Actionable Knowledge Discovery

Ambikavathi .V[1], Veeraiah.A[2] Prabhu.R[3]
[1]Department of Master of Computer Applications , [2,3] Department of Information Technology
P.S.R Engineering College, Sivakasi-626 140.

Abstract— The data mining process consists of a series of steps ranging from data cleaning, data selection and transformation, to pattern evaluation and visualization. One of the central problems in data mining is to make the knowledge actionable. Here the term actionable refers to the mined patterns suggest concrete and profitable actions to the decision-maker. In this paper, we present a formal view of actionable knowledge discovery (AKD) from the system and decision-making perspectives. AKD is a closed optimization problem- solving process from problem definition, framework/model design to actionable pattern discovery, and is designed to deliver operable business rules that can be seamlessly associated or integrated with business processes and systems. To support such processes, we correspondingly propose, formalize, and illustrate Multisource Combined-Mining-based AKD (MSCM-AKD). A real-life case study of MSCM-based AKD is demonstrated to extract debt prevention patterns from social security data. Substantial experiments show that the proposed frameworks are sufficiently general, flexible, and practical to tackle many complex problems and applications by extracting actionable deliverables for instant decision-making.

Keywords— Data Mining, Domain-Driven Data Mining, Actionable Knowledge Discovery, Decision Making.

I. INTRODUCTION

Data mining, as well as its synonyms knowledge discovery and information extraction, is frequently referred to the literature as the process of extracting interesting information or patterns from large databases. There are two major issues in data mining research and applications: patterns and interest. The identified patterns are then handed over business people for further employments. The business people cannot effectively take over and interpret the identified patterns for business use. This may result from several aspects of challenges besides the dynamic environment enclosing constraints. Many patterns mined but they are not informative and transparent to business people who do not know which are truly interesting and operable for their businesses. Business people feel confused by why and how they should care about the identified patterns may be either commonsense or of no particular interest to business needs. Business people often do not know, and are also not informed, how to interpret them and what straightforward actions can be taken on them to support business decision-making and operation.

Stating the AKD problem from system and micro- economy perspectives to define fundamental concepts of action ability and actionable patterns. Defining knowledge action ability by highlighting both technical significance and business expectations that need to be considered, balanced, and/or aggregated in AKD, proposing four general frameworks to facilitate AKD, and demonstrating the effectiveness and flexibility of the proposed frameworks in tackling real-life AKD.

The main idea of MSCM-AKD [3] such as handles AKD in either multiple data sources or large quantities of data, one of the data sets is selected for mining initial patterns, some learned patterns are then selected to guide feature construction and pattern mining on the next data set(s). The iterative mining stops when all data sets are mined, and the corresponding patterns are then merged/summarized into actionable deliverables

II. KNOWLEDGE MANAGEMENT

The knowledge management [5] is becoming a hotspot of the management research. Explicit knowledge is systematic and easy to communicate and share in the form of leaflet, scientific formula and software. Tacit knowledge is highly personal and difficult to formulize and communicate, it runs deeply in the personal action and understanding of the surrounding in the form of workmanship, special skill, product market and teamwork, including the informal technology in “know how”.

The knowledge management is the promotion of the flexible and creative competence with the team wisdom and a new method to share the explicit and tacit knowledge for the company.” The knowledge management is one kind of science to identify, obtain, evaluate, retrieve and share all of the company information through the promotion of the integration.” The development of the knowledge management must be faster with the more violent market competition.

The research of the basic tasks of the knowledge management can strengthen our understanding and promote the method and mode of the knowledge management for companies and organizations. The two basic tasks that are mentioned in the text are only one fundamental question of the knowledge management. Here it didn't go further how to complete the basic tasks and how to fulfill the other tasks of the knowledge sharing and management. However, it stays the necessary need and pressing task of the knowledge management to find the operable and systematic method of the knowledge management.

From the knowledge chain model [6], three aspects should be understood to grasp knowledge management: 1) The management is a process for companies to administer their owned knowledge resources, and the core object is knowledge. 2) Operated by the links of accumulation, sharing, learning, application and innovation, knowledge can guide enterprise's actions and continuously generate values. 3) Culture, management and technology are the three key elements in knowledge management.

