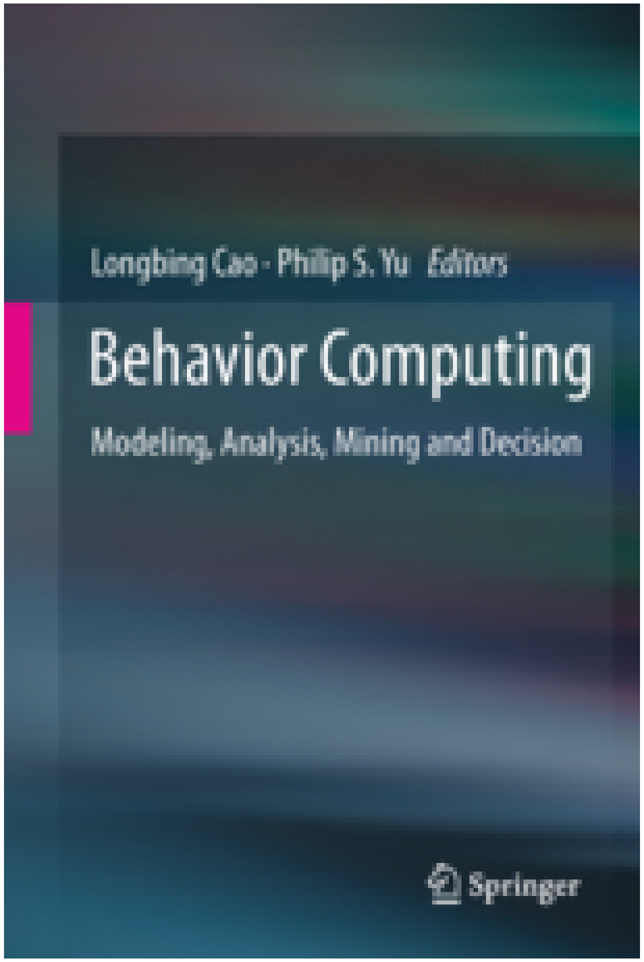
Behavior Informatics: An Overview of Related Research

Prof Longbing Cao
University of Technology Sydney

Theoretical Research

-- building the theory of behavior informatics

Related Books



The concept of behavior informatics

Information Sciences 180 (2010) 3067–3085

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

In-depth behavior understanding and use: The behavior informatics approach

Longbing Cao

Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

A R T I C L E   I N F O

Article history:
Received 10 May 2009
Received in revised form 20 March 2010
Accepted 24 March 2010

Keywords:
Informatics
Behavior analysis
Behavior informatics
Behavior computing
Decision making

A B S T R A C T

The in-depth analysis of human behavior has been increasingly recognized as a crucial means for disclosing interior driving forces, causes and impact on businesses in handling many challenging issues such as behavior modeling and analysis in virtual organizations, web community analysis, counter-terrorism and stopping crime. The modeling and analysis of behaviors in virtual organizations is an open area. Traditional behavior modeling mainly relies on qualitative methods from behavioral science and social science perspectives. On the other hand, so-called behavior analysis is actually based on human demographic and business usage data, such as churn prediction in the telecommunication industry, in which behavior-oriented elements are hidden in routinely collected transactional data. As a result, it is ineffective or even impossible to deeply scrutinize native behavior intention, lifecycle and impact on complex problems and business issues. In this paper, we propose the approach of *behavior informatics* (BI), in order to support explicit and quantitative behavior involvement through a conversion from source data to behavioral data, and further conduct genuine analysis of behavior patterns and impacts. BI consists of key components including *behavior representation*, *behavioral data construction*, *behavior impact analysis*, *behavior pattern analysis*, *behavior simulation*, and *behavior presentation* and *behavior use*. We discuss the concepts of behavior and an abstract behavioral model, as well as the research tasks, process and theoretical underpinnings of BI. Two real-world case studies are demonstrated to illustrate the use of BI in dealing with complex enterprise problems, namely analyzing exceptional market microstructure behavior for market surveillance and mining for high impact behavior patterns in social security data for governmental debt prevention. Substantial experiments have shown that BI has the potential to greatly complement the existing empirical and specific means by finding deeper and more informative patterns leading to greater in-depth behavior understanding. BI creates new directions and means to enhance the quantitative, formal and systematic modeling and analysis of behaviors in both physical and virtual organizations.

Longbing Cao: **In-depth behavior understanding and use: The behavior informatics approach.** Inf. Sci. 180(17): 3067-3085 (2010)

Editors: **Longbing Cao,** University of Technology, Sydney, longbing.cao@uts.edu.au

Behavior Informatics: A New Perspective

Longbing Cao, *University of Technology, Sydney*

Behavior is a concept increasingly recognized in broad communities spreading from social to business, online, mobile, economic, and cultural domains. However, systematic and comprehensive methodologies, theories, tools, and systems aren't ready for deeply, fully, and effectively capturing, representing, quantifying, analyzing, learning, and measuring the semantics, sequencing, networking, evolution, utility and impact of individual, group, and cohort behaviors taking place in the real world. This is becoming fundamental and critical in the age of Big Data. Here, in this installment of "Trends & Controversies," we look at how *behavior informatics* targets the development of effective methodologies and techniques to tackle these issues.

social and collaborative searching activities is needed. Gabriella Pasi presents insights on engaging behaviors in information seeking, especially considering coupled behaviors within certain contexts.

Nowadays, an increasing number of users are interested in IPTV programs online, and generate massive amounts of activities. Ya Zhang and her colleagues lead a discussion about the behaviors of IPTV users that are related to system efficiency, personalization, recommendation, and targeted advertisement.

Finally, Edoardo Serra and V.S. Subrahmanian raise an interesting question: Should behavior models of terror groups be disclosed? They share their research and arguments on strategic disclosures and consequences in tackling today's terrorism.

Longbing Cao, Thorsten Joachims, Can Wang, Éric Gaussier, Jinjiu Li, Yuming Ou, Dan Luo, Reza Zafarani, Huan Liu, Guandong Xu, Zhiang Wu,Gabriella Pasi, Ya Zhang, Xiaokang Yang, Hongyuan Zha, Edoardo Serra, V. S. Subrahmanian: **Behavior Informatics: A New Perspective.** IEEE Intelligent Systems 29(4): 62-80 (2014)

Behavior Informatics and Analytics: Let Behavior Talk

Longbing Cao
Data Sciences and Knowledge Discovery Lab
Faculty of Engineering and Information Technology,
University of Technology, Sydney, Australia
lbcao@it.uts.edu.au

Abstract

Behavior is increasingly recognized as a key component in business intelligence and problem-solving. Different from traditional behavior analysis, which mainly focus on implicit behavior and explicit business appearance as a result of business usage and customer demographics, this paper proposes the field of Behavior Informatics and Analytics *(BIA), to support explicit behavior involvement through a conversion from transactional data to behavioral data, and further genuine analysis of native behavior patterns and impacts. BIA consists of key components including behavior modeling and representation, behavioral data construction, behavior impact modeling, behavior pattern analysis, and behavior presentation. BIA can greatly complement the existing means for combined, more informative and social patterns and solutions for critical problem-solving in areas such as dealing with customer-officer interaction, counter-terrorism and monitoring online communities.*

1. Introduction

Human behavior has been increasingly highlighted for pattern analysis and business intelligence in many areas such as customer relationship management, social computing [16], intrusion detection [15], fraud detection [9], event analysis [17], outlier detection [11], and group decision-making. For instance, in customer relationship management [12], it is widely agreed that customer behavior analysis is essentially important for deeply understanding and caring for customers, and eventually boosting enterprise operation and enhancing business intelligence. Other typical examples include web usage and user preference analysis [8, 13, 14], churn analysis of telecommunication customers from one provider to another [1], credit estimation of banking customers in home loan and doing finance [2], exceptional behavior analysis of terrorist and criminals [7], and trading pattern analysis of investors in capital markets [9].

