(n)$ is the propagation delay between the two signals, $w_1(n)$ and $w_2(n)$ are assumed to be independent, white, zero mean, additive Gaussian noises, of equal variance σ^2.

Once θ(n) is estimated, the MFCV can be simply deduced by MFCV(n) = Δe/θ(n), where Δe stands for the inter-electrode distance, which is taken as 10 mm in the following. The digitization step is processed at the sampling frequency F_s = 1024 Hz. MFCV can be calculated from $\theta(n)$ through

$$MFCV(n) = \frac{F_s.\Delta e}{\theta(n)} \tag{2}$$

where Fs is the sampling frequency and e is the inter-electrode distance.

Next, we describe in detail the way for generating synthetic sEMG signals with predefined TVD functions.

2.2 Delayed Signal Generation

The signals are synthetic ones and are generated according to the following analytic Power Spectral Density (PSD) shape proposed by Shwedyk *et al.* in [18] and written in the following equation as

$$PSD(f) = \frac{k f_h^4 f^2}{\left(f^2 + f_l^2\right).\left(f^2 + f_l^2\right)^2} \tag{3}$$

An example of sEMG PSD shape is given in [5], where the low and high frequency parameters are fixed as f_l = 60 Hz and f_h = 120 Hz, respectively. The parameter k is a normalization factor. The first channel is generated by linear filtering of white Gaussian noise with the impulse response corresponding to this PSD (*i.e.* the inverse Fourier transform of the square root of the previous PSD shape. Once the first channel is generated, its delayed version is created, thanks to the sinc-interpolator [3]:

$$s(n - \theta(n)) = \sum_{i=-p}^{p} \mathrm{sinc}(i - \theta(n)) s(n - i) \tag{4}$$

The parameter p is the filter length and is fixed by p = 40. Finally, both channels are distorted by adding White Gaussian noise at a given signal-to-noise ratio (SNR) level.

3 Proposed Methods

3.1 Maximum Likelihood Estimation

This method was derived in [12]. The estimated TVD can be defined by Eq. 5. Maximize the log-likelihood function is equivalent to minimizing the following expression:

$$\hat{\theta} = \arg\min_{\theta} e_t^2(\theta) \tag{5}$$

where

$$e_t^2(\theta) = \sum_{n=1}^{N} \left(x_2\left(n - \theta(n)\right) - x_1(n)\right)^2 \tag{6}$$

3.2 Polynomial Model

The problem of estimating (n) is the same as estimating the N-dimensional vector $\theta = [\theta(1)\theta(2)\ldots.\theta(N)]$. In the case of TVD polynomial models, this problem, as expressed in Eq. 7, reduces to the estimation of a p + 1-dimensional vector $\Theta = [\theta_0\theta_1\ldots.\theta_p]$.

Thank to the Weierstrass theorem, the TVD may be decomposed up to order p on the canonical polynomial basis as:

$$\theta(n) = \mathrm{F_s} \sum_{k=0}^{p} \theta_k.n^k \tag{7}$$

where F_s is the sampling frequency.

The TVD is thus defined by a p + 1 dimensional vector with parameters $\Theta = [\theta_0\theta_1\ldots.\theta_p]$. In this work, the stochastic optimization technique, called simulated annealing was used to solve the optimization. We detail below this optimization technique.

3.3 Stochastic Optimization

The technique known as simulated annealing is motivated by an analogy to annealing in solids. The idea comes from a paper published by Metropolis *et al.* 1953 [15]. The algorithm in this paper simulates the cooling of the material in a heat bath. This is a process known as annealing. If a solid is heated to melting point and then cooling it, the structural properties obtained for the solid will depend on the cooling rate. If the liquid is cooled slowly enough, large crystals will form. However, if the liquid is rapidly cooled, the crystals contain imperfections. The cooling rate is a critical parameter in the process of crystallization. Metropolis algorithm proposes to simulate the material as a particle system and simulates the cooling process by progressively lowering the temperature of the system until it converges to a stable state frozen. The cooling rate is a critical parameter in the process of crystallization. In 1983, Kirkpatrick *et al.* [8] took up the idea of the Metropolis algorithm and applied to optimization problems. The idea is to use simulated annealing to search for feasible solutions and converge to an optimal solution. A Markov process is used to sample the "objective" function $l\left(\Theta^{(k)}\right)$, seeking a new solution generated in a neighborhood of the current solution, with the objective of minimizing this function. The exploration of the solution space is controlled by a Boltzmann distribution parameter T [4] which describes the behavior of a system "thermodynamic equilibrium" at a certain temperature T. The simulated annealing method, iterative, based on the rule of acceptance Metropolis-Hastings of accepting or refusing a vector solution especially in a function of temperature T. A significant temperature allows to temporarily accept the solutions that move away from the minimum. This allows us to explore more broadly the solution space and thus out of a local minimum. The equilibrium is then achieved with slowly decrease in temperature. For a temperature T and iteration, the algorithm is as follows:

1. Generate vector $\Delta l = l(\Theta^{(k)}) - l(\Theta^{(k-1)})$
2. Apply the Metropolis acceptance rule that involves:
   - If $\Delta l \leq 0$: accept $\Theta^{(k)}$, increment k then iterating step 1
   - If $\Delta l > 0$: accept $\Theta^{(k)}$ with probability $P = e^{\left(-\frac{\Delta l}{T}\right)}$:
     Randomize a random number R between 0 and 1 according to a uniform law.
     if $R \leq P$: accept $\Theta^{(k)}$; increment k and go to step 1.
     If $R > P$: refuse $\Theta^{(k)}$ and go to step 1.

Simulated annealing parameters. The implementation of the simulated annealing procedure requires the setting of several parameters, which play a decisive role in its efficiency.

1. Research neighborhood
   The search for a new solution $\Theta^{(k)}$ is done in the neighborhood of the current solution $\Theta^{(k-1)}$. We have used for the generation of $\Theta^{(k)}$ a normal distribution centered on $\Theta^{(k-1)}$.
   $$\Theta^{(k)} = \Theta^{(k-1)} + \delta \tag{8}$$
   where δ Is an agitation vector generated by a normal distribution $\mathcal{N}(0, \sigma^2)$. In order to unify the standard deviation σ for all the parameters of the vector Θ, we propose to normalize the orthogonal basis of Legendre polynomial. It makes it possible to produce variations of the parameters θ_i of the same order. For a large exploitation of the search space, we take an initial σ of the order of 1/4 of the range of variation. σ decreases with decreasing temperature to obtain a better accuracy of the results.
2. Initial temperature
   It must be high enough so that most degradation are authorized at the beginning of the procedure to allow the location of the region of the global minimum.
   An study [2,16] has shown that an initial value of the temperature of the order of that of the initial "objective" function leads to good results. Thus, the proposed initialization is defined by:
   $$T_0 = \frac{l\left(\Theta^{(0)}\right)}{A} \tag{9}$$
   where $l\left(\Theta^0\right)$ Represents the value of the "objective" function for the initial solution Θ^0 and A is a parameter to be fixed.
3. Diagram of temperature decrease
   It represents the number of solutions tested before applying the diagram of temperature decrease. It is assumed to be constant for all temperature levels. If this length is too small, the exploration of the search space may be too partial, whereas a too large value may have the effect of slowing down or even blocking the search.
   $$T_k = T_{k-1}.C \tag{10}$$
   with $0 < C < 1 (C = 0.95)$.