The characteristics of KMS are knowledge management is the extension and development of information management. KMS and traditional management information system has the following differences: 1. The target of collection, processing and dissemination is different. The traditional management information system is for collecting, processing and dissemination the information that reflects the real world objects' attribute. But the target, what knowledge management systems have to collect, disseminate and dispose, is the knowledge that carried by human brain, and has implied characteristic. 2. Target object has different processing depth. Traditional management information systems treat the information most on the surface processing, such as calculation, merger, compilation, linking, etc. 3. Products are in different forms. Traditional management information system's output is usually in form of reports, documents, statements or other summary data. But knowledge management system's output is varied, which is multimedia output integrated text, graphics, audio and video. 4. The value orientation of system's product is different. The information products of the traditional management information system have higher demand of timeliness, accuracy. And the knowledge products provided by knowledge management system emphasize innovation, scientific, experience and skills. 5. Different measurement. The traditional management information system takes the computer hardware, software, network connectivity and enterprises' income as the measurement. But knowledge management system develops the knowledge investment, knowledge-intensive and the breadth and depth in excavating tacit knowledge as measurement.

III. RELATED WORK

Data mining and knowledge discovery has emerged to be one of the most vivacious areas in information technology during the last decade. It has boosted a major academic and industrial campaign crossing many traditional areas such as database, statistics, business as well as emergent disciplines.

In [1] have explained that action hierarchy which is defined as a tree of actions with patterns and pattern templates (data mining queries) assigned to its nodes. A method for discovering actionable patterns is presented and various techniques for optimizing the discovery process are proposed. To define actionability as a measure of interestingness of patterns based on the concept of action hierarchy.

In [2] presented that most data mining algorithms is to facilitate the discovery of concise and interpretable information from large amounts of data. However, many of the current formalizations of data mining algorithms have not quite reached this goal. One of the reasons for this is that the focus on using purely automated techniques has imposed several constraints on data mining algorithms.

In [3] have proposed that Domain-driven KDD represents a paradigm shift from a research-centered discipline to a practical tool for actionable knowledge. Despite many open issues, deployed systems are already showing ways to transmit reliable research in forms that satisfy business needs with direct support for decisions. It represents a paradigm shift from a research-centered discipline to a practical tool for actionable knowledge.

In [4] presented the current data mining algorithms and tools often stop at the delivery of patterns satisfying expected technical interestingness. Business people are not informed about how and what to do to take over the technical deliverables. D3M aims to construct next-generation methodologies, techniques and tools for a possible paradigm shift from data-centered hidden pattern mining to domain-driven actionable knowledge delivery.

Actionable Knowledge Discovery Applications [7] are Customer Relationship Management, Supplier Selection, Crime Identification and Business Intelligence.

Customer Relationship Management (CRM) consists of four dimensions: Customer identification, customer attraction, customer retention, and customer development. Customer satisfaction is the central concern for customer retention. Supplier selection is one of the most parts in supply chain management. Strategic partnership with better performing suppliers should be integrated into the manufacturing to improve the performance in various aspects including reducing costs by eliminating wastes, continuously improving quality to achieve zero defects, reducing lead time at different

stages of the manufacturing. Crime detection for credit applications is so popular in bank industry. Phua et al. [8] presented an updated adaptive Communal Analysis Suspicion Scoring (CASS) algorithm. CASS adaptively changes the appropriate parameter setting to trade off efficiency and effectiveness. Their approach is validated with three sets of experiments on real credit applications. Business Intelligence (BI) refers to skills, processes, technologies, applications and practices used to support decision-making. BI technologies provide historical, current, and predictive views of business operations. The approach uses domain knowledge to filter indicators, and enables incremental adjustment of underlying domain model thorough involving domain knowledge.

IV. PROPOSED METHODOLOGY

MSCM-AKD Enterprise applications often involve multiple-subsystems-based and heterogeneous data sources that cannot be integrated, or are too costly to do so. Another common situation is that the data volume is so large that it is too costly to scan the whole data set. Mining such complex and large volumes of data challenges existing data mining approaches

The MSCM-AKD framework can also be instantiated into a number of mutations. For instance, for a large volume of data, MSCD-AKD can be instantiated into data partition + unsupervised + supervised-based AKD by integrating data partition into combined mining. An example is as follows: First, the whole data set is partitioned into several data subsets based on the data/business understanding and domain knowledge jointly by data miners and domain experts, say data sets 1 and 2. Second, an unsupervised learning method is used to mine one of the preference data sets, say data set 1. Supervised learning is further conducted on data set 2 to generate actionable patterns by checking both technical and business interestingness. Finally, the individual patterns mined from both data subsets are combined into deliverables.