To the best of our knowledge, the above behavior-oriented analysis was usually conducted on customer demographic and transactional data directly. For instance, in telecom churn analysis, customer demographic data and service usage data are analyzed to classify customers into loyal and non-loyal groups based on the dynamics of usage change; while in outlier mining of trading behavior, price movement is usually focused to detect abnormal behavior. In activity monitoring [10], static and appearance-oriented data is focused. In scrutinizing the datasets used in the above examples, we realize that the so-called behavior-oriented analysis is actually not on customer behavior-oriented elements, rather on straightforward customer demographic data and business usage related transactions accumulated during business processes (altogether transactional data).

In general, customer demographic and transactional data is not organized in terms of behavior but entity relationships. Entities and their relationships collected in transactions reflect those objects closely related to particular business problems. For instance, in stock market, orderbook transactions in trading engines mainly record and manage price, volume, value and index information related to traders' decisions. Such data is normally seen and analyzed by both financial and IT researchers and practitioners.

Consequently, human behavior is *implicit* in normal transactional data. Such *behavior implication* indicates the limitation or even ineffectiveness of supporting behavior-oriented analysis on transactional data directly. The main reasons include the following aspects.

- First, the behavior implication in transactional data determines that it cannot support in-depth analysis on *behavior interior* which is surrounded by behavioral elements, but on *behavior exterior* that excludes behavioral elements from average data such as service usage.
- Second, with behavior implied in transactional data, it is not possible to scrutinize behavioral intention and impact on business appearance and problems; while behavior may play important roles in the appearance

978-0-7695-3503-6/08 $25.00 © 2008 IEEE
DOI 10.1109/ICDMW.2008.95

IEEE computer society

Longbing Cao:
Behavior Informatics and Analytics: Let Behavior Talk. ICDM Workshops 2008: 87-96

Issues Addressed

- What is behavior?
- What is an abstract behavior model?
- What are the key behavioral factors?
- What is the conceptual map of behavior informatics?
- Major challenges and research issues of behavior informatics
- Case studies of behavior analysis

Behavior representation

Formalization and Verification of Group Behavior Interactions

Can Wang, Longbing Cao, *Senior Member, IEEE*, and Chi-Hung Chi

Abstract—Group behavior interactions, such as multirobot teamwork and group communications in social networks, are widely seen in both natural, social, and artificial behavior-related applications. Behavior interactions in a group are often associated with varying coupling relationships, for instance, conjunction or disjunction. Such coupling relationships challenge existing behavior representation methods, because they involve multiple behaviors from different actors, constraints on the interactions, and behavior evolution. In addition, the quality of behavior interactions are not checked through verification techniques. In this paper, we propose an ontology-based behavior modeling and checking system (OntoB for short) to explicitly represent and verify complex behavior relationships, aggregations, and constraints. The OntoB system provides both a visual behavior model and an abstract behavior tuple to capture behavioral elements, as well as building blocks. It formalizes various intra-coupled interactions (behaviors conducted by the same actor) via transition systems (TSs), and inter-coupled behavior aggregations (behaviors conducted by different actors) from temporal, inferential, and party-based perspectives. OntoB converts a behavior-oriented application into a TS and temporal logic formulas for further verification and refinement. We demonstrate and evaluate the effectiveness of the OntoB in modeling multirobot behaviors and their interactions in the Robocup soccer competition game. We show, that the OntoB system can effectively model complex behavior interactions, verify and refine the modeling of complex group behavior interactions in a sound manner.

Index Terms—Behavior interaction, coupling relationship, group behavior, model checking.

I. Introduction

BEHAVIOR refers to the action or reaction of any material under given circumstances and environment. It is intrinsic in many areas, and behavior analysis has become a fundamental topic which has been increasingly investigated as an essential activity in many fields, from social and behavioral sciences to computer science [1], [2]. In Google, the keyword "behavior" attracts 379 000 000 hits while "behavior interaction" achieves 202 000 000 results, searched on 4th Dec. 2014.

In both natural and social sciences and applications, multiple behaviors from one or multiple actors often interact with one another, which are called coupled behaviors or group behavior interactions. They play important roles in group-based activities such as social networking and multirobot teamwork. These coupled behaviors and behavior interactions may form interior driving forces that shape underlying businesses, such as in online community and social networks [3], or may even cause challenging problems like group-based manipulation by a group of traders [4] or serious traffic jams resulting from haphazard interactions between vehicles traveling in different directions toward an intersection. With the deepening and widening of complex networking, coupled behaviors, or group behavior interactions are increasingly seen in both mainstream and emerging situations, in particular, in enterprise applications, organizations, complex systems, online, and social communities.

We illustrate coupled behaviors and behavior interactions using the example of multirobot soccer game in Fig. 1. As shown in Fig. 1, two teams participate in a Robocup soccer competition (http://www.robocup.org/) with four Sony AIBO robots in each group. The robot players operate on their own without any external control, either by humans or by computers. They communicate with each other by wireless or by using the speakers and microphones. Their interactions include the collaborations among different actions of the same robot, e.g., one of the robots kicks the ball after it gets a message; and distinct operations conducted by different robots, such as sending messages between different players. As shown in the scenario described by Ros and Veloso [5], a team of robots intelligently cooperate with one another and self-adjust their own activities; the successful task execution and problem resolution rely on the proper implementation of an individual robot's activities as well as collaborative interactions between robots. If a robot undertakes tasks without appropriate arrangement and coordination with the other robots, the Robocup is likely to be unsuccessful, even though every robot performs perfectly. This example shows that group actors and behaviors by the same or different actors within the group are often coupled in different forms of interactions [6], and it is essential to identify, represent, and verify how the robots interact to ensure the performance of a multirobot system.

To enable the above behavior interaction-oriented systems to work properly, a fundamental task is to develop effective behavior representation tools to capture, formalize, and verify behavioral elements, coupling relationships, and interactions between behaviors, in both qualitative and quantitative

Manuscript received February 24, 2014; revised June 10, 2014; accepted November 8, 2014. Date of publication February 24, 2015; date of current version July 15, 2015. This paper was recommended by Associate Editor Y. Wang.

C. Wang and C.-H. Chi are with the Digital Productivity Flagship, Commonwealth Scientific and Industrial Research Organisation, Sandy Bay, TAS 7005, Australia (e-mail: canwang613@gmail.com).

L. Cao is with the Advanced Analytics Institute, University of Technology, Sydney, NSW 2007, Australia.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

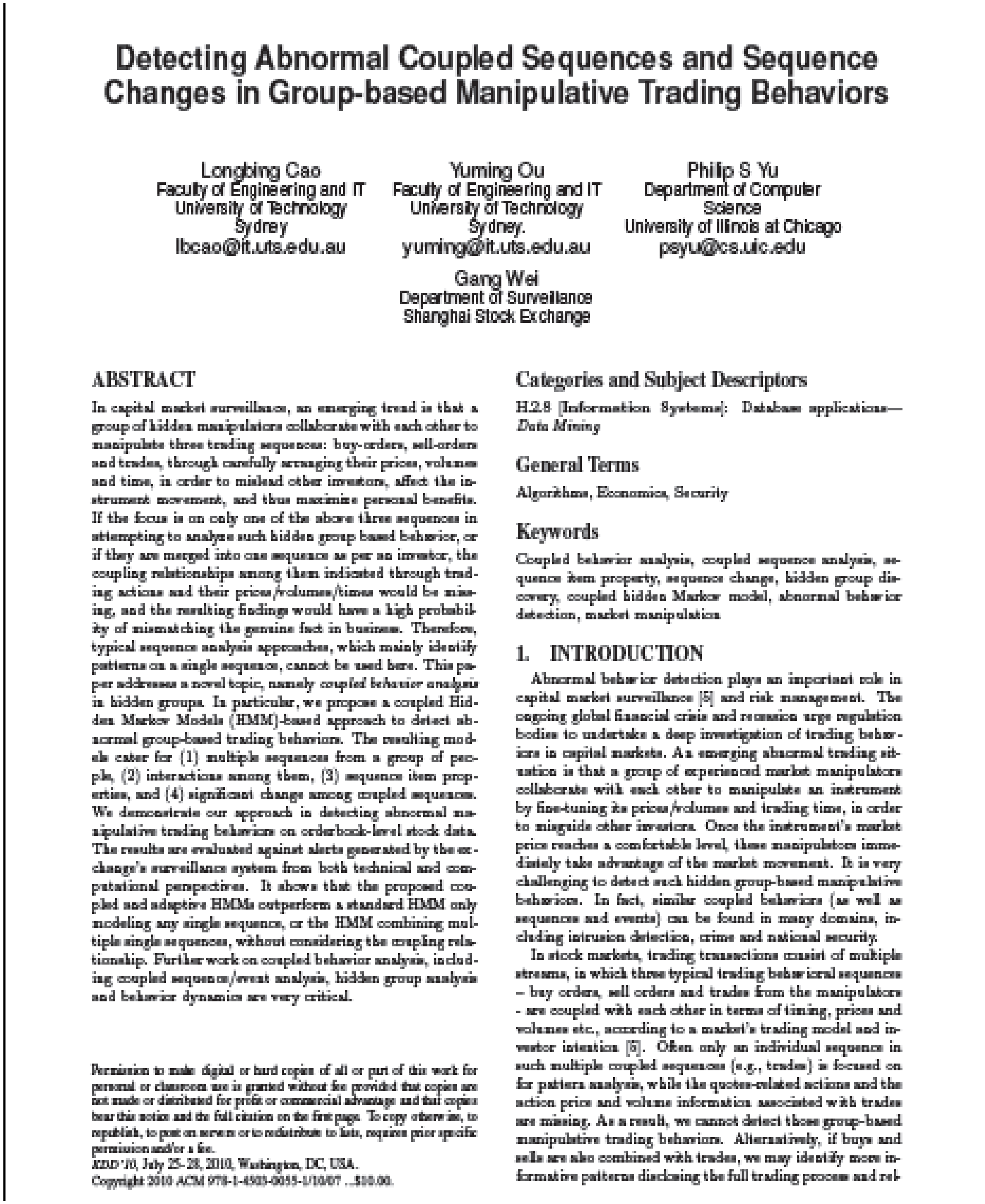
Digital Object Identifier 10.1109/TSMC.2015.2399862

