



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY
Volume 11 Issue 23 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Data Driven Data Mining to Domain Driven Data Mining

By Mitu Kumari

Kurukshetra University, Kurukshetra, Haryana, India

Abstract - In the preceding decade data mining has come into sight as one of the largely energetic areas in information technology. Traditional data mining is seriously dependent on data itself, and relies on data oriented methodologies. So, there is a universal necessity in bridging the space among academia and trade is to provide all-purpose domain-related matters in surrounding real-life applications. Domain-Driven Data Mining try to build up general principles, methodologies, and techniques for modelling and reconciling wide-ranging domain-related factors and synthesized ubiquitous intelligence adjacent problem domains with the data mining course of action, and discovering knowledge to hold up business decision-making.

Keyterms : *Data Mining, Domain driven data mining, decision-making.*

GJCST Classification : *H.2.8*



Strictly as per the compliance and regulations of:



Data Driven Data Mining to Domain Driven Data Mining

Mitu Kumari

Abstract - In the preceding decade data mining has come into sight as one of the largely energetic areas in information technology. Traditional data mining is seriously dependent on data itself, and relies on data oriented methodologies. So, there is a universal necessity in bridging the space among academia and trade is to provide all-purpose domain-related matters in surrounding real-life applications. Domain-Driven Data Mining try to build up general principles, methodologies, and techniques for modelling and reconciling wide-ranging domain-related factors and synthesized ubiquitous intelligence adjacent problem domains with the data mining course of action, and discovering knowledge to hold up business decision-making.

Keywords : Data Mining, Domain driven data mining, decision-making.

I. INTRODUCTION

In the last ten years, data mining is a field which becomes the most active, dynamic and lively area in information and communication technologies. The rapid growth of the global economy and heavy usage of computing and networking across every sector and business, results in data and its deep analysis becomes a particularly important issue for the soft control of an organization, and also equally important for the production system, decision making powers and performance of the organization. Now these days, there is a rapid increase in the applications of the data mining in various fields like business, government, social networks and the like ones. But due to the data driven data mining's limited decision support power in the real world, it hinders from playing a strategic decision support role in all these areas. In order to sort out this problem, a new approach Domain Driven Data Mining is evolved, this new approach will handle all the issues which are faced by the traditional data mining and also tackle the findings, thoughts and lessons learned in conducting several large scaled real world data mining business applications. The motivation of Domain Driven Data Mining is to study effective and efficient methodologies, techniques, tools and applications that can discover and deliver actionable knowledge that can be passed directly to the business people for the direct decision making and action taking.

If we apply current data mining algorithms and techniques on to the real world problem solving and

decision making tasks then we have to face the crucial need to lessen the differences between academic world and commerce. Also we have to tackle the space between estimation systems and real business requirements. Not only this, we also have to manage the inequilibrium between the huge number data mining algorithms existing in the market in opposition to those few data mining algorithm that are in fact deployed in problem areas and resulting in those patterns which are of real use and these patterns can be suggested for decision support actions.

Real world data mining applications have projected critical desires for discovering actionable knowledge of foremost interest to real user and business wishes. As the actionable knowledge discovery is noteworthy and also very demanding.

In order to overpass the above mentioned gaps, it is vital to boost the decision support power of data mining and knowledge findings. It is crucial to expand the actionability of the discovered patterns and to make available the results that can sustain decision making, in the right and beneficial route.

Domain driven data mining provides an efficient overview of the issues in discovering actionable knowledge and advocates the methodology of mining the actionable knowledge in constraint based context through human mining cooperation in a loop closed iterative improvement manner. It is valuable for promoting the paradigm shift from data driven hidden pattern mining to area driven actionable data discovery. Further, progress in studying domain driven data mining methodologies and applications can facilitate the deployment swing from standard and artificial data set based testing to genuine data and business atmosphere based back testing and development.

II. DATA DRIVEN DATA MINING

A distinctive feature of traditional data mining is that KDD (Knowledge Discovery From Data). One of the elementary objectives of KDD is to discover knowledge that is of key concentration to genuine business requirements and user preferences but KDD is a presumed and preset process. It targets the production of predefined and automatic algorithms and tools. As a consequence, the algorithms and tools developed have no potential to adapt to external environment constraints. Millions of patterns and algorithms have

Author : Kurukshetra University, Kurukshetra, Haryana, India.
E-mail : mitukajal81@gmail.com

been available in literature but sorry to say that a small number of them have been transferred into real business.

In data driven data mining lots of patterns are generated according to the problems but they are not enlightening and clear to business individuals. They can't straightforwardly acquire truly remarkable and operable patterns for their business. A large fraction of the indentified patterns may be either commonsense or of no particular attention to business desires. Business grassroots are puzzled as to why and how they should be concerned regarding those conclusions. Activities extracted or summarized through post investigation and post processing without in view of business concerns do not replicate the authentic expectations of business desires. Therefore they cannot bear smart decision making. Business people often don't know and also not well-versed regarding, how to understand and utilize the discovered patterns and what undemanding activities can be taken to engage those discovered patterns in business functioning systems and decision making.

