Behavior Informatics and Analytics: A New and Promising Area

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

of problems, their roles have been weakened or even ignored as a potential factor in traditional customer behavior analysis.

Why and what does behavior make difference in pattern analysis and business intelligence?

- First, in many cases, behavior plays the role as internal driving forces or causes for business appearance and problems. Most of business problems such as mobile customer churning can be better understood and investigated if customer behavior can be combined and scrutinized.

- Second, when behavior is disclosed and taken as an extra factor in problem-solving solutions, it can greatly complement traditional pattern analysis solely relying on demographic and transactional data, and disclose extra information and relationship between behavior and target business problem-solving. In this way, a multiple-dimensional viewpoint and solution may exist that can uncover problem-solving evidence from not only demographic and transactional but behavioral (including intentional, social and impact aspects) perspectives. As a result, the identified patterns are combined, more informative and social for problem understanding and solving.

In order to support genuine behavior analysis on behavior interior, it is essentially important to make behavior 'explicit' by squeezing out behavior elements hidden in transactional data. For that, *a conversion from transactional space to behavior feature space is necessary*. The conversion extracts, transforms and presents behavior-related elements, and reorganizes them into behavior data that caters for behavior-oriented analysis. This is the process of *behavior modeling and mapping*. As a result, in behavior data, behavior is *explicit*, and is mainly organized in terms of behavior, behavior relationship and impact. This leads to *behavior explication*. On the behavior data, we can then explicitly and more effectively analyze behavior patterns and behavior impacts than on transactional data.

The *behavior modeling and representation*, *construction of behavioral data*, *behavior impact modeling*, *behavior pattern analysis*, and *behavior presentation* consist of the main goals and tasks of behavior informatics and analytics (BIA). On top of the studies on mining activity data and activity sequential patterns [4, 5], and market microstructure behavior analysis for market surveillance [6], this paper presents an overall framework and key concepts of BIA from the perspective of setting up a new scientific field. Please note, limited to the objectives, BIA is mainly from the perspectives of information technology and data analysis rather than from social behavior aspect.

Let's still use the example of churn analysis of mobile customers to distinguish behavior analysis on behavioral data from traditional customer behavior analysis on transactional data. With BIA, besides the analysis on demographic and service usage data, we can further analyze behavior sequences of a customer, including activities happened from his/her registration and activation of a new account into a network, the distribution (such as frequency and duration) of making calls during an observation period, to the characteristics of making payments to the date leaving the network. Obviously, analysis on such data can explore much more fruitful information about mobile holder's intention, activity change, usage dynamics, and payment profile than simply on demographic and service usage data. These are important for disclosing reasons and drivers of churning and loyalty change.

The paper is organized as follows. In Section 2.2, the concepts of behavior and an abstract behavioral model are discussed. Section 3 introduces the content of BIA studies and reasons for proposing it. Section 4 proposes the theoretical underpinnings of BIA. Research issues are discussed in Section 5. We illustrate several typical BIA application areas in Section 6. We conclude the paper in Section 7.

2. Behavior and Behavioral Model

2.1. What Is Behavior About

Under the scope of Behavior Informatics and Analytics (BIA), *behavior* refers to those activities that present as actions, operations or events as well as activity sequences conducted by human beings under certain context and environment. Even though generally behavior may also refer to other actions by organisms such as animals or more physical activities such as the movement of a robot, we are particularly interested in the informatics and analytics for *symbolic behavior* and the analytics of *mapped behavior*.

- Those social activities recorded into computer systems, which present as symbols representing human interaction and operation with a particular object or object system; a typical example is customer behavior, for instance, an investor places an order into a trading system, then the behavior of the trader is to 'place an order', some other examples include web user behavior, game user behavior and intelligent agent behavior; typically, such human behavior happens under certain social context, and therefore presents social characteristics. We call such behavior 'symbolic behavior'.

The symbolic behavior is our main focus in Behavior Informatics and Analytics. However, there is another type of behavior widely seen in areas such as computer vision and pattern recognition. In this case, we see physical behavior either direct or indirectly.

- Those physical activities recorded by sensors into computer systems, which present as the virtual version of an object's actions in the physical world; a typical example is human activities captured by video surveillance systems; some other example includes robot's behavior and organism's behavior in game systems; such captured behavior is the direct or indirect mapping of physical behavior in a virtual world. We call such behavior 'mapped behavior'.

For mapped behavior, our focus is on its analysis rather than modeling. We address this in the Section 3.1.