Can Wang, Longbing Cao, Chi-Hung Chi: **Formalization and Verification of Group Behavior Interactions.** IEEE T. Systems, Man, and Cybernetics: Systems 45(8): 1109-1124 (2015)

Can Wang, and Longbing Cao. **Modeling and Analysis of Social Activity Process**, in Longbing Cao and Philip S Yu (eds) Behavior Computing, 21-35, Springer, 2012

Issues Addressed

- How to represent behaviors?
- Behavior ontology
- Behavior process and interaction
- Behavior interaction relationship
- How to model check behavior models built?
- Case studies of behavior modeling



Detecting Abnormal Coupled Sequences and Sequence Changes in Group-based Manipulative Trading Behaviors

Longbing Cao
Faculty of Engineering and IT
University of Technology Sydney
lbcao@it.uts.edu.au

Yuming Ou
Faculty of Engineering and IT
University of Technology Sydney.
yuming@it.uts.edu.au

Philip S Yu
Department of Computer Science
University of Illinois at Chicago
psyu@cs.uic.edu

Gang Wei
Department of Surveillance
Shanghai Stock Exchange

ABSTRACT

In capital market surveillance, an emerging trend is that a group of hidden manipulators collaborate with each other to manipulate three trading sequences: buy-orders, sell-orders and trades, through carefully arranging their prices, volumes and time, in order to mislead other investors, affect the instrument movement, and thus maximize personal benefits. If the focus is on only one of the above three sequences in attempting to analyze such hidden group based behavior, or if they are merged into one sequence as per an investor, the coupling relationships among them indicated through trading actions and their prices/volumes/times would be missing, and the resulting findings would have a high probability of mismatching the genuine fact in business. Therefore, typical sequence analysis approaches, which mainly identify patterns on a single sequence, cannot be used here. This paper addresses a novel topic, namely *coupled behavior analysis* in hidden groups. In particular, we propose a coupled Hidden Markov Models (HMM)-based approach to detect abnormal group-based trading behaviors. The resulting models cater for (1) multiple sequences from a group of people, (2) interactions among them, (3) sequence item properties, and (4) significant change among coupled sequences. We demonstrate our approach in detecting abnormal manipulative trading behaviors on orderbook-level stock data. The results are evaluated against alerts generated by the exchange's surveillance system from both technical and computational perspectives. It shows that the proposed coupled and adaptive HMMs outperform a standard HMM only modeling any single sequence, or the HMM combining multiple single sequences, without considering the coupling relationship. Further work on coupled behavior analysis, including coupled sequence/event analysis, hidden group analysis and behavior dynamics are very critical.

KDD'10, July 25–28, 2010, Washington, DC, USA.

Categories and Subject Descriptors

H.2.8 [Information Systems]: Database applications—*Data Mining*

General Terms

Algorithms, Economics, Security

Keywords

Coupled behavior analysis, coupled sequence analysis, sequence item property, sequence change, hidden group discovery, coupled hidden Markov model, abnormal behavior detection, market manipulation

1. INTRODUCTION

Abnormal behavior detection plays an important role in capital market surveillance [8] and risk management. The ongoing global financial crisis and recession urge regulation bodies to undertake a deep investigation of trading behaviors in capital markets. An emerging abnormal trading situation is that a group of experienced market manipulators collaborate with each other to manipulate an instrument by fine-tuning its prices/volumes and trading time, in order to misguide other investors. Once the instrument's market price reaches a comfortable level, these manipulators immediately take advantage of the market movement. It is very challenging to detect such hidden group-based manipulative behaviors. In fact, similar coupled behaviors (as well as sequences and events) can be found in many domains, including intrusion detection, crime and national security.

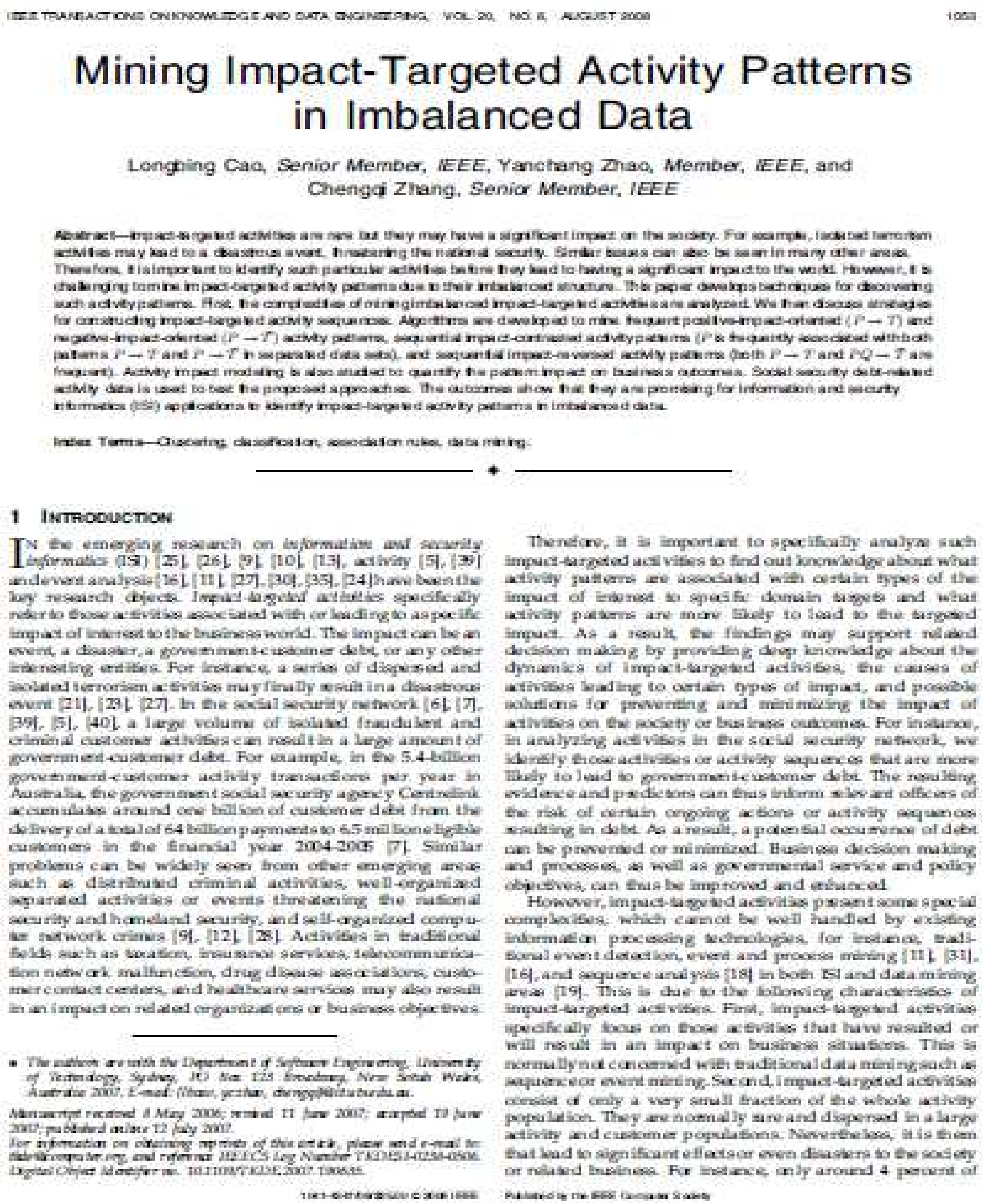
In stock markets, trading transactions consist of multiple streams, in which three typical trading behavioral sequences – buy orders, sell orders and trades from the manipulators - are coupled with each other in terms of timing, prices and volumes etc., according to a market's trading model and investor intention [5]. Often only an individual sequence in such multiple coupled sequences (e.g., trades) is focused on for pattern analysis, while the quotes-related actions and the action price and volume information associated with trades are missing. As a result, we cannot detect those group-based manipulative trading behaviors. Alternatively, if buys and sells are also combined with trades, we may identify more informative patterns disclosing the full trading process and rel-