4. Length of temperature bearing
   The temperature is lowered slowly by marking bearings. The temperature change of the bearing T_k to $T_{(k-1)}$ is carried out according to a given decay pattern. The most commonly used schemes are the arithmetic, geometric, logarithmic or exponential laws.
5. Criteria for stopping the program
   To stop the program, several criteria can be used, separately or often together:
   The temperature becomes lower than a given value.
   The function "objective" ceases to evolve during the decrease of the temperature.
   The non-evolution of the current solution on D consecutive steps.

Setting the parameters of simulated annealing. The setting of the simulated annealing algorithm is done empirically. Several tests are made to arrive at the right parameters which give the best results. These parameters (described in Sect. 3.3) are:

1. The initial search point, chosen equal to $\Theta^{(0)} = [1, 1, \ldots, 1]$ because the "objective" function has a high value for this point. The stirring vector δ of law $\mathcal{N}(0, \sigma^2)$ is initialized with $\sigma = 1$ (depending on the problem data, F_s and distance IE). The standard deviation decreases with the temperature according to: $\sigma_k = 0.98$ (k is the index of the temperature bearing).
2. The temperature parameters: the initial temperature is determined by Eq. 9 with the fixed parameter $A = 0.75$. The diagram of temperature decrease used is a geometric law described in Eq. 10, with C is a fixed parameter $C = 0.95$, because it allows to have a rapid decrease at the beginning of the optimization and slow close to the convergence.
3. The length of the bearing temperature, set at 100 iterations.
4. A counter to stop the program, set to zero at the beginning. The search stops when the counter reaches a certain threshold ($D = 10$). At the end of a stage, the counter is incremented if the current solution does not evolve and it is reset if the quality of the best solution has evolved during the bearing.

The simulated annealing method has the advantage of being flexible with respect to evolution of the problem and easy to implement. It has produced excellent results for a large number of problems, most often large. On the other hand, this algorithm has the disadvantage of having a high number of parameters (initial temperature, decrease in temperature, duration of temperature steps, etc.) whose settings often remain fairly empirical.

4 Results and Discussions

In this section, we detailed statistical study of the performances of the proposed methods. First, the statistical tool are detailed. Then, the simulation strategy is described. The results of the proposed methods compared with the CRLB are also detailed.

4.1 Statistical Tool

To evaluate the performance of the estimators, the useful tool are the normalized bias and the root mean square error (RMSE). The variance of these estimators are also useful to compare with the CRLB.

The normalized bias is expressed by:

$$Bias(\widehat{\theta}(n)) = \frac{\left|E\left[\widehat{\theta}(n)\right] - \theta(n)\right|}{\theta(n)} \tag{11}$$

And the instantaneous RMSE is defined by:

$$RMSE(\widehat{\theta}(n)) = \sqrt{E(\left(\widehat{\theta}(n) - \theta(n)\right)^2)} \tag{12}$$

where $\widehat{\theta}(n)$ is the estimator of $\theta(n)$. The symbol $E[.]$ denote the expectation operator.

4.2 Monte-Carlo Simulation

A Monte-Carlo simulation with 150 independent runs was performed for each signal to noise ratio (SNR) value in order to study the noise impact of these estimators. In this work, two synthetic sEMG signals have the same value of SNR = 10, 20 dB respectively. Duration of the signals is 1 s, the sampling frequency was set as $F_s = 1024$ Hz and the inter-electrode distance is $\Delta e = 10$ mm.

4.3 Simulation Results

Inverse sinusoidal model. This model has been previously proposed in [9]. It takes into account reasonable physiological variations of MFCV that may be encountered during dynamical exercise situations. In particular, the minimum and maximum MFCV values are $2\,\mathrm{m.s^{-1}}$ and $8\,\mathrm{m.s^{-1}}$ respectively. The maximum acceleration value is $2.5\,\mathrm{m.s^{-2}}$. One period of the sine wave is considered corresponding to 1 s observation duration or to equivalently 1024 data samples.

$$\theta(\mathrm{n}) = \mathrm{F_s}\frac{10.10^{-3}}{5 + 3\sin\left(0.2\mathrm{n}2\pi/\mathrm{F_s}\right)}. \tag{13}$$

Applying the model 13 on the delay $\theta(n)$. The approximations $\theta_a(n)$ of $\theta(n)$ for different orders are shown in Fig. 1. The resulting parameters up to order 9 are: $\Theta = [4.18;\ 1.21;\ -0.09;\ -1.01;\ 0.33;\ 0.15;\ 0.24;\ 0.07;\ -0.05;\ -0.06]$.

We propose to estimate the delay with an order $P = 7$. This choice is justified by a compromise between an acceptable approximation error and a lowest possible order. An mean square error (MSE), between the theoretical delay $\theta(n)$ and $\theta_a(n)$ was set at 5.10^{-2}. The estimation results are shown in Fig. 2 (150 simulations of Monte Carlo).

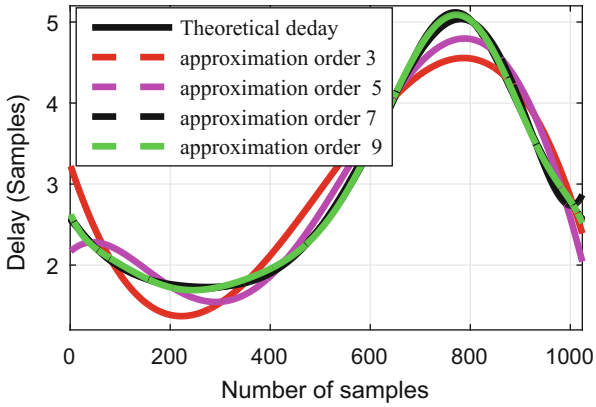


Fig. 1. $\theta(n)$, and and its approximations $\theta_a(n)$ by the model 7 for P = 3, 5, 7 and 9.

