Behavior Informatics: An Informatics Perspective for Behavior Studies

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

***Abstract*—Behavior is increasingly recognized as a key entity in business intelligence and problem-solving. Even though behavior analysis has been extensively investigated in social sciences and behavior sciences, in which qualitative and psychological methods have been the main means, nevertheless to conduct formal representation and deep quantitative analysis it is timely to investigate behavior from the informatics perspective. This article highlights the basic framework of *behavior informatics*, which aims to supply methodologies, approaches, means and tools for formal behavior modeling and representation, behavioral data construction, behavior impact modeling, behavior network analysis, behavior pattern analysis, behavior presentation, management and use. Behavior informatics can greatly complement existing studies in terms of providing more formal, quantitative and computable mechanisms and tools for deep understanding and use.**

***Index Terms*—Behavior, Behavior Informatics.**

I. Introduction

While behavior has been intensively studied in social sciences and behavioral sciences, the current research methodologies and approaches are derived mainly from the social and psychological aspects. Behavioral sciences [7], [11], [9] abstract empirical data to investigate the decision processes and communication strategies within and between organisms in a social system [2]. This involves fields like psychology and social neuroscience (psychiatry), and genetics among others. Qualitative analysis and experiments followed by psychological explanation and reasoning are mainly conducted on human and animal behavior.

Behavioral sciences include two broad categories [2]: neural-decision sciences and social-communication sciences. Decision sciences involve those disciplines primarily dealing with the decision processes and individual functioning used in the survival of an organism in a social environment. These include psychology, cognitive science, organization theory, psychobiology, management science, operations research (not to be confused with business administration) and social neuroscience. On the other hand, communication sciences include those fields which study the communication strategies used by organisms and the dynamics between organisms in an environment. These include fields like anthropology, organizational behavior, organization studies, sociology and social networks.

This work is sponsored in part by Australian Research Council Discovery Grants (DP0988016, DP0773412, DP0667060) and ARC Linkage Grant (LP0989721, LP0775041).

Longbing Cao is with the Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. E-mail: lbcao@it.uts.edu.au. Philip S Yu is with the Department of Computer Science, University of Illinois at Chicago. E-mail: psyu@cs.uic.edu.

With the emergence of new behavioral data, for instance, web usage, vehicle movements, market dynamics, ubiquitous transactional data recorded in computerized software systems, and agentized behavior, behavioral data including human behavior is largely electronically recorded. Behavioral sciences cannot support the formal representation and deep understanding of such behavioral data.

With the increasing needs and focus on social network analysis and social computing, it is very timely to develop behavior representation and analysis from the informatics perspective. Behavior informatics (including analytics, BI or BIA) is proposed for and aimed at the development of effective methodologies, approaches, tools and applications for formal and quantitative behavior representation and modeling, and deep analysis of behavior networks, impacts and patterns. This differentiates the aims and tasks of behavior informatics from those of behavioral sciences. This article outlines the area of behavior informatics. Behavior informatics [1] has the potential for designing and supplying new and practical mechanisms, tools and systems for deep behavior understanding and use. This will greatly complement behavioral sciences and behavior studies in social sciences. It can be widely used in many areas and domains, including understanding the Internet network, human community behavior and its evolution in the Internet, the deep understanding of human, animal, agentized and computerized organism behavior, and in widespread domains such as counter-terrorism, crime prevention, network analysis, intrusion detection, fraud and risk control, intelligent transport systems, trading agents, market dynamics, e-commerce, and financial transactions.

In fact, many researchers have started to develop deep analysis techniques for understanding behavior-related data in relevant domains. Typical examples include sequence analysis [15], event mining, crime mining, and activity mining [3], [4] and monitoring [8], as well as specific methods proposed to handle intrusion detection [13], fraud detection, outlier detection, customer relationship management [10], web usage mining [12], and so on. Behavior informatics is a scientific field consolidating these efforts and further studies on open issues toward a systematic and rigorous formalization and mechanism for behavior representation, analysis, presentation and use. With the power of behavior informatics, many traditional methods and domains can be further investigated from the behavioral perspective. In [14], facial behavioral data is analyzed, combined with facial expression information, which has shown great opportunities for expanding facial recognition capabilities and performance by considering facial behavior. [5] further reports the use of behavior informatics

in deeply analyzing microstructure-based trading behavior in stock markets, which has demonstrated very impressive advantages compared to traditional methods in understanding low-level driving forces of exceptional market dynamics. The remainder of this article is organized as follows. Section II describes what behavior informatics is. Section III argues why we need behavior informatics. The theoretical underpinnings are discussed in Section IV. Section V lists various research issues related to behavior informatics. We conclude this paper in Section VI.