Behavior sequence analysis

Longbing Cao, Yuming Ou, Philip S YU, Gang Wei. **Detecting Abnormal Coupled Sequences and Sequence Changes in Group-based Manipulative Trading Behaviors**, KDD2010, 85-94.

Issues Addressed

- Behavior properties described by attribute vector
- How to construct behavior sequences?
- How to handle multiple sequences coupled with each other?
- How to model vector-based behavior sequences?
- How to map vector-based behavior sequences to Coupled Hidden Markov Model?
- How to detect pool manipulation by identifying abnormal coupled sequences?



Mining Impact-Targeted Activity Patterns in Imbalanced Data

Longbing Cao, *Senior Member, IEEE*, Yanchang Zhao, *Member, IEEE*, and Chengqi Zhang, *Senior Member, IEEE*

Abstract—Impact-targeted activities are rare but they may have a significant impact on the society. For example, isolated terrorism activities may lead to a disastrous event, threatening the national security. Similar issues can also be seen in many other areas. Therefore, it is important to identify such particular activities before they lead to having a significant impact to the world. However, it is challenging to mine impact-targeted activity patterns due to their imbalanced structure. This paper develops techniques for discovering such activity patterns. First, the complexities of mining imbalanced impact-targeted activities are analyzed. We then discuss strategies for constructing impact-targeted activity sequences. Algorithms are developed to mine frequent positive-impact-oriented ($P \rightarrow T$) and negative-impact-oriented ($P \rightarrow \bar{T}$) activity patterns, sequential impact-contrasted activity patterns (P is frequently associated with both patterns $P \rightarrow T$ and $P \rightarrow \bar{T}$ in separated data sets), and sequential impact-reversed activity patterns (both $P \rightarrow T$ and $PQ \rightarrow \bar{T}$ are frequent). Activity impact modeling is also studied to quantify the pattern impact on business outcomes. Social security debt-related activity data is used to test the proposed approaches. The outcomes show that they are promising for information and security informatics (ISI) applications to identify impact-targeted activity patterns in imbalanced data.

Index Terms—Clustering, classification, association rules, data mining.

1 Introduction

In the emerging research on *information and security informatics* (ISI) [25], [26], [9], [10], [13], activity [5], [39] and event analysis [16], [11], [27], [30], [33], [24] have been the key research objects. *Impact-targeted activities* specifically refer to those activities associated with or leading to a specific impact of interest to the business world. The impact can be an event, a disaster, a government-customer debt, or any other interesting entities. For instance, a series of dispersed and isolated terrorism activities may finally result in a disastrous event [21], [23], [27]. In the social security network [6], [7], [39], [5], [40], a large volume of isolated fraudulent and criminal customer activities can result in a large amount of government-customer debt. For example, in the 5.4-billion government-customer activity transactions per year in Australia, the government social security agency Centrelink accumulates around one billion of customer debt from the delivery of a total of 64 billion payments to 6.5 million eligible customers in the financial year 2004-2005 [7]. Similar problems can be widely seen from other emerging areas such as distributed criminal activities, well-organized separated activities or events threatening the national security and homeland security, and self-organized computer network crimes [9], [12], [28]. Activities in traditional fields such as taxation, insurance services, telecommunication network malfunction, drug disease associations, customer contact centers, and healthcare services may also result in an impact on related organizations or business objectives.

Therefore, it is important to specifically analyze such impact-targeted activities to find out knowledge about what activity patterns are associated with certain types of the impact of interest to specific domain targets and what activity patterns are more likely to lead to the targeted impact. As a result, the findings may support related decision making by providing deep knowledge about the dynamics of impact-targeted activities, the causes of activities leading to certain types of impact, and possible solutions for preventing and minimizing the impact of activities on the society or business outcomes. For instance, in analyzing activities in the social security network, we identify those activities or activity sequences that are more likely to lead to government-customer debt. The resulting evidence and predictors can thus inform relevant officers of the risk of certain ongoing actions or activity sequences resulting in debt. As a result, a potential occurrence of debt can be prevented or minimized. Business decision making and processes, as well as governmental service and policy objectives, can thus be improved and enhanced.

However, impact-targeted activities present some special complexities, which cannot be well handled by existing information processing technologies, for instance, traditional event detection, event and process mining [11], [31], [16], and sequence analysis [18] in both ISI and data mining areas [19]. This is due to the following characteristics of impact-targeted activities. First, impact-targeted activities specifically focus on those activities that have resulted or will result in an impact on business situations. This is normally not concerned with traditional data mining such as sequence or event mining. Second, impact-targeted activities consist of only a very small fraction of the whole activity population. They are normally rare and dispersed in a large activity and customer populations. Nevertheless, it is them that lead to significant effects or even disasters to the society or related business. For instance, only around 4 percent of

- *The authors are with the Department of Software Engineering, University of Technology, Sydney, PO Box 123 Broadway, New South Wales, Australia 2007. E-mail: {lbcao, yczhao, chengqi}@it.uts.edu.au.*

Manuscript received 8 May 2006; revised 11 June 2007; accepted 19 June 2007; published online 12 July 2007.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDE-0238-0506. Digital Object Identifier no. 10.1109/TKDE.2007.190635.

Behavior impact analysis

Longbing Cao. Zhao Y., Zhang, C. **Mining Impact-Targeted Activity Patterns in Imbalanced Data**, IEEE Trans. on Knowledge and Data Engineering, 20(8): 1053-1066, 2008.

Issues Addressed

- What is impact of behavior?
- How to model behavior impact?
- How to construct impact-based behavior sequences?
- How to identify high impact behavior sequences?
- How to identify combined behavior sequences associated with impact?
- How to manage behavior patterns through combined impact-targeted behavior sequences?

USpan: An Efficient Algorithm for Mining High Utility Sequential Patterns

Junfu Yin
Advanced Analytics Institute
University of Technology,
Sydney, Australia
junfu.yin@student.uts.edu.au

Zhigang Zheng
Advanced Analytics Institute
University of Technology,
Sydney, Australia
zgzheng@it.uts.edu.au

Longbing Cao*
Advanced Analytics Institute
University of Technology,
Sydney, Australia
longbing.cao@uts.edu.au

ABSTRACT

Sequential pattern mining plays an important role in many applications, such as bioinformatics and consumer behavior analysis. However, the classic frequency-based framework often leads to many patterns being identified, most of which are not informative enough for business decision-making. In frequent pattern mining, a recent effort has been to incorporate utility into the pattern selection framework, so that high utility (frequent or infrequent) patterns are mined which address typical business concerns such as dollar value associated with each pattern. In this paper, we incorporate utility into sequential pattern mining, and a generic framework for high utility sequence mining is defined. An efficient algorithm, USpan, is presented to mine for high utility sequential patterns. In USpan, we introduce the lexicographic quantitative sequence tree to extract the complete set of high utility sequences and design concatenation mechanisms for calculating the utility of a node and its children with two effective pruning strategies. Substantial experiments on both synthetic and real datasets show that USpan efficiently identifies high utility sequences from large scale data with very low minimum utility.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining

General Terms

Algorithms

Keywords

High utility sequential pattern mining, Sequential pattern mining

*Author for correspondence.