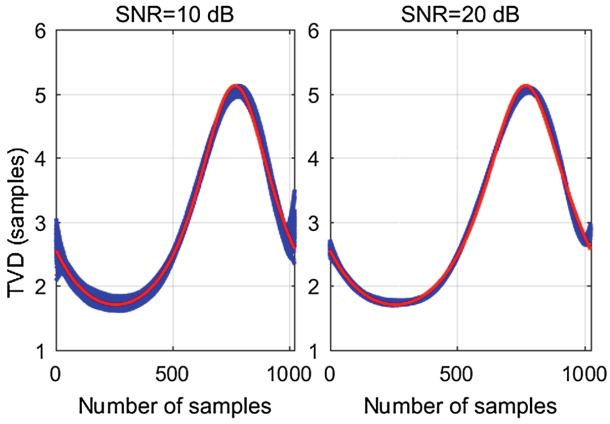


Fig. 2. Delay theoretical (red) $\theta(n)$ and its estimate (blue) with P = 7, (a) SNR = 10 dB, (b) $SNR = 20$ dB. (Results of 150 Monte Carlo simulation, N = 1024 samples, $F_s = 1024$ Hz, $\Delta e = 10$ mm) (Color figure online)

Indeed, the simulated annealing algorithm estimates the approximation of the delay $\theta_a(n)$ of $\theta(n)$. There are two types of errors: estimation errors (Fig. 3) and errors due to the applied model (Fig. 4).

The bias shown in Fig. 5 represents the total error committed (including the modeling error). This bias is low (<5%). For $SNR = 20$ dB, this bias is independent of the SNR because the error due to the estimate is negligible compared to the error due to the model. Increasing the P order reduces modeling errors, but the difficulty of the estimation problem increases.

The MLE method using the model (7) is compared with a method that exploits the information of the phase of the time-frequency plane (phase coherency).

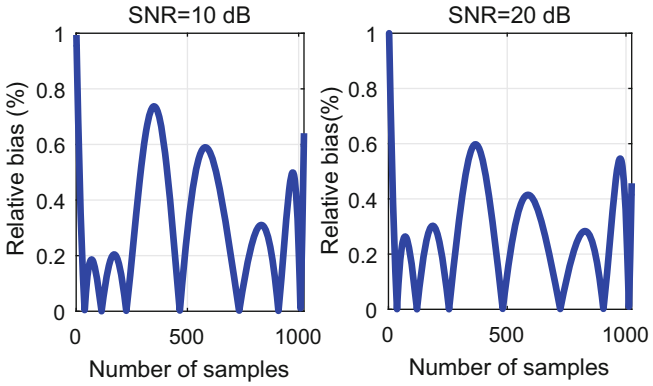


Fig. 3. Relative bias (relative to $\theta_a(n)$): $SNR = 10\,\text{dB}$, (b) $SNR = 20\,\text{dB}$

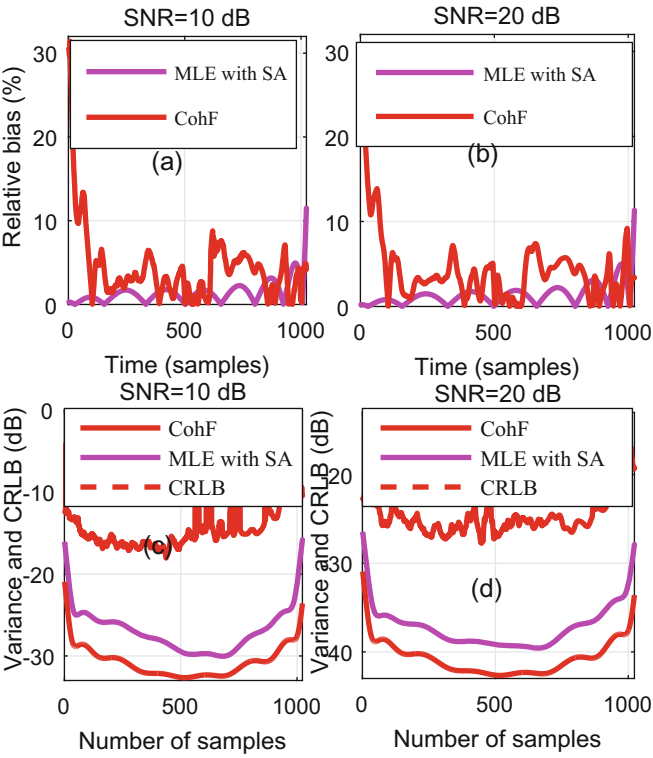


Fig. 4. Statistical results for two estimation methods: MLE (with simulated annealing and an order 7) and phase coherency (CohF) [17] (time-frequency plane calculated on windows of 256 samples (0.25 s), number of fft = 1024. (a), (b) Relative bias (relative to the theoretical delay ($\theta(n)$); (c), (d) variance and CRLB, $F_s = 1024\,\text{Hz}$, $\Delta e = 10\,\text{mm}$.

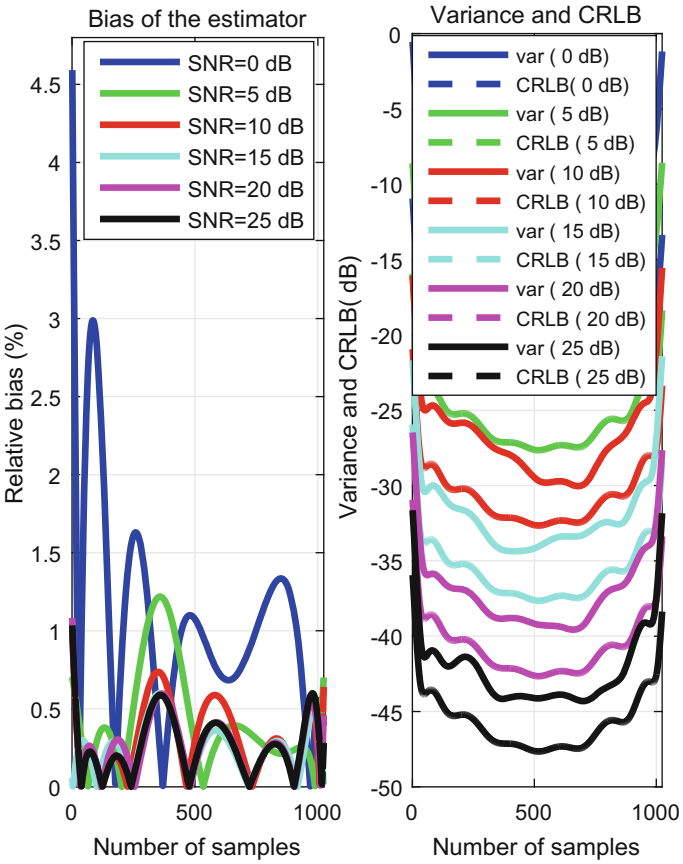


Fig. 5. Statistical results of the $\widehat{\theta}(n)$: (a) relative bias (relative to $\theta_a(n)$ approximation of order 7 of $\theta(n)$, (b) variance (continuous curve) compared with CRLB (discontinuous curve) and each color corresponds to an SNR, 150 runs of Monte-Carlo simulation, N = 1024 samples, $F_s = 1024\,\text{Hz}$, $\Delta e = 10\,\text{mm}$. (Color figure online)

For the method MLE optimized by simulated annealing, the estimator is weakly biased (a very low relative bias <1%). Moreover, the variance of the estimator is very close to the CRLB (a difference of 3 dB). In general, the estimator can be said to be effective.