II. What Is Behavior Informatics?

Behavior Informatics 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 networking, behavioral patterns, behavioral impacts, the formation of behavior-oriented groups and collective intelligence, and behavioral intelligence emergence. In more detail, behavior informatics 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 *behavioral feature space*. In the behavioral feature space, behavioral elements are presented in behavioral itemsets. Figure 1 illustrates the mapping and conversion from transactional data to behavioral data.
- Behavioral representation and modeling: The goal 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: For analyzing behavioral data, we are particularly interested in those behavioral instances that are associated with having a 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 conducting 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 (for instance, the emergence of swarm intelligence from a group of interactive agents). 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.

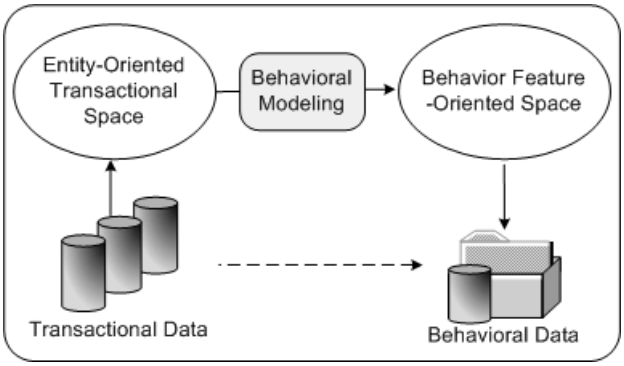

Fig. 1. From Transactional Data to Behavioral Data

- Behavioral network: Multiple sources of behavior may form into certain behavioral networks. Particular human behavior is normally embedded into such a network to fulfill its roles and effects in a particular situation. Behavioral network analysis seeks to understand the 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 the above mechanisms that may exist in behavioral data, simulation can play an important role for observing the dynamics, the impact of rules/protocols/patterns, behavioral intelligence emergence, and the formation and dynamics of social behavioral networks.
- Behavioral presentation: From analytical and business intelligence perspectives, behavioral presentation seeks to explore presentation means and tools that can effectively describe the motivation and 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, and it is of high interest to analyze behavioral patterns in a visual manner.

In essence, the purpose of Behavior Informatics is to deliver technologies and tools for understanding behavior and social behavior networks. In this sense, we also call it *behavioral computing*.

III. Why Behavior Informatics?

First of all, deep and quantitative behavior analysis cannot be supported by methodologies and techniques in traditional behavioral sciences. 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 constraints. In addition, multiple relevant behavior instances make up a 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 regarding 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, and entities are connected through 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. An example would be the trading transactions recorded in stock markets, in which an investor's trading behaviors, such as buy quote, sell quote, trade, withdrawal etc., are separately recorded into different tables, while they are actually closely related to each other. We certainly lose the full picture of an investor's overall behavior if we only look at any single aspect of them rather than putting them together. Therefore, in general, behavior is *implicit* and often *dispersed* in transactional data. It is not effective to straightforwardly analyze the interior driving force of human behavior on normal transactional data. To effectively understand such driving forces, we need to make behavior *explicit* for further behavior-oriented pattern analysis. For the example of trading behavior, if we consider the coupling relationships amongst quotes, trades, withdrawals etc. regulated by trading rules and market mechanisms, and analyze the coupled multiple behavior sequences, it is very likely that we can generate a much more informative and natural picture of trading behaviors. For this purpose, we extract quotes, trades, withdrawals etc. behavioral elements, and their properties including timepoints, prices and volumes when we detect exceptional trading behavior [5], [6].

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 a 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. Even though general data preprocessing on behavior element-oriented data is helpful, it is not effective enough nor sophisticated enough to mine such data for explicit behavior patterns and impact. Straightforward behavioral data is expected in order to cater for behavior analytics smoothly. Further, to mine for behavior and impact patterns, new issues and corresponding techniques have to be addressed.