2.2. An Abstract Behavioral Model

As an abstract concept, behavior (γ) indicates the following attributes and properties:

- Subject (s): The entity (or entities) that issues the activity or activity sequence;
- Object (o): The entity (or entities) on which a behavior is imposed on;
- Context (e): The environment in which a behavior is operated; context may include pre-condition and post-condition of a behavior;
- Goal (g): Goal represents the objectives that the behavior subject would like to accomplish or bring about;
- Belief (b): Belief represents the informational state and knowledge of the behavior subject about the world;
- Action (a): Action represents what the behavior subject has chosen to do or operate;
- Plan (l): Plans are sequences of actions that a behavior subject can perform to achieve one or more of its intentions;
- Impact (f): The results led by the execution of a behavior on its object or context;
- Constraint (c): Constraint represents what conditions are taken on the behavior; constraints are instantiated into specific factors in a domain;
- Time (t): When the behavior occurs;
- Place (w): Where the behavior happens;

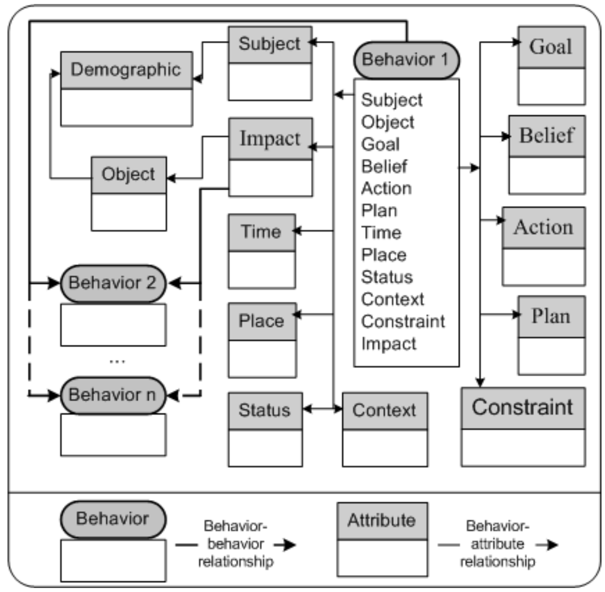

Figure 1. An Abstract Behavioral Model

- Status (u): The stage where a behavior is currently located; for instance, status may refer to *passive* (not triggered), *active* (triggered but not finished yet) or *done* (finished); in some other cases, status may include *valid* or *invalid*;
- Associate (m): Other behavior instances or sequences of actions that are associated with the target one; behavior associates exist that maybe because a behavior has impact on another or behaviors are related through interaction and business process to form a behavior network.

Figure 1 illustrates an abstract behavioral model with above attributes and properties. The model also shows the following features of behavior.

- Demographics of behavioral subjects and objects are closely related to a behavior;
- Associates of a behavior may form into certain behavior sequences or network;
- Social behavioral network consists of sequences of behaviors that are organized in terms of certain social relationships or norms.

The above abstract behavioral model aims to capture the major features as a generic behavior. In deploying this abstract model into describing behaviors in different domains, some of those attributes and properties may not take place, while for some other situations some of them may be embodied into specific features or multiple variables.

In modeling behavior to build behavioral data, behavior can be represented in terms of vector sequence. A generic behavior instance (γ) can be represented in a behavior vector $\vec{\gamma}$ as follows.

$$\vec{\gamma} = \{s, o, e, g, b, a, l, f, c, t, w, u, m\} \quad (1)$$

As we discussed before in defining behavior, the behavior vector is heterogeneous, which may consist of categorical and numerical data.

In this behavior vector, it only consists of basic properties of a behavior instance such as time and place, but also social and organizational factors including context, constraints, and impact.

Further, the behavior sequence Γ of a customer can be represented in terms of a vector sequence $\vec{\Gamma}$, which consists of all behavior instances represented in vectors.

$$\vec{\Gamma} = \{\vec{\gamma_1}, \vec{\gamma_2}, ..., \vec{\gamma_n}\} \quad (2)$$

With the vector-based behavior sequences, further analysis on such vectors can identify vector-oriented patterns. Compared to traditional sequential pattern mining, such vector-oriented behavior pattern analysis is much more comprehensive. It basically includes the following categories of information.