The MLE method yielded good results with respect to the Fourier Coherency (CohF) method. For CohF method, the estimator is biased. Therefore, the variance of the estimator is very far from the CRLB (a difference of 20 dB). This method is very sensitive to noise (limited at $SNR = 10\,\text{dB}$).

5 Conclusions

In this paper, we have investigate the TVD estimator using the parametric approach. First, the polynomial model for TVD have been proposed. Second, MLE

estimation of the parameter model have been performed using the optimization technique called simulated annealing. The simulation results shown that the MLE method combined with simulated annealing technique is an unbiased estimator. It has a variance 3 dB higher than the CRLB. SA technique also outperformed compared with the CohF method [17]. To increase the performance of estimation, we have to increase the polynomial order of TVD model, but the execution time of also increase. In the future, we will compromise between the execution time and the order of the polynomial model. We will also investigated an other optimization technique in order to improve the performance of the estimator. The PSD shapes of sEMG signals will be also considered in order to take into account the fatigue effect.

Acknowledgment. The authors would like to thank the Posts and Telecommunications Institute of Technology, Ho Chi Minh City Branch for funding this research in project 03-*HV*-2017-*RD*-*CB*2.

References

1. Allen, D.C., Arunachalam, R., Mills, K.R.: Critical illness myopathy: further evidence from muscle-fiber excitability studies of an acquired channelopathy. Muscle Nerve **37**(1), 14–22 (2008)
2. Athier, G., Floquet, P., Pibouleau, L., Domenech, S.: Synthesis of heat-exchanger network by simulated annealing and NLP procedures. AIChE J. **43**(11), 3007–3020 (1997)
3. Chan, Y., Riley, J., Plant, J.: Modeling of time delay and its application to estimation of nonstationary delays. IEEE Trans. Acoust. Speech Signal Process. **29**(3), 577–581 (1981)
4. Dréo, J., Siarry, P.: Métaheuristiques d'optimisation vues sous l'angle de l'échantillonnage de distribution. J. Européen des Systèmes Automatisés **42**(1), 9 (2008)
5. Farina, D., Merletti, R.: Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions. J. Electromyogr. Kinesiol. **10**(5), 337–349 (2000)
6. Farina, D., Pozzo, M., Merlo, E., Bottin, A., Merletti, R.: Assessment of average muscle fiber conduction velocity from surface EMG signals during fatiguing dynamic contractions. IEEE Trans. Biomed. Eng. **51**(8), 1383–1393 (2004)
7. Gerdle, B., Ostlund, N., Grnlund, C., Roeleveld, K., Karlsson, J.S.: Firing rate and conduction velocity of single motor units in the trapezius muscle in fibromyalgia patients and healthy controls. J. Electromyogr. Kinesiol. **18**(5), 707–716 (2008)
8. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**(4598), 671–680 (1983)
9. Leclerc, F., Ravier, P., Buttelli, O., Jouanin, J.C.: Comparison of three time-varying delay estimators with application to electromyography. In: 2007 15th European Signal Processing Conference, pp. 2499–2503, September 2007
10. Leclerc, F., Ravier, P., Farina, D., Jouanin, J.C., Buttelli, O.: Time-varying delay estimation with application to electromyography. In: 2008 16th European Signal Processing Conference, pp. 1–5, August 2008

11. Luu, G.T.: Contributions à lestimation de la vitesse de conduction des signaux électromyographiques. Ph.D. dissertation, Orleans University (2013)
12. Luu, G.-T., Ravier, P., Buttelli, O.: Comparison of maximum likelihood and time frequency approaches for time varying delay estimation in the case of electromyography signals. In: Biomedical Engineering International Conference (BMEiCON), pp. 1–5 (2012)
13. Merletti, R., Knaflitz, M., De Luca, C.J.: Myoelectric manifestations of fatigue in voluntary and electrically elicited contractions. J. Appl. Physiol. **69**(5), 1810–1820 (1990). http://jap.physiology.org/content/69/5/1810
14. Merletti, R., Conte, L.R.L.: Surface EMG signal processing during isometric contractions. J. Electromyogr. Kinesiol. **7**(4), 241–250 (1997)
15. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**(6), 1087–1092 (1953)
16. Pibouleau, L., Domenech, S., Davin, A., Azzaro-Pantel, C.: Expérimentations numériques sur les variantes et paramètres de la méthode du recuit simulé. Chem. Eng. J. **105**(3), 117–130 (2005)
17. Ravier, P., Farina, D., Buttelli, O.: Time-varying delay estimators for measuring muscle fiber conduction velocity from the surface electromyogram. Biomed. Signal Process. Control **22**, 126–134 (2015)
18. Shwedyk, E., Balasubramanian, R., Scott, R.N.: A nonstationary model for the electromyogram. IEEE Trans. Biomed. Eng. **BME–24**(5), 417–424 (1977)
19. Yaar, I., Niles, L.: Muscle fiber conduction velocity and mean power spectrum frequency in neuromuscular disorders and in fatigue. Muscle Nerve **15**(7), 780–787 (1992). https://doi.org/10.1002/mus.880150706

The Optimal Solution of Communication Resource Allocation in Distributed System Integrated on Cloud Computing

Hung Vi Dang[1(✉)], Tien Sy Nguyen[2], Van Son Le[1], and Xuan Huy Nguyen[3]

[1] Department of Information Technology, Danang University of Education, Da Nang, Vietnam
dhungvi@ued.udn.vn
[2] Information and Communications Publishing House, Hanoi, Vietnam
[3] Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

Abstract. Cloud computing has been growing rapidly in the world over the past decade. The Studies and development of this system has met the demand of large number of users in the world. In order to share shared resources, most applications are deployed in the cloud under the control of distributed systems. The distributed system deployment on the SaaS layer responds to the maximum user access through coordination between servers. This coordinate control messages moving across servers to ensure conherence, transparency for user. However, disadvantage of coordination is that communication between servers in the cloud occupies large bandwidth; not to mention overlap of information at destination by multicast transmission. In this paper, we present optimal solution of communication resource allocation (CRA) in distributed system integrated on cloud computing based on network coding technique to ensure maximum throughput and avoid overlap of information at destination.