1. INTRODUCTION

Sequential pattern mining has emerged as an important topic in data mining. It has proven to be very essential for handling order-based critical business problems, such as behavior analysis, gene analysis in bioinformatics and weblog mining. For example, sequence analysis is widely employed in DNA and protein to discover interesting structures and functions of molecular or DNA sequences. The selection of interesting sequences is generally based on the frequency/support framework: sequences of high frequency are treated as significant. Under this framework, the downward closure property (also known as *Apriori property*) [1] plays a fundamental role for varieties of algorithms designed to search for frequent sequential patterns [10, 14, 5].

In practice, many patterns are identified by frequent sequential pattern mining algorithms. Most of them may not be informative to business decision-making, since they do not show the business value and impact. In some cases, such as fraud detection, some truly interesting sequences may be filtered because of their low frequencies. For example, in retail business, selling a car generally leads to much higher profit than selling a bottle of milk, while the frequency of cars sold is much lower than that of milk. In online banking fraud detection, the transfer of a large amount of money to an unauthorized overseas account may appear once in over one million transactions, yet it has a substantial business impact. Such problems cannot be tackled by the frequency/support framework.

This brings about an interesting question: how to mine sequential patterns of business interest? In the related area, utility is introduced into frequent pattern mining to mine for patterns of high utility by considering the quality (such as profit) of itemsets. This has led to high utility pattern mining [13], which selects interesting patterns based on minimum utility rather than minimum support. Let us use a toy example to illustrate. Table 1 shows the items and their respective weights or profit (quality) appearing in an online retail store. Table 2 collects several shopping sequences with quantities; each transaction in the sequence consists of one to multiple items, and each item is associated with a quantity showing how many of this item were purchased. For instance, the first sequence $\langle(c,5)[(c,2)(f,1)](h,2)\rangle$ shows three itemsets $(c,5)$, $[(c,2)\ (f,1)]$ and $(h,2)$, and the quantity purchased of item, e.g. the quantity of c is 5. Following the high utility pattern mining concept, a possible calculation of utility of an itemset is to consider its total profit.

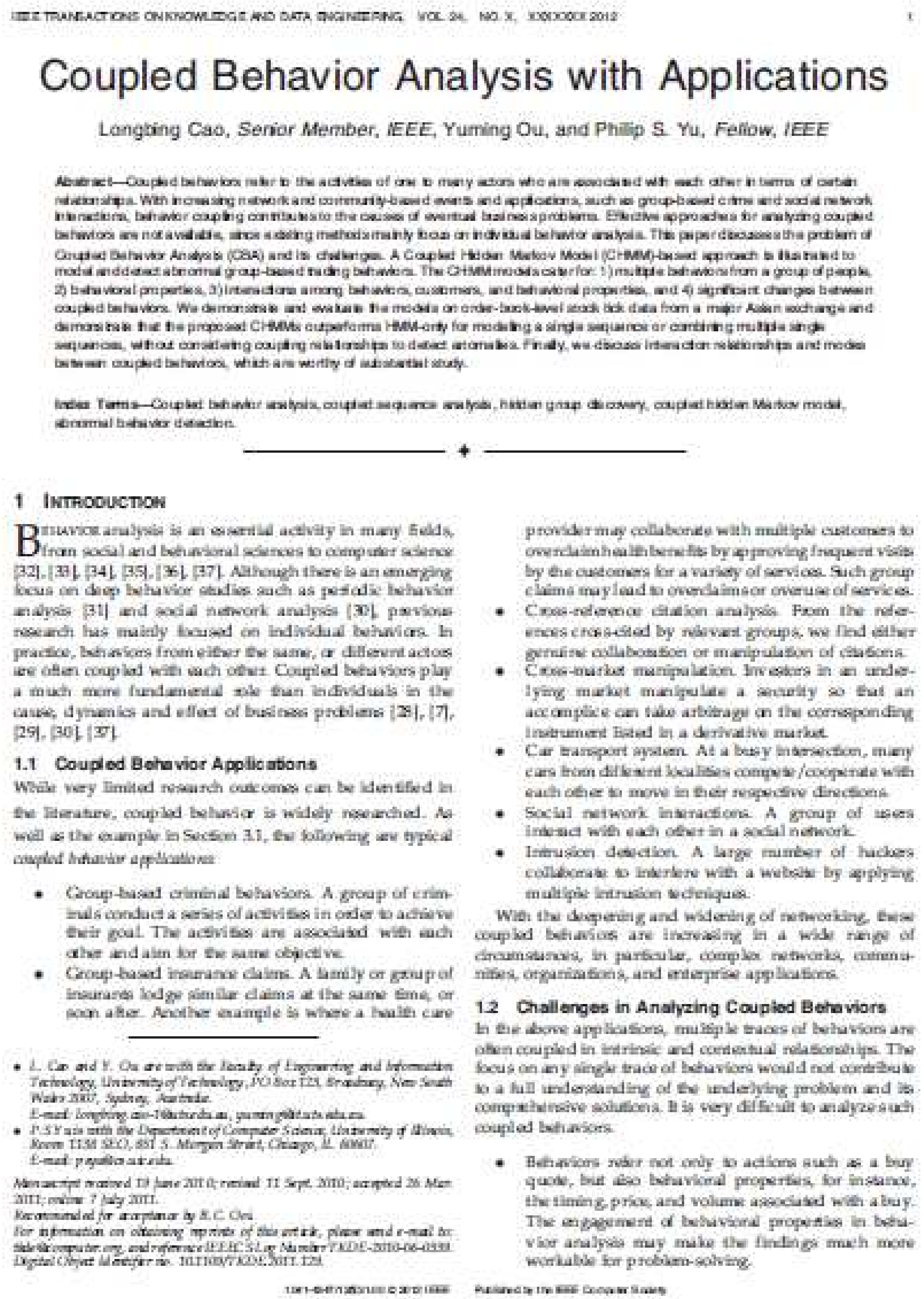
Behavior utility analysis

Junfu Yin, Zhigang Zheng, Longbing Cao. **USpan: An Efficient Algorithm for Mining High Utility Sequential Patterns**, KDD 2012, 660-668.

Junfu Yin, Zhigang Zheng, Longbing Cao, Yin Song, Wei Wei: **Efficiently Mining Top-K High Utility Sequential Patterns.** ICDM 2013: 1259-1264

Issues Addressed

- What is utility of a sequence?
- How to quantify sequence utility?
- How to define high utility sequences?
- How to identify high utility sequences?



Coupled Behavior Analysis with Applications

Longbing Cao, *Senior Member, IEEE*, Yuming Ou, and Philip S. Yu, *Fellow, IEEE*

Abstract—Coupled behaviors refer to the activities of one to many actors who are associated with each other in terms of certain relationships. With increasing network and community-based events and applications, such as group-based crime and social network interactions, behavior coupling contributes to the causes of eventual business problems. Effective approaches for analyzing coupled behaviors are not available, since existing methods mainly focus on individual behavior analysis. This paper discusses the problem of Coupled Behavior Analysis (CBA) and its challenges. A Coupled Hidden Markov Model (CHMM)-based approach is illustrated to model and detect abnormal group-based trading behaviors. The CHMM models cater for: 1) multiple behaviors from a group of people, 2) behavioral properties, 3) interactions among behaviors, customers, and behavioral properties, and 4) significant changes between coupled behaviors. We demonstrate and evaluate the models on order-book-level stock tick data from a major Asian exchange and demonstrate that the proposed CHMMs outperforms HMM-only for modeling a single sequence or combining multiple single sequences, without considering coupling relationships to detect anomalies. Finally, we discuss interaction relationships and modes between coupled behaviors, which are worthy of substantial study.

