ISSN No. 0976-5697

Volume 2, No. 1, Jan-Feb 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Domain Driven Data Mining (D3M) – A General Perspective

S.Maheswari*
System Programmer
Alagappa University
Karaikudi, Tamilnadu India
mahivh@gmail.com

Dr.S.S.Dhenakaran
Asst. Professor in Computer Science
Alagappa University
Karaikudi, Tamilnadu India

Abstract: Domain driven data mining (D3M or DDDM) is aimed to extract actionable and domain related pattern from the large voluminous databases. Many research issues are raised at mining of actionable pattern. This paper discussed some existing works related to DDDM and actionable pattern mining. This paper compared two existing actionable pattern mining algorithm that are mutually dependent pattern mining algorithm (MDPMA) and actionable hierarchy based actionable pattern mining algorithm (AHBPMA).

Keywords: Domain data driven mining, actionable pattern, AKD

I. INTRODUCTIONS

Domain Driven Data Mining (D3M for short) targets the development of next-generation data mining methodologies, frameworks, algorithms, evaluation systems, tools and decision support, which aim to promote the paradigm shift from data-centered hidden pattern mining to domain-driven actionable knowledge discovery (AKD). To this end, D3M needs to involve and integrate human intelligence, domain intelligence, data intelligence, network intelligence, organizational and social intelligence, and the meta-synthesis of the above ubiquitous intelligence. As a result of the D3M research and development, the AKD system can deliver business-friendly and decision-making rules and actions that are of solid technical and business significance.

In data mining community, there is a big gap between academic objectives and business goals, and between academic outputs and business expectations. However, this runs in the opposite direction of KDD's original intention and its nature. It is also against the value of KDD as a discipline, which generates the power of enabling smart businesses and developing business intelligence for smart decisions in production and living environment. From both macro-level and micro-level, we can find reasons asking for new methodology and paradigm shift such as domain driven data mining. On the macro-level, issues related to methodological and fundamental aspects include An intrinsic difference existing in academic thinking and business deliverable expectation; for example, researchers usually are interested in innovative pattern types, while practitioners care about getting a problem solved; The paradigm of KDD, whether as a hidden pattern mining process centered by data, or an AKD-based problem-solving system; the latter emphasizes not only innovation but also impact of KDD deliverables. The micro-level issues are more related to technical and engineering aspects, for instance, If KDD is an AKD-based problem-solving system, we then need to care about many issues such as system dynamics, system environment, and interaction in a system; If AKD is the target, we then have to cater for real world aspects such as business processes, organizational factors, and constraints.

II. RESEARCH ISSUES

To effectively synthesize the above ubiquitous intelligence in AKD-based problem-solving systems, many research issues need to be studied or revisited.

- Typical research issues and techniques in *Data Intelligence* include mining in-depth data patterns, and mining structured knowledge in unstructured data.
- Typical research issues and techniques in *Domain Intelligence* consist of representation, modelling and involvement of domain knowledge, constraints, organizational factors, and business interestingness.
- Typical research issues and techniques in *Network Intelligence* include information retrieval, text mining, web mining, semantic web, ontological engineering techniques, and web knowledge management.
- Typical research issues and techniques in *Human Intelligence* include human-machine interaction, representation and involvement of empirical and implicit knowledge.
- Typical research issues and techniques in *Social Intelligence* include collective intelligence, social network analysis, and social cognition interaction.
- Typical issues in *intelligence metasynthesis* consist of building meta synthetic interaction as working mechanism, and metasynthetic space as an AKD-based problem-solving system.
- Typical issues in actionable knowledge discovery through m-spaces consist of Mechanisms for acquiring and representing unstructured and ill-structured, uncertain knowledge such as empirical knowledge stored in domain experts' brains, such as unstructured knowledge representation and brain informatics; Mechanisms for acquiring and representing expert thinking such as imaginary thinking and creative thinking in group heuristic

discussions; Mechanisms for acquiring and representing group/collective interaction behaviour and impact emergence, such as behaviour informatics and analytics; Mechanisms for modelling learning-of-learning, i.e., learning other participants' behaviour which is the result of self-learning or ex-learning, such as learning evolution and intelligence emergence.