Keywords: Communication resource allocation · Distributed system
Cloud computing · Multicast · Network coding

1 Introduction

Cloud computing is the large system that resources allocation to users through services. The flexibility of cloud computing is a resources allocator on demand, which facilitates the maximum utilization of system resources available. For cloud computing, physical resources are used as a virtual computer combined. This provides an environment in which applications perform independently without regard to any physical resources.

Cloud computing uses dynamic resources, hence, the issue of resource allocation for this system is more complex than resource allocation in distributed systems. Virtualization that will allow for cloud computing. Virtual machine systems (VMS) combine physical resources to resource allocation for virtual machines. Characteristic of VMS is make the most of the available resources in the system and serve large number of users. In addition, VMS can deploy one or more distributed systems on it to shared resource allocation to users.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 226–236, 2018.
https://doi.org/10.1007/978-3-319-77818-1_22

Shared resources in the distributed system are concerned by the limited amount of resources, while the number of requests is greater than the number of requests leading to concurrency and deadlock in provisioning. The coordination between the servers in the distributed system is based on message passing mechanisms [10]. Through this mechanism, messages are controlled, routed based on structure to ensure data consistency. Studies about communications in distributed systems refer to multicast transmission mechanisms that allow a group of servers to exchange messages with each other to control shared resource allocation [4, 6, 10]. The continuous exchange of messages between servers leads to appropriate of large communication resources, thus, there should be an optimal solution in CRA.

Studies in [8, 9] point out that the disadvantage of multicast transmission is that packets that arrive at destination set can be overlapped, leading to waste of communication resources. One of the research directions to solve this solution is network coding. The goal research of network coding is to establish multicast connections and avoid overlap information at destination set. The solution for network coding technique are processing on multicast trees to achieve maximum throughput at destination set through XOR operation. This solution is built and implemented in the communication resource allocator.

Building a communication resource allocator is one of the essential conditions for control, monitoring, and resource allocation. In this paper, we present the optimal solution to CRA in distributed systems integrated on cloud computing based on network coding techniques. The main contents of the article are summarized as follows:

- Introduction a general solution CRA in distributed systems integrated in cloud computing.
- This system described the optimization algorithm CRA in VMS.
- Simulates the solution optimization of communication resources based on the network coding technology versus multicast transmission solution.

The sections of the paper are organized as follows: Sect. 2 presents the related studies, we explore, evaluate and analyze the current model to draw the advantages and disadvantages in CRA. Section 3 presents the optimal solution model and algorithms for resource allocation. Section 4 provides simulation results for evaluating a solution model from which to draw conclusions as outlined in Sect. 5.

2 Relate Works

The basic model of resource allocation in cloud computing is shown in Fig. 1, based on the specific structure of each type to classify as cloud types: private, hybrid, public and community. Services of cloud computing resource allocation to users for a variety of purposes based on demand.

In the process of resource allocation, CRA system can achieve optimum bandwidth consumption by observing real-time bandwidth consumption and making decisions about reallocation.

Studies in [13] presents an overview of resource allocation strategies in cloud computing. Discussion in the studies refers to the integration of allocation operations with utilities and services in limited resource conditions to meet user needs and cloud applications.

Studies in [5, 12] presents studies on virtualization solutions that focus on Data Center. Studies in [7, 14] present new strategies in resource allocation that studies adaptation to the inheritance used in cloud computing environments.

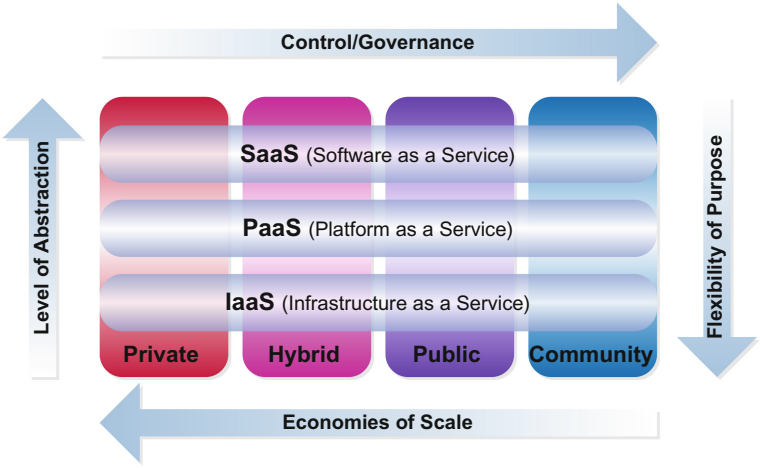


Fig. 1. The basic model of cloud computing

In addition, the approach to process of CRA on the cloud based on a network virtualization perspective. The final product of network virtualization is the virtual network. Studies in [1] presents solutions that focus on optimizing bandwidth consumption. Communication environment is virtualized by the virtual network to ensure the connection to the virtual server [3].

As shown in Fig. 2, each VM can deploy one or more distributed systems over the virtual network through a communications environment. Therefore, the optimization of communication resources in the virtual network environment should be considered.

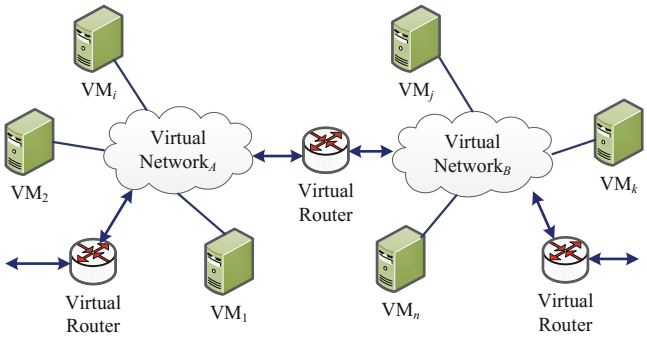


Fig. 2. Virtual distributed network

The processing that takes place inside the system is the set of virtual hosts $VM = \sum S_i\{i = 1..n\}$, where S_i is a virtual machine connected through the distributed communication system as shown in Fig. 2. Virtual machines coordinate control through message passing mechanisms to respond resources. These messages can also be considered as processes are moving in the network to request shared resources. The execution of processes depends on the communication resource allocator for routing from source to destination.

Studies in [4, 11, 13] have shown that optimal communication is based on tree-based algorithms in multicast transmissions. The disadvantage of this mechanism, however, is overlap data in communication if multiple servers provide the same information to the destination servers [8, 9]. To overcome this shortcoming, the challenge in research is offer a solution to eliminate duplicate packets. In this paper, we present a network coding technique to solve duplicate packets at the destination set.

Network coding implements concurrent data transmission over the channels to the destination of multicast transmissions by encoding packets together [2]. A single multicast session i has the communication rate between information flow and physical flow denoted tl_i to transmit packets from source to destination node.