Index Terms—Coupled behavior analysis, coupled sequence analysis, hidden group discovery, coupled hidden Markov model, abnormal behavior detection.

1 Introduction

Behavior analysis is an essential activity in many fields, from social and behavioral sciences to computer science [32], [33], [34], [35], [36], [37]. Although there is an emerging focus on deep behavior studies such as periodic behavior analysis [31] and social network analysis [30], previous research has mainly focused on individual behaviors. In practice, behaviors from either the same, or different actors are often coupled with each other. Coupled behaviors play a much more fundamental role than individuals in the cause, dynamics and effect of business problems [28], [7], [29], [30], [37].

1.1 Coupled Behavior Applications

While very limited research outcomes can be identified in the literature, coupled behavior is widely researched. As well as the example in Section 3.1, the following are typical *coupled behavior applications*:

- Group-based criminal behaviors. A group of criminals conduct a series of activities in order to achieve their goal. The activities are associated with each other and aim for the same objective.
- Group-based insurance claims. A family or group of insurants lodge similar claims at the same time, or soon after. Another example is where a health care provider may collaborate with multiple customers to overclaim health benefits by approving frequent visits by the customers for a variety of services. Such group claims may lead to overclaims or overuse of services.
- Cross-reference citation analysis. From the references cross-cited by relevant groups, we find either genuine collaboration or manipulation of citations.
- Cross-market manipulation. Investors in an underlying market manipulate a security so that an accomplice can take arbitrage on the corresponding instrument listed in a derivative market.
- Car transport system. At a busy intersection, many cars from different localities compete/cooperate with each other to move in their respective directions.
- Social network interactions. A group of users interact with each other in a social network.
- Intrusion detection. A large number of hackers collaborate to interfere with a website by applying multiple intrusion techniques.

With the deepening and widening of networking, these coupled behaviors are increasing in a wide range of circumstances, in particular, complex networks, communities, organizations, and enterprise applications.

1.2 Challenges in Analyzing Coupled Behaviors

In the above applications, multiple traces of behaviors are often coupled in intrinsic and contextual relationships. The focus on any single trace of behaviors would not contribute to a full understanding of the underlying problem and its comprehensive solutions. It is very difficult to analyze such coupled behaviors.

- Behaviors refer not only to actions such as a buy quote, but also behavioral properties, for instance, the timing, price, and volume associated with a buy. The engagement of behavioral properties in behavior analysis may make the findings much more workable for problem-solving.

- *L. Cao and Y. Ou are with the Faculty of Engineering and Information Technology, University of Technology, PO Box 123, Broadway, New South Wales 2007, Sydney, Australia.*
- *P.S. Yu is with the Department of Computer Science, University of Illinois, Room 1138 SEO, 851 S. Morgan Street, Chicago, IL 60607.*

Manuscript received 18 June 2010; revised 11 Sept. 2010; accepted 26 Mar. 2011; online 7 July 2011.
Recommended for acceptance by B.C. Ooi.
Digital Object Identifier no. 10.1109/TKDE.2011.129.

Published by the IEEE Computer Society

Group behavior analysis

Longbing Cao, Yuming Ou, Philip S. Yu: **Coupled Behavior Analysis with Applications.** IEEE Trans. Knowl. Data Eng. 24(8): 1378-1392 (2012)

Yin Song, Longbing Cao: **Graph-based coupled behavior analysis: A case study on detecting collaborative manipulations in stock markets.** IJCNN 2012: 1-8

Yin Song, Longbing Cao, Xindong Wu, Gang Wei, Wu Ye, Wei Ding: **Coupled behavior analysis for capturing coupling relationships in group-based market manipulations.** KDD 2012: 976-984

Issues Addressed

- What is group behavior?
- How to model coupling relationships between behaviors of one actor?
- How to model coupling relationships between behaviors of multiple actors?
- What is the problem of coupled behavior analysis for understanding group behaviors?
- As a case study, how to use coupled behavior analysis to understand pool manipulation in stock market?

Combined behavior analysis

Focus Article

Combined mining: Analyzing object and pattern relations for discovering and constructing complex yet actionable patterns

Longbing Cao*

Combined mining is a technique for analyzing object relations and pattern relations, and for extracting and constructing actionable knowledge (patterns or exceptions). Although combined patterns can be built within a single method, such as combined sequential patterns by aggregating relevant frequent sequences, this knowledge is composed of multiple constituent components (the left hand side) from multiple data sources, which are represented by different feature spaces, or identified by diverse modeling methods. In some cases, this knowledge is also associated with certain impacts (influence, action, or conclusion, on the right hand side). This paper presents an abstract high-level picture of combined mining and the combined patterns from the perspective of object and pattern relation analysis. Several fundamental aspects of combined pattern mining are discussed, including feature interaction, pattern interaction, pattern dynamics, pattern impact, pattern relation, pattern structure, pattern paradigm, pattern formation criteria, and pattern presentation (in terms of pattern ontology and pattern dynamic charts). We also briefly illustrate the concepts and discuss how they can be applied to mining complex data for complex knowledge in either a multifeature, multisource, or multimethod scenario.

How to cite this article:
WIREs Data Mining Knowl Discov 2013, 3: 140–155 doi: 10.1002/widm.1080

INTRODUCTION

In this paper, we introduce the concept of combined (pattern) mining. Combined mining is mainly suitable for handling the complexity of employing multifeature sets, multi-information sources, constraints, multimethods, and multimodels in data mining, and for analyzing complex relations between objects or descriptors (attributes, sources, methods, constraints, labels, and impacts) or between identified patterns during the learning process. Combined patterns may be formed through analysis of the internal relations between objects or pattern constituents obtained by a single method on a single dataset, for instance, combined sequential patterns formed from analyzing the relations within a discovered sequential pattern space.

With the exception of object and pattern relation analysis, which is a very new topic in the data mining community, many approaches and algorithms are available in the literature on other aspects of the above combinations. The main contribution of combined mining is that it enables the extraction, discovery, construction, and induction of knowledge, which consists of not simply discriminant objects but also of interactions and relations between objects, as well as their impact. They are referred to as *complex but actionable patterns*, because they reflect pattern elements and relations, which form certain pattern structures and dynamics, and indicate decision-making actions.

Combined mining provides an overall solution for meeting the challenge of mining complex knowledge in complex data.[1] It also substantially builds upon other individual approaches such as conceptual inductive learning[2,3] and inference, generalization, aggregation, and summarization,[4,5] in order to

*Correspondence to: longbing.cao@uts.edu.au
Advanced Analytics Institute, University of Technology Sydney, Australia

DOI: 10.1002/widm.1080

Jingyu Shao, Junfu Yin, Wei Liu, Longbing Cao. **Mining actionable combined patterns of high utility and frequency**. DSAA 2015: 1-10.

Longbing Cao. **Combined Mining: Analyzing Object and Pattern Relations for Discovering and Constructing Complex but Actionable Patterns**, WIREs Data Mining and Knowledge Discovery, 3(2): 140-155, 2013.

Longbing Cao. Zhao Y., Zhang, C. **Mining Impact-Targeted Activity Patterns in Imbalanced Data**, IEEE Trans. on Knowledge and Data Engineering, 20(8): 1053-1066, 2008.

Yanchang Zhao, Huaifeng Zhang, Longbing Cao, Chengqi Zhang. **Combined Pattern Mining: from Learned Rules to Actionable Knowledge**, LNCS 5360/2008, 393-403, 2008.

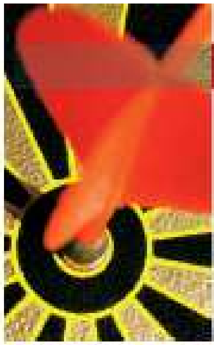
Issues Addressed

- What are the challenges of behavior interactions?
- What are the types, forms and relations of interactions between behaviors?
- How do behaviors of one actor or multiple actors combine with each other?
- What are combinations of behavior patterns?
- What are issues in group behavior combinations?
- How to make behavior patterns actionable?