- Behavior performer: Information about the subject who conducts the behavior, such as *subject* (s), *action* (a) and the *time* (t), *place* (w) of the action;
- Social information: Information about the behavior's social and organizational factors, such as *object* (o), *context* (e), *constraints* (c), and *associates* (m);
- Intentional information: information about the subject's intention on the behavior, such as *goal* (g), *belief* (b) and *plan* (l);
- Behavior performance: Information about the outcomes of behavior, such as behavior *impact* (f) and subject *status* (u).

To mine for patterns in such complex data structure, it is not possible for the existing data mining techniques to be deployed directly.

3. Concept of Behavior Informatics and Analytics

3.1. What Is Behavior Informatics and Analytics

Behavior Informatics and Analytics is a scientific field, which aims to develop methodologies, techniques and practical tools for representing, modeling, analyzing, understanding and/or utilizing symbolic and/or mapped behavior, behavioral interaction and network, behavioral patterns, behavioral impacts, the formation of behavior-oriented groups and collective intelligence, and behavioral intelligence emergence.

In more detail, BIA addresses the following key aspects.

- Behavioral data: In preparing behavioral data, behavioral elements hidden or dispersed in transactional data need to be extracted and connected, and further converted and mapped into a behavior-oriented feature space, or called *behavioral feature space*. In the behavioral feature space, behavioral elements are presented into behavioral itemsets. Figure 2 illustrates the mapping and conversion from transactional data to behavioral data.
- Behavioral representation: or called behavioral modeling, is to develop behavior-oriented specifications for describing behavioral elements and the relationships amongst the elements. The specifications reshape the behavioral elements to suit the presentation and construction of behavioral sequences. Behavioral modeling also provides a unified mechanism for describing and presenting behavioral elements, behavioral impact and patterns.
- Behavioral impact analysis: In analyzing behavioral data, we are particularly interested in those behavioral instances that are associated with high impact on business processes and/or outcomes. Behavioral impact analysis features the modeling of behavioral impact.
- Behavioral pattern analysis: There are in general two ways of behavioral pattern analysis. One is to discover behavioral patterns without the consideration of behavioral impact, the other is to analyze the relationships between behavior sequences and particular types of impact.
- Behavioral intelligence emergence: To understand behavioral impact and patterns, it is important to scrutinize behavioral occurrences, evolution and life cycles, as well as the impact of particular behavioral rules and patterns on behavioral evolution and intelligence emergence. An important task in behavioral modeling is to define and model behavioral rules, protocols and relationships, and their impact on behavioral evolution and intelligence emergence.
- Behavioral network: Multiple sources of behavior may form into certain behavioral network. Particular human behavior is normally embedded into such a network to fulfill its roles and effects in a particular situation. Behavioral network analysis is to understand the

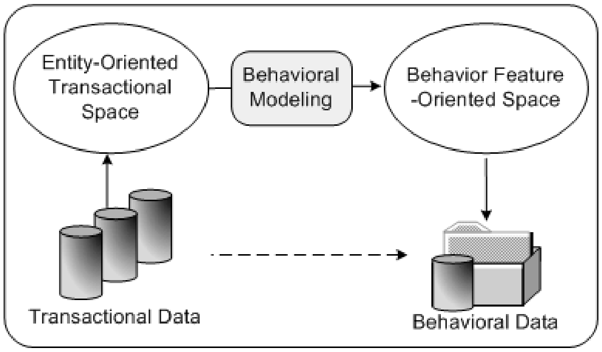

Figure 2. From Transactional Data to Behavioral Data

intrinsic mechanisms inside a network, for instance, behavioral rules, interaction protocols, convergence and divergence of associated behavioral itemsets, as well as their effects such as network topological structures, linkage relationships, and impact dynamics.

- Behavioral simulation: To understand all of the above mechanisms that may exist in behavioral data, simulation can play important roles to observe the dynamics, the impact of rules/protocols/patterns, behavioral intelligence emergence, and the formation and dynamics of social behavioral network.
- Behavioral presentation: From analytical and business intelligence perspectives, behavioral presentation is to explore presentation means and tools that can effectively describe the motivation and the interest of stakeholders on the particular behavioral data. Besides the traditional presentation of patterns such as associations, visual behavioral presentation is a major research topic. It is of high interest to analyze behavioral patterns in a visual manner.

In essence, behavior informatics and analytics is to deliver technologies and tools for understanding behavior and social behavior network. In this sense, we also call it *behavioral computing*.