III. APPLICATIONS

D3M can bring about the effective and practical development of many challenging data mining applications in every area. Based on the collaborations with our business partners, we have the experience in developing and deploying *D3M* in areas such as capital markets and social security area. In capital markets, we develop actionable trading agents, actionable trading strategies, and exceptional market microstructure behaviour patterns. In social security area, we propose the concept of activity mining and combined mining.

IV. EXISTING WORKS

Adeyemi Adejuwon and Amir Mosavi [1] are applied D3M in business. Conventional data mining applications face serious difficulties in solving complex real-life business decision making problems when practically deployed. Their work in order to improve the operations in a collection of business domains aims to suggest solutions by reviewing and studying the latest methodological, technical, practical progresses and some cases studies of data mining via domain driven data mining . The presented paper tries to answer this question: "what can domain driven data mining do for real-life business applications?" Moreover this work attempts to provide information and abilities to fill the existing gap between academic researches and real-world business problems.

Thomas Piton, Julien Blanchard, Henri Briand and Fabrice Guillet [2] applied DDDM to improve promotional campaign ROI and Select Marketing Channels. The trading activities of materials retail is concerned with an extremely competitive market. However, business people are not well informed about how to proceed and what to do during marketing activities. Data mining methods could be interesting to generate substantial profits for decision makers and to optimize the choice of different marketing activities. They proposed an actionable knowledge discovery methodology, for one-to-one marketing, which allows contacting the right customer through the right communication channel. This methodology first requires a measurement of the tendency for the customers to purchase a given item, and second requires an optimization of the Return On Investment by selecting the most effective communication channels for attracting these customers. Our methodology has been applied to the VM Mat´eriaux Company. Thanks to the collaboration between data miners and decision makers, we present a domain-driven view of knowledge discovery satisfying real business needs to improve the efficiency and outcome of several promotional marketing campaigns.

Qiang Yang, Charles X. Ling and Jianfeng Gao applied the Mining Web Logs for Actionable Knowledge. Every day, popular Web sites attract millions of visitors. These visitors leave behind vast amount of Web site traversal information in the form of Web server and query logs. By analyzing these logs, it is possible to discover various kinds of knowledge, which can be applied to improve the performance of Web services. A particularly useful kind of knowledge is knowledge that can be immediately applied to the operation of the Web site; we call this type of knowledge the actionable knowledge. In this paper, they presented three examples of actionable Web log mining. The first method is to mine a Web log for Markov models that can be used for improving caching and prefetching of Web objects. A second method is to use the mined knowledge for building better, adaptive user interfaces. The new user interface can adjust as the user behavior changes with time. Finally, we present an example of applying Web query log knowledge to improving Web search for a search engine application.

J. L. Hellerstein, S. Ma, C.-S. Perng [3] are mined actionable pattern from event data. Applications such as those for systems management and intrusion detection employ an automated real-time operation system in which sensor data are collected and processed in real time. Although such a system effectively reduces the need for operation staff, it requires constructing and maintaining correlation rules. Currently, rule construction requires experts to identify problem patterns, a process that is time consuming and error-prone. They proposed reducing this burden by mining historical data that are readily available. Specifically, they presented efficient algorithms to mine three types of important patterns from historical event data: event bursts, periodic patterns, and mutually dependent patterns. They then discuss a framework for efficiently mining events that have multiple attributes. Last, they presented Event Correlation Constructor—a tool that validates and extends correlation knowledge.

V. RESULTS AND COMPARISONS

This work also compares two actionable pattern mining algorithms. The algorithms are developed by J. L. Hellerstein, S. Ma, C.-S. Perng and Gediminas Adomavicius, Alexander Tuzhilin. Both algorithms are mined actionable pattern from chess dataset. The chess dataset is downloaded from FIMI dataset. The mutually dependenet pattern mining algorithm (MDPMA) is mined the pattern based on mutually dependency between pattern. The Actionable hierarchy based pattern mining algorithm (AHBPMA) is mined pattern based on actionablility measure. The figure 1 shows the comparison of execution time between both algorithms. From this figure, this work studies MDPMA is performed better than AHBPMA.