Nonoccurring behavior analysis

EXPERT OPINION

Editor: Daniel Zeng, University of Arizona and Chinese Academy of Sciences, zengdaniel@gmail.com



Nonoccurring Behavior Analytics: A New Area

Longbing Cao, *University of Technology Sydney*
Philip S. Yu, *University of Illinois at Chicago*
Vipin Kumar, *University of Minnesota*

Behavior-related studies and applications, such as behavior analysis, data mining, machine learning, and behavioral science, have generally focused on behaviors that have occurred or will occur. Such behaviors are called *positive behaviors* (PBs) or *occurring behaviors* (OBs). Related work has focused on behavioral patterns, anomalies, impact, and dynamics. This constitutes the area of behavior analytics, which focuses on understanding, analyzing, learning, predicting, and managing past, present, and future behaviors. When behavior representation and modeling are also considered, we use the term *behavior informatics* or *behavior computing*[1] to describe the new perspective of modeling, reasoning about, verifying, analyzing, learning, and evaluating behaviors. This has emerged as an important and demanding area for comprehensively and deeply handling ubiquitous behaviors online, in business, government services, scientific activities, social activities, and economic and financial business.

Limited research has been conducted on analyzing, detecting, or predicting nonoccurring behaviors (NOBs), those that did not or will not occur. NOBs are also called *negative behaviors*, which are not straightforward, since they usually are hidden and difficult to understand, or one usually is not concerned with them. That NOBs are overlooked does not mean they are unimportant. For instance, if a patient misses an appointment with a specialist, and thus misses the opportunity to receive immediate and appropriate treatment for a health problem, the patient's health could worsen. Additionally, in many situations, failure to follow rules or policies could result in administrative or even legal obligations. Therefore, it is important to build a theoretical foundation for NOB study.

Unfortunately, few research outcomes of NOB study can be identified in the literature. Relevant work includes event analysis; negative association rule mining,[2] which identifies patterns comprising nonoccurring items; and negative sequential patterns,[3–8] which comprise sequential elements that do not appear in the business process. No systematic work has been conducted to understand, model, formalize, analyze, learn, detect, predict, intervene, and manage NOBs.

NOB is not a trivial problem. Some may argue that it is simple to treat an NOB as a special OB, and that all relevant techniques can then be used directly for NOB analytics. Unfortunately, this often does not work for reasons related to the different natures and complexities of occurring and nonoccurring behaviors. In this article, we outline the concept of NOBs and related complexities, draw a picture of NOB analytics, and present our view of NOB research directions and prospects.

What Is NOB?

We briefly discuss the essence, intrinsic characteristics, and complexities of NOBs, and the forms that NOBs can take, in order to understand the concept of NOB.

Intrinsic Characteristics

NOBs refer to those behaviors that should occur but do not for some reason. They are hidden but are widely seen in behavioral applications in business, economics, health, cyberspace, social and mobile networks, and natural and human systems. Many businesses, services, applications, and systems involve NOBs, including healthcare and

1541-1672/15/$31.00 © 2015 IEEE
Published by the IEEE Computer Society

Longbing Cao, Philip S. Yu, Vipin Kumar. **Nonoccurring Behavior Analytics: A New Area**. IEEE Intelligent Systems 30(6): 4-11 (2015).

Xiangjun Dong, Zhigang Zhao, Longbing Cao, Yanchang Zhao, Chengqi Zhang, Jinjiu Li, Wei Wei, Yuming Ou. **e-NSP: Efficient Negative Sequential Pattern Mining Based on Identified Positive Patterns Without Database Rescanning**, CIKM 2011, 825-830.

Zhigang Zheng, Yanchang Zhao, Ziye ZuoLongbing Cao, Huaifeng Zhang, Yanchang Zhao, Chengqi Zhang. **An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns**, PAKDD2010, 262-273.

Yanchang Zhao, Huaifeng Zhang, Longbing Cao, Chengqi Zhang and Hans Bohlscheid. **Mining Both Positive and Negative Impact-Oriented Sequential Rules From Transactional Data**, PAKDD2009, pp.656-663.

Yanchang Zhao, Huaifeng Zhang, Shanshan Wu, Jian Pei, Longbing Cao, Chengqi Zhang and Hans Bohlscheid. **Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns**, ECML/PKDD2009, 648-663, 2009.

Issues Addressed

- What is non-occurring behavior?
- Why do we care about non-occurring behaviors?
- What are issues in understanding non-occurring behaviors?
- What are the problems with existing behavior study in addressing non-occurring behaviors?
- Research opportunities and prospects of non-occurring behavior study

Behavior non-IIDness analysis

The Computer Journal Advance Access published August 22, 2013

doi:10.1093/comjnl/bxt084

Non-IIDness Learning in Behavioral and Social Data

LONGBING CAO*

Advanced Analytics Institute, University of Technology, Sydney, Australia
**Corresponding author: longbing.cao@uts.edu.au*

Most of the classic theoretical systems and tools in statistics, data mining and machine learning are built on the fundamental assumption of IIDness, which assumes the independence and identical distribution of underlying objects, attributes and/or values. However, complex behavioral and social problems often exhibit strong couplings and heterogeneity between values, attributes and objects (i.e., non-IIDness). This fundamentally challenges the IIDness-based learning methodologies and techniques. This paper presents a high-level overview of the needs, challenges and opportunities of non-IIDness learning for handling complex behavioral and social problems. By reviewing the nature and issues of classic IIDness-based algorithms in frequent pattern mining, clustering and classification to complex behavioral and social applications, concepts, structures, frameworks and exemplar techniques are discussed for non-IIDness learning. Case studies, related work and prospects of non-IIDness learning are presented. Non-IIDness learning is also a fundamental issue in big data analytics.

Keywords: non-IIDness learning; IIDness, IID data; non-IID data; coupling; behavior informatics; social informatics

Received 13 February 2013; revised 5 July 2013
Handling editor: Guandong Xu

1. INTRODUCTION

Behavioral and social applications are ubiquitous, ranging from business and online applications to social and organizational applications and domains. With the increasing and continuous development of such applications, an emerging need is to develop an in-depth understanding of the underlying working mechanism, driving force, dynamics and evolution of a behavioral and/or social system, as well as the impact on business and context. To this end, building on the classic theories and tools available in behavioral science and social science, behavior informatics [1, 2] and social informatics [3][1] have recently been studied to 'formalize', 'quantify' and 'compute' complex behavioral and social applications.

As an emerging area of research, behavior and social informatics is in its earliest stage and features many challenges and opportunities. A canonical trend is to develop theories, tools and algorithms based on the classic outcomes available in extant disciplines including statistics, data mining and machine learning. Typically, frequent pattern mining, clustering and classification of behavioral and social applications are conducted by expanding the corresponding existing theories and algorithms. In this paper, we discuss the potential issues and risk in pursuing this path for complex behavioral and social applications by explicitly or implicitly taking the IIDness assumption, and thus reveal the need for developing non-IIDness learning for behavior and social informatics.

Arguably, most of the existing theories, tools and systems in statistics, data mining and machine learning are built on the IIDness assumption, which assumes the independence and identical distribution of the underlying objects, attributes and/or values. Based on a high-level abstraction, it is assumed that objects, attributes and values are independent and identically distributed, with most of existing learning theories, models and algorithms proposed on the basis of this assumption. This works well in simple business applications and abstract problems with weakened and avoidable relations and heterogeneity, and serves as the foundation of classic analytics, mining and learning theories, algorithms, systems and tools.

Complex behavioral and social applications often exhibit strong coupling relations (which are beyond the usual dependency relation) and heterogeneity between objects, object

[1] See more from the IEEE Task Force on Behavior and Social Informatics and Computing: www.bsic.info

Longbing Cao. **Non-IIDness Learning in Behavioral and Social Data**, The Computer Journal, 57(9): 1358-1370 (2014).

Longbing Cao. **Coupling Learning of Complex Interactions**, Journal of Information Processing and Management, 51(2): 167-186 (2015).

Issues Addressed

- Behaviors are non-IID, namely not independent and identically distributed (IID)
- What is the non-IIDness of behavior-related problems?
- What are coupling relationships of non-IID behaviors?
- What are heterogeneity of non-IID behaviors?
- What are opportunities and prospects of non-IID behavior and social problem study?