3.2. What is the General BIA Process

With the BIA components discussed in the above, we can sketch a generic process of BIA as shown in Figure 3. From the concept perspective, BIA is a process converting entity relationships (DB) oriented transactional data (Ψ) to behavior feature-oriented data ($\vec{\Gamma}$) through behavior modeling ($\Theta(\vec{\Gamma})$), analyzing behavior patterns ($P(\vec{\Gamma})$) and impacts ($I(\vec{\Gamma})$) in terms of developing behavior pattern mining methods (Ω), presenting behavior patterns ($V(\vec{\Gamma})$), and transforming into decision-support business rules ($\widetilde{R}$). The outcomes of BIA consist of behavioral patterns $\widetilde{P}$ and further corresponding business rules $\widetilde{R}$ for business decision-making.

$$BIA : \Psi(DB) \xrightarrow{\Theta(\vec{\Gamma})} \vec{\Gamma} \xrightarrow{\Omega,e,c,t_i()} \widetilde{P} \xrightarrow{\Lambda,e,c,b_i()} \widetilde{R} \qquad (3)$$

Following the principle of actionable knowledge discovery [3], this process can be further decomposed and modeled in terms of the following steps.

BIA PROCESS: The Process of Behavior Informatics and Analytics
INPUT: original dataset Ψ;
OUTPUT: behavior patterns $\widetilde{P}$ and operationalizable business rules $\widetilde{R}$;
Step 1: Behavior modeling $\Theta(\vec{\Gamma})$;
  Given dataset Ψ;
  Develop behavior modeling method θ ($\theta \in \Theta$) with technical interestingness $t_i()$;
  Employ method θ on the dataset Ψ;
  Construct behavior vector set $\vec{\Gamma}$;
Step 2: Converting to behavioral data $\Phi(\vec{\Gamma})$;
  Given behavior modeling method θ;
  FOR $j = 1$ to ($count(\Psi)$)
    Deploy behavior modeling method θ on dataset Ψ;
    Construct behavior vector $\vec{\gamma}$;
  ENDFOR
  Construct behavior dataset $\Phi(\vec{\Gamma})$;
Step 3: Analyzing behavioral patterns $P\vec{\Gamma}$;
  Given behavior data ($\Phi(\vec{\Gamma})$;
  Design pattern mining method $\omega \in \Omega$;
  Employ the method ω on dataset $\Phi\vec{\Gamma}$;
  Extract behavior pattern set $\widetilde{P}$;
Step 4: Converting behavior patterns $\widetilde{P}$ to operationalizable business rules $\widetilde{R}$;
  Given behavior pattern set $\widetilde{P}$;
  Develop behavior modeling method Λ;
  Involve business interestingness $b_i()$ and constraints c in the environment e;
  Generate business rules $\widetilde{R}$;

3.3. Why Behavior Informatics and Analytics

In understanding and solving many issues and problems, *behavior* emerges as a key component, in both artificial societies (such as computerized business-support systems) and human societies. Behavior connects to many entities and objects in businesses, such as business objects, behavior subjects and objects, causes, impacts, scenarios and con-

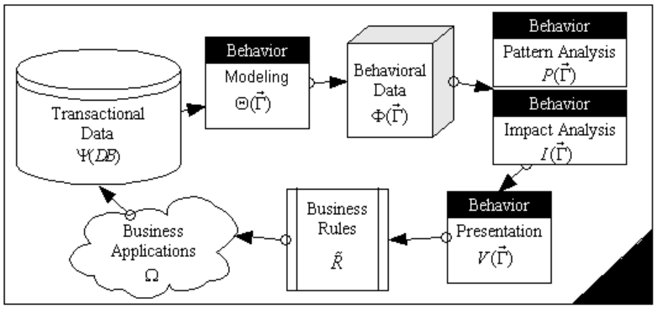

Figure 3. BIA Process

straints. In addition, multiple relevant behavior instances consist of social behavior network, which involves social and organizational factors, and collective intelligence. Therefore, it is highly likely that behavior-oriented analysis can provide extra information, in particular interior principles, causes and impact about the formation and movement of exterior business objects and appearances.

In current business management information systems, the above behavior-related factors are normally hidden in transactional data. Transactional data is usually entity-oriented, entities are connected through inter or intra-keys, which form a *transactional entity space*. In such transactional entity spaces, behavioral elements are dispersed and hidden in multiple transactions with weak or no direct linkages. Therefore, in general, behavior is *implicit* in transactional data. It is not effective to straightforwardly analyze human behavior interior on transactional data. To effectively analyze genuine behavior patterns, it is essentially important to squeeze out behavioral elements from transactions, and make behavior *explicit* for further behavior-oriented pattern analysis.