As a result, with the development of foundations and technical tools for behavior informatics, it is possible for us to understand and scrutinize business processes, problems and potential solutions from a perspective different from the traditional ones of 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 demographics, human actions, environment and behavioral impact. With such a capability, behavior informatics is likely to further expand the opportunities of problem-solving, and stimulate promising prospects. Behavior informatics can complement classic behavioral analytical methods. This makes it possible to more effectively understand, model, represent, analyze and utilize behavior and social behavior networks toward more comprehensive and effective problem solving and understanding. This includes but is not limited to behavior understanding, exceptional behavior analysis, taking advantage of opportunities, behavior pattern analysis, behavior impact analysis, and cause-effect analysis.

IV. Theoretical Underpinnings

Behavior Informatics is a multidisciplinary research field. Its theoretical underpinnings involve analytical, computational and social sciences as shown in Figure 2. We interpret the theoretical infrastructure for behavior informatics from the following perspectives: (1) Methodological support, (2) Fundamental technologies, and (3) Supporting techniques and tools. From the methodological support perspective, behavior informatics needs to draw 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 the large scale of a behavior network. Cognitive sciences incorporate principles and methods for understanding human behavior belief, and the intention and goal of human behavior. Psychology can play an important role in understanding human behavior motivation and evolution. The social sciences supply foundations for conceiving the organizational and social factors and business processes that surround behavior and are embedded in behavior networks. Areas such as economics and finance are also important for understanding and measuring behavior impact. Methodologies from the science of complexities are 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 formation and cognitive science are essentially important. They can not only represent behavioral elements, but also contribute to the mapping from the transactional entity space to the 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,

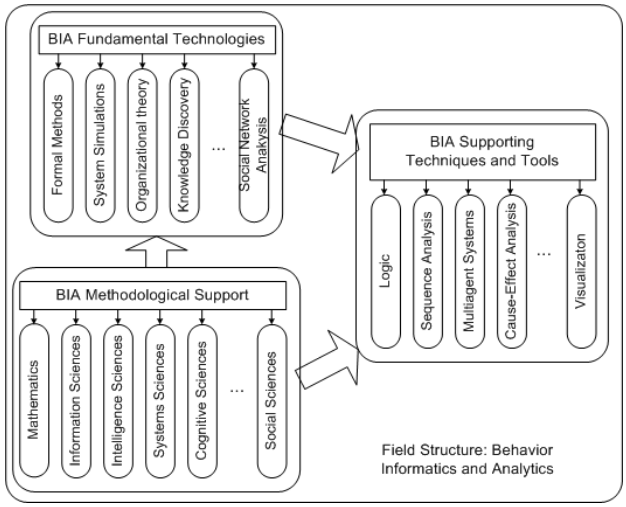

Fig. 2. Field Structure of Behavior Informatics

artificial intelligence and machine learning can contribute a great deal. In simulating behavior, behavioral impact and behavior networks, 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. In addition, the scale and complexity related to behavioral data used to be a critical issue in social science studies. We now have the ability to collect a huge amount of data in a continuous fashion (just think about Facebook). An analogy is bioinformatics. Even the simulation can produce a large amount of data to predict behavior changes over a long period of time. The studies on complex sequence analysis can provide effective tools for handling complex behavior. Adaptive and active learning offers capabilities for dealing with behavior changes in a dynamic and online environment. From the operationalization aspect, behavior informatics 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. 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.

V. Research Issues

As behavior informatics is at its beginning stage, many open issues are worthy of 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 these by listing some key research topics for each of the above research issues, although certainly there may be other issues.

(1) *Behavioral Data*: In many cases, it may be necessary to convert normal transactional data into a 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*: The building of behavior models will enable the understanding of interaction, convergence, divergence, selection, decision, and evolution of behavior sequences and behavior networks. To achieve this, 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 behavior informatics, namely to identify patterns in behavior sequences or behavior networks. For this, we need first to understand behavior structures, semantics and dynamics in order to further explore behavior patterns. We then need toinvestigate 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*: Simulation can play an essential role in the deep understanding of behavior working mechanisms, interaction amongst behavior instances, dynamics and the formation of behavior group and behavior intelligence emergence, etc. For example, simulation

can be conducted on large-scale behavior networks, 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 that has a 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 networks 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 3 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 study behavior.