In network coding technique, each message sent in the outcoming link of a node may have some calculation functions such as "*mixing*" other messages in incoming link of node as shown in Fig. 3 [2]. Thus, network coding technique is transmission, mixing (or encoding) and re-mixing (or re-encoding) of incoming messages at a node. As a result, messages can be republished (or decoded) at destination node. Three main advantages of network coding for communication are shown as optimal throughput, minimal power per bit, and reduced latency.

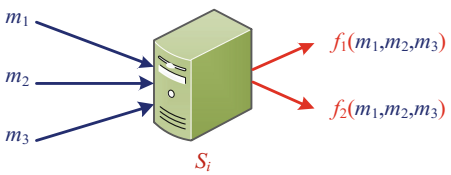


Fig. 3. Node may have some calculation functions for incoming messages

Which network coding, linear equations in the form of encoded packets with code coefficients stored in the encoded packet header, undefined variables are the actual contents of the packet. Therefore, network coding can be integrated with the software defined infrastructure (SDI) in the virtual system to increase throughput, while avoiding duplication of data at the destination node.

3 Systems Model

In order to achieve performance over multicast transmission, the solution of us is built on multicast communication rate control with network coding on the communication channels.

The description of the basic components in Fig. 4 sets out the general problem of cloud resource allocation as follows:

1. Terminals in part 5 require shared resources $YC = \sum yc_i\{i = 1..n\}$ to resource allocator of distributed system in part 2 through communication environment in part 4.
2. After receiving and processing shared resources at resource allocator of distributed system, *YC* send to network coding resources allocator in part 3 to processing.
3. The resource allocator in part 3 divides the *YC* into vectors, encoded and transmitted to servers in the distributed system through the communications resource allocator in part 2.
4. The vectors that are transmitted to destination set S_i in the distributed system for receiving, analyzing, calculating, processing and responding resources to communications resource allocator.
5. After completing the processing of shared resources, communications resource allocator to return result to terminal state of shared resource through resource allocator of distributed system.

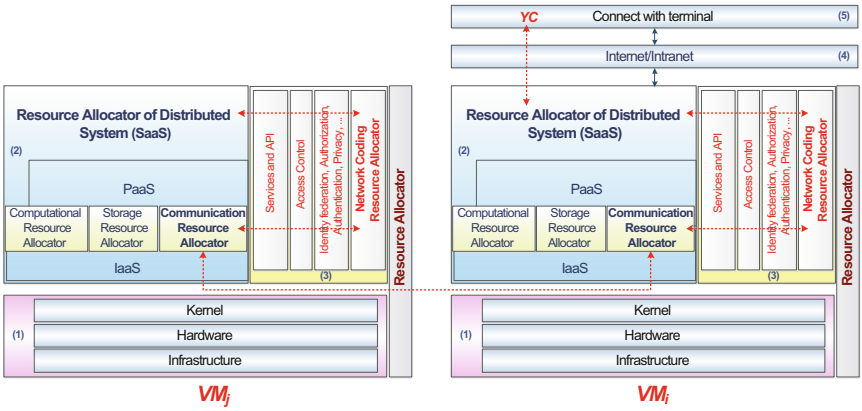


Fig. 4. The overview model of resources allocation in virtual system

This general problem sets out two basic requirements: ensuring requires of shared resources allocation and optimize communication in the system. In this paper, we focus on the optimal communication solution that ensures packet transmitted from source to destination set.

Proposition 1: Consider graph $G(U, V)$: weighted, directed, non-cyclic; denote h is mincut between source S_0 and any intermediate $S_t((t \subset T) \in U)$. Algorithm 1.1 generates a linear multicast network coding on a finite field F. Algorithm 1.1 in Fig. 5 has an execution time $O(|V| \cdot |T| \cdot h(h + |T|))$. Any finite field size $|\mathrm{F}| \geq |T|$ can be used to represent symbols sent along the edges. The results of linear codes from Algorithm 1.1 have the following characteristics:

- The source given incoming messages h.
- Node S_i needs time to execute $O\left(\min\left(\left|V_I\left(S_{ij}\right)\right|, |T|\right)\right)$ to compute messages sent along a discrete edge. The source needs time to execute $O(h)$ for each edge.
- Each intermediate can recover the original messages h in time $O(h^2)$.

Prove:

Let $start\left(S_{ij}\right)$ is a node for edge starting S_{ij}, for each S_{ij} we denote length of local encoding vector $\left|V_I\left(start\left(S_{ij}\right)\right)\right|$ according to formula (1):

$$\overrightarrow{m_{S_{ij}}} = V_I\left(start\left(S_{ij}\right)\right) \to P^{\left|V_I\left(start\left(S_{ij}\right)\right)\right|} \tag{1}$$

formula (1) is a vector that determines linear combination of messages on $V_I\left(start\left(S_{ij}\right)\right)$ generates messages on S_{ij}. $y\left(S_{ij}\right)$ was message carried edge S_{ij} according to formula (2) as follows:

$$y(S_{ij}) = \sum_{p \in V_I\left(start\left(S_{ij}\right)\right)} \overrightarrow{m_{S_{ij}}(p)} y(p) \tag{2}$$

Our task to find coefficients so that all intermediate can reconstruct original packets from messages that achieve them. We present the parallel edges $S_{12}, \ldots, S_{ih}$ from some new source $S_0^{'}$ to S_0; edges carried messages from S_0.

We can describe the efficiency of all local encoding vectors on independent edge S_{ij} of a message using global encoding vector $\overrightarrow{b\left(S_{ij}\right)} \in \mathsf{F}^h$. Length vector h is denoted by $\overrightarrow{b\left(S_{ij}\right)}$, it is linear combination of message $y\left(S_{ij}\right)$. Thus, $\overrightarrow{b\left(S_{ij}\right)} = \left[0^{i-1}, 1, 0^{h-i}\right]$ (length vector h with 1 at coordinate i) is calculated by formula (3) as follows:

$$\overrightarrow{b(S_{ij})} = \sum_{p \in V_I\left(start\left(S_{ij}\right)\right)} \overrightarrow{m_{S_{ij}}(p)} y(p), \forall S_{ij} \in V \tag{3}$$

vector $\overrightarrow{b\left(S_{ij}\right)}$ is defined because the network is non-cyclic. Using linear basic algebra, it can be considered as linear coding scheme that used for multicast transmission from source to destination only if $\forall S_Z \in D$, vector $\left\{\overrightarrow{b(p)} : p \in V_I(t)\right\}$ in span F^h. Recover the original messages can be achieved by linear equations through the variables h.