As addressed above, the presentation of behavioral data differentiates from that of normal transactional data. To effectively understand and analyze behavior and its impact, it is essentially important to squeeze out behavioral elements from transactions, and to map behavior-oriented elements in transactional data into a behavior-oriented feature space to form the behavioral data. Such extrusion and transformation from transactional space to behavioral space makes behavior shift from implicit to explicit for more effective analysis of behavior patterns and impacts.

To support the mapping from transactional space to behavioral space, it is vitally important to build formal methods and workable tools for behavior representation, processing and engineering, namely the sciences of behavior informatics and analytics.

With the development of foundations and technical tools for BIA, it is possible for us to understand and scrutinize the business processes, problems and potential solutions from a perspective different from traditional ones, namely target behavior and behavioral network perspective. In fact, due to the intrinsic integration of behavior and its subjects and objects, the in-depth understanding of behavior can actually promote a much deeper understanding of the roles and effects of comprehensive factors surrounding a business problem, for instance, human being's demographics, human actions, environment and behavioral impact. With such a capability, BIA likely further expand the opportunities of problem-solving, and stimulate promising prospects.

With BIA, and the complementation with classic analytical methods, it is possible to more effectively understand, model, represent, analyze and utilize behavior and social behavior network toward more comprehensive and effective problem understanding and solving. This includes but is not limited to behavior understanding, exceptional behavior analysis, opportunities use, behavior pattern analysis, behavior impact analysis, and cause-effect analysis.

4. Theoretical Underpinnings

Behavior Informatics and Analytics is a multidisciplinary research field. Its theoretical underpinnings involve analytical, computational and social sciences as shown in Figure 4. We interpret the theoretical infrastructure for BIA from the following perspectives: (1) Methodological support, (2) Fundamental technologies, and (3) Supporting techniques and tools.

From the methodological support perspective, BIA needs support from multiple fields, including information sciences, intelligence sciences, system sciences, cognitive sciences, psychology, social sciences and sciences of complexities. Information and intelligence sciences provide support for intelligent information processing and systems. System sciences furnish methodologies and techniques for behavior and behavioral network modeling and system simulation, and large scale of behavior network. Cognitive sciences incorporate principles and methods for understanding human behavior belief, intention and goal of human behavior. Psychology can play important roles in understanding human behavior motivation and evolution. Social sciences supply foundations for conceiving organizational and social factors and business processes surrounding behavior and embedded in behavior network. Areas such as economics and finance are also important for understanding and measuring behavior impact. Methodologies from the science of complexities is essential for group behavior formation and evolution, behavior self-organization, convergence and divergence, and behavior intelligence emergence.

Fundamental technologies are necessary for behavioral modeling, pattern analysis, impact analysis, and behavior simulation. To support behavior modeling, technologies such as user modeling, formal methods, logics, representation, ontological engineering, semantic web, group forma-

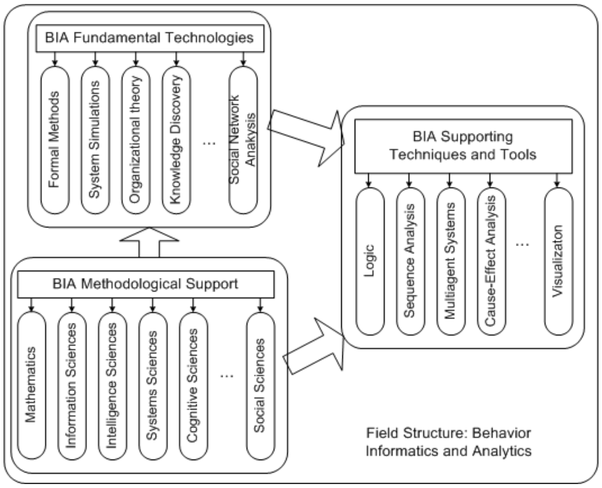

Figure 4. Field Structure of Behavior Informatics and Analytics

tion and cognitive science are essentially important. They can not only represent behavioral elements, but also contribute to the mapping from transactional entity space to behavioral feature space. The modeling of behavior impact needs to refer to technologies in areas such as risk management and analysis, organizational theory, sociology, psychology, economics and finance. For the analysis of behavioral patterns, technologies such as data mining and knowledge discovery, artificial intelligence and machine learning can contribute a lot. In simulating behavior, behavioral impact and behavior network, we refer to techniques and tools in fields like system simulation, artificial social system, open complex systems, swarm intelligence, social network analysis, reasoning and learning. The presentation of behavior evolution and behavior patterns can benefit from areas of visualization and graph theory.