Table 1. MSCM-AKD Algorithm

INPUT: Target Data Sets DB, Business Problem Ψ, and Thresholds ($t_{o,0}$, $t_{s,0}$, $b_{o,0}$ and $b_{s,0}$)

OUTPUT: Actionable Patterns $\tilde{P}$ and business rules $\tilde{R}$.

Step 1: Identify or partition whole source data into N data sets DB_n (n = 1,..., N);

Step 2: *Data Set-n mining*: Extracting general patterns P_n on data set/subset DB_n;

FOR l = n to (N)

Develop modeling method m_n with technical interestingness $t_{i,n}()$ (i.e., $t_o()$, $t_b()$) or unified $i_{i,n}()$ Employ method m_n on the environment e and data DB_n engaging meta-knowledge Ω_m;

Extract the general pattern set P_n;

ENDFOR

Step 3: *Pattern merger*: Extracting actionable patterns $\tilde{P}$;

FOR l = n to N

Design the pattern merger functions $\oplus^N P_n$ to merge all patterns into $\tilde{P}$ by involving domain and meta knowledge Ω_d and Ω_m, and business interestingness $b_i()$;

Employ the method $\oplus P_n$ on the pattern set P_n;

Extract the actionable pattern set $\tilde{P}$;

ENDFOR

Step 4: Converting patterns $\tilde{P}$ to business rules $\tilde{R}$.

Finally, combined patterns can be transformed into operable business rules that may indicate direct actions for business decision-making. For instance, for the above combined association, it actually connects key business elements with segmented customer characteristics, and we can generate the following business rule by extending the Business Rule specification.

Table 2. Delivering Business Rules

For All customer i ($i \in I$ is the number of valid customers)
Condition:
satisfies S/he is a debtor aged 65 or plus;
relates
S/he is under arrangement of "withholding" and
"irregularly",
and
His/her favorite Repayment method is "withholding";
Operation:
Alert = "S/he has 'High' risk of paying off debt in a very long timeframe."
Action = "Try other arrangements and repayments in R_2, such as trying to persuade her/him to repay under 'irregular' arrangement with 'cash or post.'"
End-All

CONCLUSION

This framework is supporting the controlled data of employee's regarding withheld and payment process and it is very effective data from social welfare. The framework includes, collect and maintain the organizational and non-organizational data regarding their salary. Proposed system can be hardly applied in real cases if the service provider cannot protect the data it has acquired from competitors and selects the trusted parties from which it wants to receive information.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Discovery of Actionable Patterns in Databases: The Action Hierarchy Approach," Proc. Int'l Conf. Knowledge Discovery and Data Mining (KDD '97), pp. 111-114, 1997.

[2] C. Aggarwal, "Towards Effective and Interpretable Data Mining by Visual Interaction," ACM SIGKDD Explorations Newsletter, vol. 3, no. 2, pp. 11-22, 2002.

[3] L. Cao, "Domain-Driven Actionable Knowledge Discovery," IEEE Intelligent Systems, vol. 22, no. 4, pp. 78-89, July/Aug. 2007.

[4] L. Cao, "Domain-Driven Data Mining: Empowering Actionable Knowledge Delivery," Proc. Pacific-Asia Conference, 2008.

[5] J.F. Boulicaut and B. Jeudy, "Constraint-Based Data Mining," The Data Mining and Knowledge Discovery Handbook, pp. 399-416, Springer, 2005.

[6] S. Yoon, L. Henschen, E. Park, and S. Makki, "Using Domain Knowledge in Knowledge Discovery," Proc. Eighth Int'l Conf. Information and Knowledge Management, pp. 243-250, 1999.

[7] Jiying Li and Jianwu Dang "A Survey on Actionable Knowledge Discovery Applications," IEEE Conference, 2010.

[8] C. Phua, V. Lee, K. S. Miles and R. Gayler, "Adaptive communal detection in search of adversarial identity crime," Proceedings of the 2007 International Workshop on Domain Driven Data Mining, pp.1–10, San Jose, USA, August 2007.