References

- Longbing Cao, Philip S Yu (Eds). Behavior Computing: Modeling, Analysis, Mining and Decision, Springer, 2012.
- Longbing Cao; Hiroshi Motoda, Jaideep Srivastava, Ee-Peng Lim, Irwin King, Philip S. Yu, Wolfgang Nejdl, Guandong Xu, Gang Li, Ya Zhang (Eds.). Behavior and Social Computing, Proceedings of International Workshop on Behavior and Social Informatics and Computing, Lecture Notes in Computer Science, Vol. 8178, Springer, 2013
- Can Wang,Longbing Cao, Chi-Hung Chi. Formalization and Verification of Group Behavior Interactions. IEEE T. Systems, Man, and Cybernetics: Systems 45(8): 1109-1124 (2015).
- Jingyu Shao, Junfu Yin, Wei Liu,, Longbing Cao. Mining actionable combined patterns of high utility and frequency. DSAA 2015: 1-10.
- Longbing Cao, Philip S. Yu, Vipin Kumar. Nonoccurring Behavior Analytics: A New Area. IEEE Intelligent Systems 30(6): 4-11 (2015).
- Wei Cao, Longbing Cao. Financial Crisis Forecasting via Coupled Market State Analysis, IEEE Intelligent Systems, 30(2): 18-25 (2015).
- Philippe Fournier-Viger, Cheng-Wei Wu, Vincent S. Tseng, Longbing Cao, Roger Nkambou. Mining Partially-Ordered Sequential Rules Common to Multiple Sequences, IEEE Trans. Knowledge and Data Engineering, 27(8): 2203-2216 (2015).
- Longbing Cao. Behavior Informatics: A New PerspectiveIEEE Intelligent Systems (Trends and Controversies), 29(4): 62-80, 2014.
- Longbing Cao and Thorsten Joachims. Behavior Computing, IEEE Intelligent Systems, 29(4): 62-66, 2014.

References

- Longbing Cao, Yu, Philip S; Motoda, Hiroshi; Williams, Graham. Special issue on behavior computing (editorial), Knowledge and Information Systems, 37(2): 245-249, 2013.
- Yin Song, Longbing Cao, et al. Coupled Behavior Analysis for Capturing Coupling Relationships in Group-based Market Manipulation, KDD 2012, 976-984.
- Junfu Yin, Zhigang Zheng, Longbing Cao, Yin Song, Wei Wei: Efficiently Mining Top-K High Utility Sequential Patterns. ICDM 2013: 1259-1264
- Junfu Yin, Zhigang Zheng, Longbing Cao. USpan: An Efficient Algorithm for Mining High Utility Sequential Patterns, KDD 2012, 660-668.
- Yin Song and Longbing Cao. Graph-based Coupled Behavior Analysis: A Case Study on Detecting Collaborative Manipulations in Stock Markets, IJCNN 2012, 1-8.
- Longbing Cao, Yuming Ou, Philip S Yu. Coupled Behavior Analysis with Applications (KDD2010 extension), IEEE Trans. on Knowledge and Data Engineering, 24(8): 1378-1392 (2012).
- Zhong She, Can Wang, and Longbing Cao. A Coupled Framework of Clustering Ensembles, AAAI2012 (poster)
- Can Wang, and Longbing Cao.Modeling and Analysis of Social Activity Process, in Longbing Cao and Philip S Yu (eds) Behavior Computing, 21-35, Springer, 2012
- Can Wang, Mingchun Wang, Zhong She, Longbing Cao. CD: A Coupled Discretization Algorithm, PAKDD2012, 407-418
- Can Wang, Longbing Cao, Minchun Wang, Jinjiu Li, Wei Wei, Yuming Ou. Coupled Nominal Similarity in Unsupervised Learning, CIKM 2011, 973-978.

References

- Xiangjun Dong, Zhigang Zhao, Longbing Cao, Yanchang Zhao, Chengqi Zhang, Jinjiu Li, Wei Wei, Yuming Ou.e-NSP: Efficient Negative Sequential Pattern Mining Based on Identified Positive Patterns Without Database Rescanning, CIKM 2011, 825-830.
- Longbing Cao, In-depth Behavior Understanding and Use: the Behavior Informatics Approach, Information Science, 180(17); 3067-3085, 2010.
- Longbing Cao, Yuming Ou, Philip S YU, Gang Wei. Detecting Abnormal Coupled Sequences and Sequence Changes in Group-based Manipulative Trading Behaviors, KDD2010, 85-94.
- Zhigang Zheng, Yanchang Zhao, Ziye ZuoLongbing Cao, Huaifeng Zhang, Yanchang Zhao, Chengqi Zhang. An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns, PAKDD2010, 262-273.
- Longbing Cao, Philip S Yu, Behavior Informatics: An Informatics Perspective for Behavior Studies, The Intelligent Informatics Bulletin, 10(1): 6-11, 2009.
- Zhigang Zheng, Yanchang Zhao, Ziye Zuo, Longbing Cao. Negative-GSP: An Efficient Method for Mining Negative Sequential Patterns, AusDM 2009: 63-67.
- Shanshan Wu, Yanchang Zhao, Huaifeng Zhang, Chengqi Zhang, Longbing Cao, Hans Bohlscheid. Debt Detection in Social Security by Adaptive Sequence Classification, KSEM 2009: 192-203.
- Yanchang Zhao, Huaifeng Zhang, Shanshan Wu, Jian Pei,Longbing Cao, Chengqi Zhang and Hans Bohlscheid.Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns, ECML/PKDD2009, 648-663, 2009.
- Yanchang Zhao, Huaifeng Zhang, Longbing Cao, Chengqi Zhang and Hans Bohlscheid. Mining Both Positive and Negative Impact-Oriented Sequential Rules From Transactional Data, PAKDD2009, pp.656-663.

References

- Longbing Cao, Behavior Informatics and Analytics: Let Behavior Talk, DDDM2008 joint with ICDM2008, 87 - 96.
- Longbing Cao Yuming Ou. Market Microstructure Patterns Powering Trading and Surveillance Agents. Journal of Universal Computer Sciences, 14(14): 2288-2308, 2008.
- Yanchang Zhao, Huaifeng Zhang, Longbing Cao, Chengqi Zhang and Hans Bohlscheid. Efficient Mining of Event-Oriented Negative Sequential Rules, WI 08, pp. 336-342.
- Huaifeng Zhang, Yanchang Zhao, Longbing Cao, Chengqi Zhang and Hans Bohlscheid. Customer Activity Sequence Classification for Debt Prevention in Social Security, Journal of Computer Science and Technology, 24(6): 1000-1009 (2009).
- Yanchang Zhao, Huaifeng Zhang, Shanshan Wu, Jian Pei,Longbing Cao, Chengqi Zhang and Hans Bohlscheid. Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns, ECML/PKDD2009, 648-663, 2009.
- Longbing Cao. Zhao Y., Zhang, C. Mining Impact-Targeted Activity Patterns in Imbalanced Data, IEEE Trans. on Knowledge and Data Engineering, 20(8): 1053-1066, 2008.
- Longbing Cao, Yanchang Zhao, Chengqi Zhang, Huaifeng Zhang. Activity Mining: from Activities to Actions, International Journal of Information Technology & Decision Making, 7(2): 259-273, 2008
- Longbing Cao, Behavior Informatics and Analytics: Let Behavior Talk, DDDM2008 joint with ICDM2008.
- Chengqi Zhang, Longbing Cao. Keynote: Activity Mining to Strengthen Debt Prevention, Pacific Asia Conf. on Intelligence and Security Informatics (PAISI), 2007.
- Longbing Cao, Yanchang Zhao, Fernando Figueiredo, Yuming Ou, Dan Luo. Mining High Impact Exceptional Behavior Patterns, PAKDD2007 industry track, LNCS4819, 56-63, 2007.
- Longbing Cao.Activity mining: challenges and prospects. ADMA2006, LNAI4093, 582-593.

Where to download the references

- http://www-staff.it.uts.edu.au/~lbcao/publication/publications.htm

Comments and questions

- Prof Longbing Cao
  University of Technology Sydney
  Longbing.Cao@gmail.com