With many intermediate, our approach is to add more flow S_0–S_t. Steps of Algorithm through $S_i \in U$ in topological order. This ensures that global encoding vector of all edges to S_i are known when local encoding vectors of the edges outgoing of S_i are defined. The algorithm calculates coefficients of $\overrightarrow{ms_{ij}}$ for $S_{ij} \in V_O(S_i)$, one edge at a time. There may be multiple paths to intermediate through edge S_{ij}. Let $T\left(S_{ij}\right)$ denote the intermediate set using S_{ij} in some paths p^t and

$$P(S_{ij}) = \{p^t \leftarrow (S_{ij}) : t \in T(S_{ij})\} \tag{4}$$

formula (4) is preprocessing S_{ij} in the corresponding flow path. Non-zero coefficients for $\overrightarrow{m_{S_{ij}}}$ are chosen for $P(S_{ij})$.

Line 15 in Algorithm 1.1 implemented random selection of vectors $\overrightarrow{m_{S_{ij}}}$ with support of $P(S_{ij})$ with condition $\forall t \in T(S_{ij}) : (B_t \backslash \{b(p^t \leftarrow (S_{ij}))\}) \cup \{\overrightarrow{b(S_{ij})}\}$, algorithm can execute at time $O(|V| \cdot |T| \cdot h^2)$ and return encoding message with time $O(h^2)$ at each intermediate.

Algorithm 1.1: Linear Network Coding Determine Algorithm

Input : Graph $G(U,V)$, source S_0, destination set D

Output: Disjoint path h from S_0 to D, vector $\overrightarrow{m_{S_{ij}}}$, intermediate $t \in T$ and weight $c \in C$, Galois field F

1 $h := \min_{d \in D} \min \{|C| : C$ is mincut from S_0 to $S_Z\}$;
2 insert new source S_0' into U;
3 insert h parallel edge $S_{12}, .., S_{ih}$ from S_0' to S_0 into U;
4 p^t denote the disjoint path h from S_0 to S_t;
5 Galois field F;
6 **foreach** i **do**
7 $\quad \overrightarrow{b(S_i)} = [0^{i-1}, 1, 0^{h-i}]$;
8 $\quad c_i := c_{S_{ij}}$;
9 **foreach** $t \in T$ **do**
10 $\quad C_T := \{c_{S_{01}}, .., c_{S_{it}}\}$;
11 $\quad B_T := \{\overrightarrow{b(S_{01})}, .., \overrightarrow{b(S_{it})}\}$;
12 $\quad$ **foreach** $c_t \in C_T$ **do**
13 $\quad\quad a_t(c_t) := \overrightarrow{b(c_t)}$;
14 **foreach** $S_i \in U \backslash \{S_0'\}$ **do**
15 $\quad$ **foreach** $V_O(S_{ij})$ **do**
16 $\quad\quad \overrightarrow{b(S_{ij})} = \sum_{p \in V_I(start(S_{ij}))} \overrightarrow{m_{S_{ij}}(p)} y(p)$;
17 $\quad\quad \forall t \in T(S_{ij}) : (B_t \backslash \{\overrightarrow{b(p^t \leftarrow (S_{ij}))}\}) \cup \{\overrightarrow{b(S_{ij})}\}$;
18 $\quad\quad$ **foreach** $t \in T(S_{ij})$ **do**
19 $\quad\quad\quad$ advance set of edge C_T;
20 $\quad\quad\quad$ update B_T;
21 $\quad\quad\quad$ update a_t;
22 $\quad\quad\quad$ **foreach** $c_t \in C_T \backslash \{p^t \leftarrow (S_{ij})\}$ **do**
23 $\quad\quad\quad\quad$ update a_t';
24 **return** $(h, \{\overrightarrow{m_{S_{ij}}} : S_{ij} \in V\}, \{(c_t, a_t) : t \in T\}, \mathrm{F})$;

Fig. 5. The Linear network coding determine algorithm

Consider pair $(x_i, y_i) \in F^h \times F^h$ with $x_i \times y_i \neq 0$, $1 \leq i \leq n (n < |F|)$. There exists a linear combination u of $x_1, x_2, \ldots, x_n$ to $u \times y_i \neq 0$, $1 \leq i \leq n$, vector u can be found in the execution time $O(|T|^2 h)$.

Thus, we have total time of execution according to formula (5):

$$O(|V|(|T| \cdot h^2 + |T|^2 h)) = O(|V| \cdot |T| \cdot h(h + |T|)) \tag{5}$$

The total execution time of the network coding compared to multicast transmission is shown in Table 1:

Table 1. Execution time of the network coding and multicast transmission

	Multicast	Network coding
Node	$O(h^2)$	$O(h)$
Edge	$O(h^2 T)$	$O(h^2 T)$
Network	$O(\|V\| \cdot \|T\|^2 \cdot h^4)$	$O(\|V\| \cdot \|T\| \cdot h(h + \|T\|))$

4 Simulation Results

In order to evaluate simulation of network coding, simulation requires declaration of initial variables to perform encode and decode of packets transmitted in network. Finite field value $F(2^m)$ performs operations. Block variable is initialized to transmitted message in network. Length of block variable in bytes.

```
Id: 0 Payload: A0 A0 A0 A0 A0
Id: 1 Payload: A1 A1 A1 A1 A1
Id: 2 Payload: A2 A2 A2 A2 A2
Id: 3 Payload: A3 A3 A3 A3 A3
Id: 4 Payload: A4 A4 A4 A4 A4
Id: 5 Payload: A5 A5 A5 A5 A5
Id: 6 Payload: A6 A6 A6 A6 A6
Id: 7 Payload: A7 A7 A7 A7 A7
Id: 8 Payload: A8 A8 A8 A8 A8
Id: 9 Payload: A9 A9 A9 A9 A9
```

Fig. 6. Uncoded package

Figure 6 shows the input data is the number of uncoded packets, based on variable length payload and block variables, data is divided into blocks. Each packet in the packet has an ID value to identify packet reconstruction at the destination. Figure 7 shows a linear combination of blocks represented in the encoded, each encoded packet containing an encoding vector describing uncoded packets that have been merged.

```
01 00 00 00 00 00 00 00 00 00 | 00 05 00 05 00 05 00 05 00 05
00 01 00 00 00 00 00 00 00 00 | 08 05 08 05 08 05 08 05 08 05
00 00 01 00 00 00 00 00 00 00 | 04 05 04 05 04 05 04 05 04 05
00 00 00 01 00 00 00 00 00 00 | 12 05 12 05 12 05 12 05 12 05
00 00 00 00 01 00 00 00 00 00 | 02 05 02 05 02 05 02 05 02 05
00 00 00 00 00 01 00 00 00 00 | 10 05 10 05 10 05 10 05 10 05
00 00 00 00 00 00 01 00 00 00 | 06 05 06 05 06 05 06 05 06 05
00 00 00 00 00 00 00 01 00 00 | 14 05 14 05 14 05 14 05 14 05
00 00 00 00 00 00 00 00 01 00 | 01 05 01 05 01 05 01 05 01 05
00 00 00 00 00 00 00 00 00 01 | 09 05 09 05 09 05 09 05 09 05
```