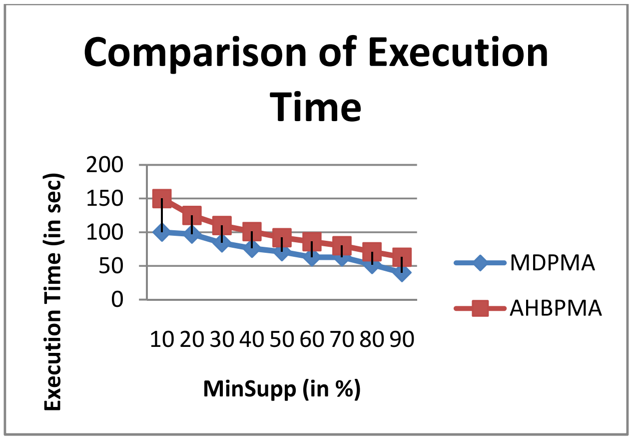

Figure 1: Comparison of execution time between MDPMA and AHBPMA

The figure 2 shows the comparison of number of frequent pattern mined by both algorithms. From this figure, this work studies MDPMA is performed better than AHBPMA. The MDPMA also reduced large set of unwanted or irrelevant pattern.

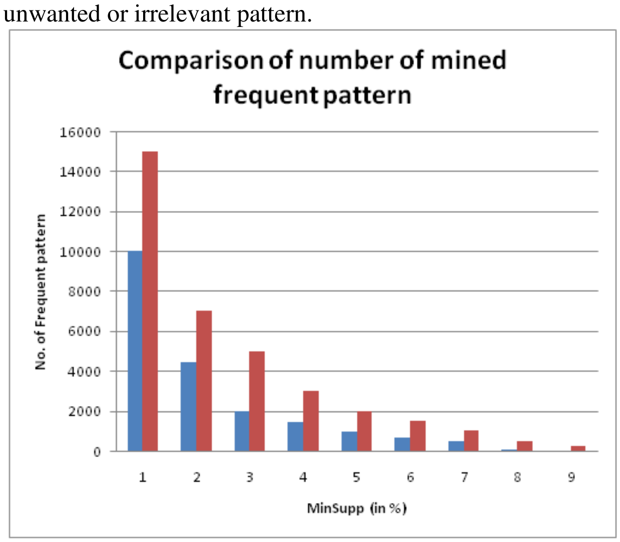

Figure 2: Comparison of number of frequent pattern mined by MDPMA and AHBPMA

VI. DISCUSSION

Actionable knowledge is particularly attractive for Web applications because they can be consumed by machines rather than human developers. Furthermore, the effectiveness of the knowledge can be immediately put to test, making the merits of the type of knowledge and methods for discovering the knowledge under more objective scrutiny than before. In this chapter, Qiang Yang, Charles X. Ling and Jianfeng Gao presented two examples of actionable Web log mining. The first method is to mine a Web log for Markov models that can be used for improving caching and prefetching of Web objects. A second method is to use the mined knowledge for building better, adaptive user interfaces. A third application is to use the mined knowledge from a query web log to improve the search performance of an Internet Search Engine. In our future work, we will further explore other types of actionable knowledge in Web applications, including the extraction of content knowledge and knowledge integration from multiple Web sites.

J. L. Hellerstein, S. Ma, C.-S. Perng described how data mining can be used to identify patterns of events that indicate underlying problems. In particular, they identify several patterns of interest to event management: event bursts, periodicities, and mutual dependencies. They provide interpretations for each, and they shown how pattern discovery can be structured to exploit a level wise search, thereby improving scalability. Scalability is particularly important because, as experience shows, tens of millions of events may need to be analyzed in order to discover actionable patterns. They also addressed the issue of related multiple attributes of events. They developed here a framework that provides a means to systematically and efficiently explore patterns with different membership definitions. Last, they presented the Event Correlation Constructor, a tool that can validate existing correlation knowledge and further construct such knowledge.

In [4], the authors mine the actionable knowledge from the viewpoint of data mining tasks and algorithms. The tasks, such as clustering, association, outliers detection etc are explained along with the actionability techniques. [5] proposed an approach to find out the best action rules. The best k-action rules are selected on the basis of maximizing the profit of moving from one decision to another. The technique used as post analysis to the rules extracted from decision tree induction algorithm. A novel algorithm is presented that suggest action to change customer status from an undesired status to desired one. In order to maximize profit based, an objective function is used to extract action rules.