From the operationalization aspect, BIA needs to develop effective techniques and tools for representing, modeling, analyzing, understanding and/or utilizing behavior. This involves many specific approaches and means. For instance, several methods such as algebra and logics may be useful for modeling behavior. The behavior pattern analysis may involve many existing tools such as classification and sequence analysis, as well as the development of new approaches. To simulate behavior impact, one may use agent-based methods for cause-effect analysis, while for presenting behavior, visualization techniques may be useful.

5. Research Issues

As BIA is at its beginning stage, many open issues are worthwhile systematic investigation and case studies from aspects such as *behavioral data*, *behavior modeling and representation*, *behavioral impact analysis*, *behavioral pattern analysis*, *behavior presentation*, and *behavior simulation*. We further expand them by listing some key research topics for each of the above research issues, certainly there may be other issues.

(1) *Behavioral Data*: In many cases, it may be necessary to convert normal transactional data into behavior-oriented feature space, in which behavior elements consist of the major proportion of the dataset.

- Behavioral data modeling
- Behavioral feature space
- Mapping from transactional to behavioral data
- Behavioral data processing
- Behavioral data transformation

(2) *Behavior Modeling*: This is to build behavior models, and to understand interaction, convergence, divergence, selection, decision, and evolution of behavior sequences and behavior networks. For these, modeling language, specifications and tools need to be developed to understand behavior dynamics.

- Behavior model
- Behavior interaction
- Collective behavior
- Action selection
- Behavior convergence and divergence
- Behavior representation
- Behavioral language
- Behavior dynamics
- Behavioral sequencing

(3) *Behavior Pattern Analysis*: This is the major focus of BIA, namely to identify patterns in behavior sequences or behavior network. For this, we need first to understand behavior structures, semantics and dynamics in order to further explore behavior patterns. Then we investigate pattern analytical tasks such as detection, prediction and prevention through approaches like correlation analysis, linkage analysis, clustering and combined pattern mining.

- Emergent behavioral structures
- Behavior semantic relationship
- Behavior stream mining
- Dynamic behavior pattern analysis
- Dynamic behavior impact analysis

- Visual behavior pattern analysis
- Detection, prediction and prevention
- Customer behavior analysis
- Behavior tracking
- Demographic-behavioral combined pattern analysis
- Cross-source behavior analysis
- Correlation analysis
- Social networking behavior
- Linkage analysis
- Evolution and emergence
- Behavior clustering
- Behavior network analysis
- Behavior self-organization
- Exceptions and outlier mining

(4) *Behavior Simulation*: To deeply understand behavior working mechanisms, interaction amongst behavior instances, dynamics and the formation of behavior group and behavior intelligence emergence, simulation can play an essential role. For example, simulation can be conducted on large-scale behavior network, convergence and divergence, evolution and adaptation of behavior through setting up artificial and computation-oriented behavior systems.

- Large-scale behavior network
- Behavior convergence and divergence
- Behavior learning and adaptation
- Group behavior formation and evolution
- Behavior interaction and linkage
- Artificial behavior system
- Computational behavior system
- Multi-agent simulation

(5) *Behavior Impact Analysis*: Behavior with high impact on business is our major interest. To analyze the behavior impact, techniques such as impact modeling, measurements for risk, cost and trust analysis, the transfer of behavior impact under different situations, exceptional behavior impact analysis would be very helpful. The analytical results will be utilized for detection, prediction, intervention and prevention of negative behavior or for opportunity use if positive cases are identified.

- Behavior impact analysis
- Behavioral measurement
- Organizational/social impact analysis
- Risk, cost and trust analysis
- Scenario analysis
- Cause-effect analysis
- Exception/outlier analysis and use
- Impact transfer patterns
- Opportunity analysis and use
- Detection, prediction, intervention and prevention

(6) *Behavior Presentation*: The presentation of dynamics of behavior and behavior network in varying aspects would assist with the understanding of behavior lifecycle and impact delivery, for instance, rule-based behavior presentation, visualization of behavior network, and visual analysis of behavior patterns.