Fig. 7. Linear combinations of blocks represented in the encoded

Intermediate performs linear combination of received packet with random coefficient. Intermediate generates many combinations of uncoded packets to ensure that encoded packets can be decodable with high probability. The output of the packets when transmitted in network are two matrices according to Fig. 8.

```
13 03 09 03 10 10 05 05 01 01 | 07 07 07 07 07 07 07 07 07 07
05 12 12 15 01 09 11 10 04 02 | 12 02 12 02 12 02 12 02 12 02
02 01 15 11 13 05 15 00 08 02 | 12 04 12 04 12 04 12 04 12 04
14 05 12 13 12 02 09 13 05 02 | 09 08 09 08 09 08 09 08 09 08
10 12 14 14 15 15 04 10 14 13 | 14 01 14 01 14 01 14 01 14 01
12 11 05 00 12 08 08 14 04 08 | 10 09 10 09 10 09 10 09 10 09
15 12 15 12 06 11 02 14 08 10 | 08 15 08 15 08 15 08 15 08 15
07 04 06 11 08 00 09 15 09 11 | 00 10 00 10 00 10 00 10 00 10
00 00 14 01 14 15 15 00 00 10 | 03 01 03 01 03 01 03 01 03 01
13 01 13 12 12 10 03 09 13 09 | 08 02 08 02 08 02 08 02 08 02
```

Fig. 8. The packet output in network

After reaching destination set, the decoder will decode the random linear combinations. When packets are received, the decoder executes as many encoded packets as possible to return the decoded blocks to recover original packets. The resulting decoded packets is shown in Fig. 9, which indicates that the packet ID may not be in order at the decoder during transmission from source to destination. We reorder to receive the original packet.

```
Id: 1 Payload: A1 A1 A1 A1 A1
Id: 0 Payload: A0 A0 A0 A0 A0
Id: 2 Payload: A2 A2 A2 A2 A2
Id: 3 Payload: A3 A3 A3 A3 A3
Id: 4 Payload: A4 A4 A4 A4 A4
Id: 5 Payload: A5 A5 A5 A5 A5
Id: 6 Payload: A6 A6 A6 A6 A6
Id: 7 Payload: A7 A7 A7 A7 A7
Id: 8 Payload: A8 A8 A8 A8 A8
Id: 9 Payload: A9 A9 A9 A9 A9
```

Fig. 9. The packet is decoded at destination

Through simulations packet transmitted based on network coding, packets are split, mixed and transmitted over network. At the destination set, packets are decoded and re-structured primitive to avoid packet overlap at destination with high decoded probability. Throughput at destination of three methods of transmission: unicast, multicast and network coding is shown in Fig. 10. The throughput of network coding is better than multicast transmission. This shows that network coding technique achieves optimal communication over multicast transmission.

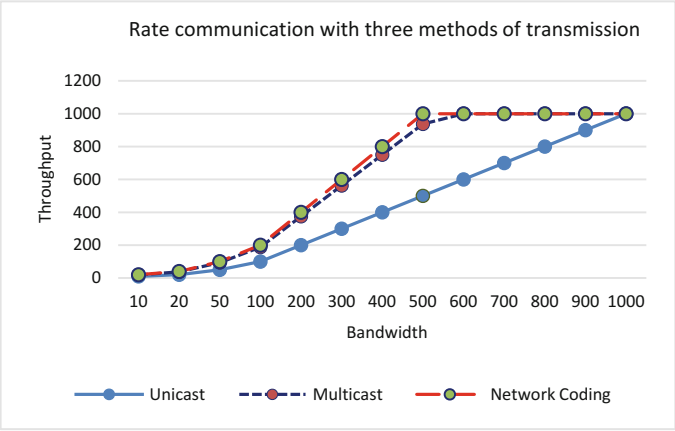


Fig. 10. Maximum throughput at the destination node with 3 methods of transmission

5 Conclusions

In this paper, we studied the optimal solution of CRA in distributed system integrated on cloud computing. We have developed an algorithm that linear network coding determine, multicast transmission combines network coding technique algorithm to ensure that the destination set receives non-overlap information and reaches maximum throughput. The solution offered is tested for accuracy and can be applied to large, complex systems. In our next research direction, the solutions are communication resources allocation with multi-source, multi-rate into multiple transmission channels. In addition, reduction of encoded, decoded complexity at intermediate and destination should also be considered.

References

1. Buyya, R., Beloglazov, A., Abawajy, J.H.: Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges. Comput. Res. Repos. **1006**(0308), 6–20 (2010)
2. Chou, P.A., Wu, Y., Jain, K.: Practical network coding. In: Allerton Conference on Communication, Control and Computing (2003)

3. Chowdhury, N.M.M.K., Boutaba, R.: Network virtualization: state of the art and research challenges. IEEE Commun. Mag. **47**(7), 20–26 (2009)
4. Erciyes, K.: Distributed Graph Algorithms for Computer Networks. Springer, Heidelberg (2013). https://doi.org/10.1007/978-1-4471-5173-9
5. Kong, X., et al.: Efficient dynamic task scheduling in virtualized data centers with fuzzy prediction. J. Netw. Comput. Appl. **34**(4), 1068–1077 (2011)
6. Kshemkalyani, A.D., Singhal, M.: Distributed Computing: Principles, Algorithms, and Systems. Cambridge University Press, Cambridge (2008)
7. Lee, Y.C., Zomaya, A.Y.: Rescheduling for reliable job completion with the support of clouds. Future Gener. Comput. Syst. **26**(8), 1192–1199 (2010)
8. Li, S.R., Yeung, R.W., Cai, N.: Linear network coding. IEEE Trans. Inf. Theory **49**(2), 371–381 (2003)
9. Li, Z., Li, B.: Network coding in undirected networks. In: Conference on Information Sciences and Systems (2004)
10. Raynal, M.: Distributed Algorithms for Message-Passing Systems. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38123-2
11. Tel, G.: Introduction to Distributed Algorithms. Cambridge University Press, Cambridge (1994)
12. Urgaonkar, R., et al.: Dynamic resource allocation and power management in virtualized data centers. In: Network Operations and Management Symposium (NOMS), pp. 479–486. IEEE (2010)
13. Vinothina, V., Sridaran, R., Ganapathi, P.: A survey on resource allocation strategies in cloud computing. Int. J. Adv. Comput. Sci. Appl. **3**, 8 (2012)
14. Toby, V., Anthony, V., Robert, E.: Cloud Computing, A Practical Approach. McGraw-Hill, Inc., New York (2010)

Author Index