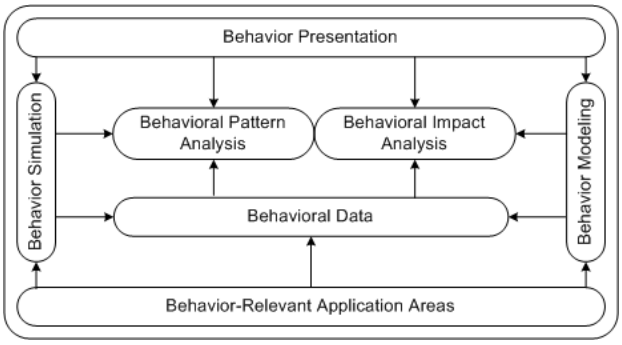

Fig. 3. Research Map of Behavior Informatics

VI. Conclusion

Behavioral sciences mainly explore the activities of and interactions among humans and animals in the natural world. For this study, qualitative, empirical, experimental and psychological methodologies and tools are generally used. With the increasing emergence of computerized and agentized behavioral data, behavioral sciences do not provide such methodologies, methods and means for formal representation and reasoning, or deep and quantitative analysis of behavior networks, impacts and patterns, from either individual or group perspectives. For this purpose, behavior informatics is proposed. Behavior informatics is essential for dealing with many behavior-related problems crossing widespread domains and areas. Typical driving forces come from Internet networks and activities, financial market dynamics, e-commerce and online businesses, human community activities and interactions, and customer relationship management.

This article highlights the framework of behavior informatics, explaining its main concepts, driving forces, theoretical underpinnings, and research issues. As a new and promising field, great efforts are expected to follow on every aspect, from formal modeling, pattern analysis, impact analysis, network analysis, behavior presentation, to behavior management and use, from fundamental, technical and practical perspectives.

References

[1] www.behaviorinformatics.org, www.behavioranalytics.org.
[2] http://en.wikipedia.org/wiki/behavioral_sciences
[3] Cao, L., Zhao, Y. and 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
[4] Cao, L., Zhao, Y., Zhang, C. and Zhang, H. Activity mining: from activities to actions, International Journal of Information Technology & Decision Making, 7(2), pp. 259 - 273, 2008.
[5] Cao, L. and Ou, Y. Market microstructure patterns powering trading and surveillance agents. Journal of Universal Computer Sciences, 2008 (to appear).
[6] Cao, L., Ou, Y., Luo, D. and Zhang, C. Adaptive detection of abnormal behavior in multiple trading activity streams, Technical Report, 2008.

[7] Devereux, G. From anxiety to method in the behaviorial sciences, The Hague, Mouton & Co, 1967.
[8] Fawcett, T. and Provost, F. Activity monitoring: noticing interesting changes in behavior, KDD'99, 53 - 62.
[9] Klemke, E., Hollinger, R. and Kline, A. (eds) Introductory readings in the philosophy of science. Prometheus Books, New York, 1980.
[10] Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E. and Kaushanky, H. Predicting subscriber dissatisfaction and improving retention in wireless telecommunications industry. IEEE Transactions on Neural Networks, 11, pp. 690-696, 2000.
[11] Smelser, N. and Baltes, P. (eds) International encyclopedia of the social & behavioral sciences, 26 v. Oxford, Elsevier, 2001.
[12] Srivastava, J., Cooley, R., Deshpande, M. and Tan, P. Web usage mining: discovery and applications of usage patterns from Web data, ACM SIGKDD Explorations Newsletter, v.1 n.2, 2000.
[13] Stolfo, S., Hershkop, S., Hu, C., Li, W., Nimeskern, O. and Wang, K. Behavior-based modeling and its application to Email analysis, ACM Transactions on Internet Technology (TOIT), 6(2), 187-221, 2006.
[14] Tsai, P., Cao, L., Hintz, T. and Jan, T. A bi-modal face recognition framework integrating facial expression with facial appearance, Pattern Recognition Letter, 30(12): 1096-1109, 2009.
[15] Weiss, G. and Hirsh, H. Learning to predict rare events in event sequences, KDD-98, pp. 359-363, 1998.