- Rule-based behavior presentation
- Flow visualization
- Sequence visualization
- Parallel visualization
- Dynamic group formation
- Dynamic behavior impact evolution
- Visual behavior network
- Behavior lifecycle visualization
- Temporal-spatial relationship
- Dynamic factor tuning, configuration and effect analysis
- Behavior pattern emergence visualization
- Distributed, linkage and collaborative visualization

Figure 5 further illustrates major research tasks/approaches and the relations among the above key research components. Behavioral data is extracted from behavior-relevant applications, and then converted into behavioral feature space. When the behavioral data is ready, behavior pattern analysis and impact analysis are conducted on the data. To support behavior pattern analysis and impact analysis effectively, behavior simulation and modeling can provide fundamental results about behavior dynamics and relevant businesses and tools for knowledge discovery. Besides supplying another point of view for behavior analysis, behavior presentation contributes techniques and means to describing and presenting behavior.

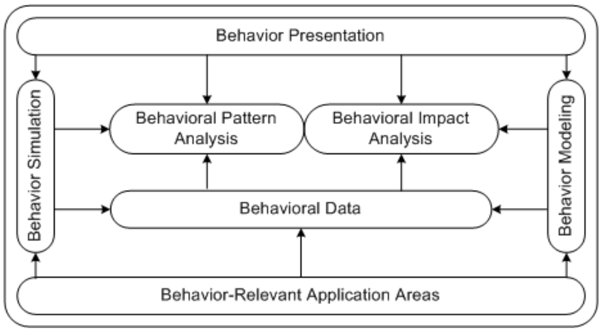

Figure 5. Research Map of Behavior Informatics and Analytics

6. Application Areas

In general, as long as there exists human actions, operations and events in the data, BIA would be very useful and presents another direction for deeply understanding the business problems and solutions. Many application areas from traditional to emergent issues can benefit from it, for instance, customer behavior in retail and online shopping businesses, web usage and interaction in the Internet, trading behavior in capital markets, and activities captured in surveillance systems. In the following, we illustrate several application areas.

(1) *Trading Behavior Analysis*: Traders in capital markets conduct sequences of actions or operations through trading systems. For instance, one may place an order, delete an order, change an order, or withdraw an order in a market. More complicated cases may include multiple traders who deal with orders distributed into multiple markets. Such sequences of actions and operations one or many markets actually reflect the belief, desire and plan of a trader. In some cases, traders' behavior discloses their exceptional intention to manipulate the market movement in order to take advantage of it. Under some other situations, such behavior is legally acceptable while creates opportunities for trading.

(2) *Customer-Officer Interaction Analysis in Social Security Areas*: In the social security area, eligible customers apply and receive governmental benefits and allowances from relevant governmental agencies as per governmental service policies. Customers therefore frequently interact with officers in relevant agencies for varying purposes. This forms a large quantity of customer-officer interaction sequences. Such interaction sequences indicate important information disclosing and resolving service quality and faulty, and understanding causes and effects. If overpayments happen on a particular customer, the customer-officer interactions can provide extra information for analyzing possible reasons, patterns of exceptional customer behavior or customer-officer interactions.

(3) *Online user behavior*: Online communities in the Internet world are getting extremely active. Online users interact with each other, computer systems (including websites, blogs, discussion boards, news rooms) for varied purposes or businesses, for instance, enjoying entertainment in the second life and doing online businesses. Analyzing such online user behavior may bring about new business opportunities. For instance, it is well known that web usage information may enclose important messages for customer caring and shopping recommendation in e-shopping. While in some other cases, the investigation of web online behavior may assist in the problem-solving of some critical issues, such as online payment fraud control that has resulted in huge amounts of revenue losses.

There are certainly many other applications and areas enclosing sequences of human behavior. For most of these problems, we cannot directly utilize the transactions available from relevant management information systems, for instance, orderbook transactions in stock markets, to analyze the behavior and its impact. Rather, behavior-oriented elements have to be extracted from the transactions or converted into some form of behavioral data, in order to further discover behavioral patterns. For this, Behavior Informatics and Analytics is the technology needed. In [6], we illustrate the use of the theory of Behavior Informatics and Analytics to analyze market microstructure behavior in stock markets. [4] illustrates the discovery of high business impact-oriented sequential activity patterns in social security area.