Furthermore, [6] has proposed a framework to quantify novelty of the discovered rules. They show that, the novelty is a good measure for actionability. Novelty of currently discovered rules is quantified with respect to domain knowledge and previously discovered knowledge based on newly discovered knowledge.

In [7] [8], they define actionability in term of defining two types of attributes, namely, stable and flexible attributes. An action is takes when a change in flexible attribute is encountered. The approach takes into consideration the changes of attributes value as well and gives suggestion of, to which attributes value should a an attribute be changed in order to get some action. For a given two rules, there was no constraints values of flexible attributes listed in both rules or in one of these rules. The proposed constraints make use of confidence of new rules, which are called as extended action rules, much higher than the confidence of corresponding to the action rules. DEAR2 is a system which is proposed by [9] as an extension to [10] and is based on a tree structure that partitions the set of rules, having the same decision value, into equivalence classes each labelled by values of stable attributes.

VII. CONCLUSION AND FUTURE ENHANCEMENT

This paper is discussed about domain driven data mining, issues and some existing works related to above one. DDDM is an important step in next generation of data mining. Most of existing works are mined the actionable pattern in many ways. Anyways the pattern should useful to decision making to the particular domain. That is achieved by some research. This paper also motives to develop new actionable pattern mining algorithm. This paper compared two existing actionable pattern mining algorithm. In this comparison, the mutually dependent pattern mining algorithm is efficient compared with actionable hierarchy based actionable pattern mining algorithm. DDDM and actionable pattern mining scope many future research directions.

VII. REFERENCES

[1]. Adeyemi Adejuwon and Amir Mosavi, "Domain Driven Data Mining – Application to Business", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 2, July 2010, pp:41-44.

[2]. Thomas Piton, Julien Blanchard, Henri Briand and Fabrice Guillet, Domain Driven Data Mining to Improve Promotional Campaign ROI and Select Marketing Channels, Proceedings of the 18th ACM Conference on Information and Knowledge Management, Hong Kong (2009).

[3]. J. L. Hellerstein, S. Ma, C.-S. Perng, "Discovering actionable patterns in event data", IBM SYSTEMS JOURNAL, VOL 41, NO 3, 2002

[4]. He., Zenyou. & Xu., Xiaofei., & deng., Shengchum. Data Mining for Actionable Knowledge: A Survey. Supported by the High TechnologyResearch & Development Program of China (No. 2003 AA4Z2170, No. 2003 AA413021) and the IBM SUR Research Fund.

[5]. Yang, Q., & Yin, J. and Ling, X. C., & Chen, T. Postprocessing Decision Trees to Extract Actionable Knowledge . In IEEE International Conference on Data Mining (ICDM'03).

[6]. A. S. Al-Hegami, V. Bhatnagar, and N. Kumar. Novelty Framework for Knowledge Discovery in Databases. In Proceedings of 6th International Conference on Data Warehousing and Knowledge Discovery.(DaWaK 2004), Spain, 2004.

[7]. Ras, Z.W. & Wieczorkowska, A.. Action Rules : How to increase profit of a company. In Principles of Data Mining and Knowledge Discovery, (Eds. D. A. Zighed, J. Komorowski, J. Zytkow) Proceedings of PKDD'00, Lyon, France, LNCS/LNAI, No. 1910, Springer-Verlag 2000, 587-592.

[8]. Silberschatz, A. & Tuzihilin, A. 1995. On Subjective Measures of Interestingness in Knowledge Discovery. In Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining.

[9]. Tsay, L.-S., & Ras, Z. W. (2004). Action Rules Discovery System DEAR, Method and Experiments. In the Special Issue on Knowledge Discovery, (Ed. Z. W. Ras), Journal of Experimental and Theoretical Artificial Intelligence, Taylor & Francis.

[10]. Tsay, L.-S., & Ras, Z. W. (2005). Action Rules Discovery System DEAR2, Method and Experiments. In the Special Issue on Knowledge Discovery, (Ed. Z. W. Ras), Journal of Experimental and Theoretical Artificial Intelligence, Taylor & Francis. Vol. 17, No.1-2, January-June 2005, 119-128.

[11]. Gediminas Adomavicius, Alexander Tuzhilin, 1996. "Discovery of Actionable Patternsin Databases: The Action Hirarchy Approach".