7. Conclusions

The effective understanding and analysis of human behaviors in businesses play essentially critical roles for disclosing interior driving forces and causes of business problems and appearance. However, in existing management information systems, behavior is implicitly hidden in transactional data, and correspondingly current so-called behavior analysis is mainly conducted on demographic and service usage data. The resulting outcomes, which focus on exterior features of business problems, cannot effectively and explicitly scrutinize human behavior patterns and impacts on businesses. In this paper, we propose the field of Behavior Informatics and Analytics (BIA) to study effective methodologies and techniques for explicit and in-depth understanding and analysis of genuine behavior-oriented actions, operations and events associated with many challenging business problems, for instance, exceptional behavior analysis of terrorists and criminals. We introduce the concept of behavior and an abstract behavior model, and a complete and clear field structure and research map of BIA. The proposed BIA techniques can actually greatly complement classic analytical approaches toward a more comprehensive and in-depth business understanding and problem-solving.

8 Acknowledgment

Thanks are given to my team members and students including Prof Chengqi Zhang, Dr Yanchang Zhao, Dr Huaifeng Zhang, Mr Yuming Ou and Ms Shanshan Wu for their comments and/or experiments on the formation and evolution of Behavior Informatics and Analytics. The work is partially sponsored by Australian Research Discovery Grant (DP0773412, DP0667060) and Linkage grant (LP0775041).

References

[1] Au, W.; Chan, K.; Yao, X. A novel evolutionary data mining algorithm with applications to churn prediction, IEEE Transactions on Evolutionary Computation, 7(6): 532 - 545, 2003.

[2] Bluhm, C., Overbeck, L., Wagner, C. An introduction to credit risk modeling. Chapman & Hall/Boca Raton, London, New York, Washington, D.C, 2003.

[3] Cao, L. Domain-Driven, Actionable Knowledge Discovery, IEEE Intelligent Systems, IEEE Computer Society, 22 (4): 78-88, 2007.

[4] Cao, L., Zhao, Y., Zhang, C. Mining Impact-Targeted Activity Patterns in Imbalanced Data, IEEE Trans. on Knowledge and Data Engineering, Vol. 20, No. 8, pp. 1053-1066, 2008

[5] Cao, L., Zhao, Y., Zhang, C., Zhang, H. Activity Mining: from Activities to Actions, International Journal of Information Technology & Decision Making, 7(2), pp. 259 - 273, 2008.

[6] Cao L., Ou, Y. Market Microstructure Patterns Powering Trading and Surveillance Agents. Journal of Universal Computer Sciences, 2008 (to appear).

[7] Chen, H., Chung, W., Xu, J., Wang, G., Qin, Y., Chau, M. Crime data mining: a general framework and some examples, Computer, 37(4): 50 - 56, 2004.

[8] Facca, F., Lanzi, P. Mining interesting knowledge from weblogs: a survey, Data & Knowledge Engineering, v.53 n.3, p.225-241, 2005.

[9] Fast, A., Friedland, L., Maier, M., Taylor, B., Jensen, D., Goldberg, H., Komoroske, J. Relational data preprocessing techniques for improved securities fraud detection, KDD'07, 2007.

[10] Fawcett, T., Provost, F. Activity monitoring: noticing interesting changes in behavior, KDD'99, 53 - 62.

[11] Hodge, V., Austin, J. A Survey of Outlier Detection Methodologies, Artificial Intelligence Review, v.22 n.2, p.85-126, 2004.

[12] Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E., Kaushanky, H. Predicting Subscriber Dissatisfaction and Improving Retention in Wireless Telecommunications Industry. IEEE Transactions on Neural Networks, 11, pp. 690-696, 2000.

[13] Srivastava, J., Cooley, R., Deshpande, M., Tan, P. Web usage mining: discovery and applications of usage patterns from Web data, ACM SIGKDD Explorations Newsletter, v.1 n.2, 2000.

[14] Stolfo, S., Hershkop, S., Hu, C., Li, W., Nimeskern, O., Wang, K. Behavior-based modeling and its application to Email analysis, ACM Transactions on Internet Technology (TOIT), 6(2), 187-221, 2006.

[15] Vigna, G., Valeur, F., Kemmerer, R. Designing and implementing a family of intrusion detection systems, ACM SIGSOFT Software Engineering Notes, 28(5), 88-97, 2003.

[16] Wang, F.Y., Carley, K.M., Zeng, D., Mao, W.J. Social Computing: From Social Informatics to Social Intelligence, IEEE Intelligent Systems, 22(2), 79-83, 2007.

[17] Weiss, G., Hirsh, H. Learning to predict rare events in event sequences, KDD-98, pp. 359-363, 1998.