Conventional KDD is a data centred and technically dominated course targeting automated hidden pattern mining. The core objective of conventional data mining research is to let data verify research innovation, track the elevated performance of the algorithms and express novel algorithms. As a consequence, the mining process stops at discovering knowledge that is primarily of importance to academic or industrial individuals.

In the real world, determining and transporting knowledge that is actionable in answering business problems has been analyzed as the fundamental nature of KDD. However, the existing data mining is principally data-centred and technically conquered, and stops at hidden pattern mining favouring technical concerns and expectation, while many other features surrounding business problems have not been thoroughly or exhaustively considered and balanced. It will be one of the great challenges to the existing and future KDD society.

A distinctive fashion in real world data mining applications is to treat a data mining system as a problem solving systems within a certain atmosphere. Looking at the problem solving from the domain driven point of view, a lot of unwrap matters and opportunities arise, demonstrating the need of next generation data mining and knowledge discovery far further than the data mining algorithms themselves. In order to sort out these troubles a new methodology is proposed i.e. Domain driven data mining. Domain driven data mining tends to create next generation methodologies, techniques and tools for a probable idea shifting from data driven hidden pattern mining to domain driven actionable data delivery.

III. DOMAIN DRIVEN DATA MINING

Intending at complementing the inadequacy of conventional data mining, in particular, reinforcing the problem-solving-oriented ability and deliverables in enterprise data mining, we recommend a realistic methodology, called Domain Driven Data Mining by following the extensively acknowledged jargon 'Data Mining'.

Domain Driven Data Mining is proposed as a methodology and a collection of techniques targeting domain driven actionable knowledge delivery to drive Knowledge Discovery from Data (i.e. KDD) toward enhanced problem-solving infrastructure and capabilities in real business state of affairs. On top of the data-centred framework, Domain Driven Data Mining aims to build up proper methodologies and techniques for targeting domain knowledge, human role and interaction, organizational and social factors, as well as capabilities and deliverables toward delivering actionable knowledge and supporting business decision-making action taking in the KDD process. In other words we can say that Domain Driven Data Mining intend to create subsequent generation methodologies, techniques and tools for a probable paradigm transfer from data centred out of sight pattern mining to domain driven actionable knowledge delivery.

As a result of the Domain Driven Data Mining investigation and development, we can deliver business-friendly and decision-making rules and actions that are of solid technical and business importance.

"Domain driven data mining refers to the set of methodologies, frameworks, approaches, techniques, tools and systems that deliver for human, domain, organizational and social, and network and web factors in the environment, for the innovation and delivery of actionable knowledge. Actionable knowledge means business responsive and comprehensible, reflects user preferences and business needs, and can be effortlessly taken over by business individuals for decision-making and action-taking"

The existing data mining methodology, usually chains self-governing pattern discovery from data. By contrast, the suggestion of domain driven knowledge discovery is to engage ubiquitous intelligence into data mining. The Domain Driven Data Mining highlights a procedure that discovers in-depth patterns from a constraint-based environment with the contribution of domain specialists and their acquaintance. Its intention is to maximally accommodate equally naive users as well as practised analysts, in addition to satisfy business goals. The patterns discovered are expected to be integrated into business systems and to be aligned with existing business rules. To formulate domain driven data mining successful, user guides and intellectual human-machine interaction interfaces are indispensable through incorporating mutually human qualitative aptitude and machine quantitative aptitude. In totalling, appropriate mechanisms are obligatory for dealing

through multiform restraints in addition to domain knowledge.

IV. KEY ELEMENTS OF DOMAIN DRIVEN DATA MINING

In domain-driven data mining, the following seven key elements play a very important role. They have capability of building a KDD which is dissimilar from the alive data-driven data mining if they are properly considered and supported from technical, procedural, and business point of view.

a) *Restraint -Based framework*

In human society, everyone is restrained either by communal regulations or by individual situations. Similarly, actionable knowledge only can be discovered in a restraint-based framework such as environmental authenticity, opportunities, and restraints in the mining procedure. Particularly, in the first section, we catalogue some types of restraints that play noteworthy roles in a process, effectively discovering knowledge actionable to business. In practice, a lot of other aspects, such as data stream and the scalability and effectiveness of algorithms, may be enumerated. They consist of domain-specific, functional, nonfunctional, in addition to environmental restraints. These ubiquitous restraints create a restraint-based framework for actionable knowledge discovery. All of the preceding restraints to varying degrees have to be considered in significant phases of real-world data mining. In this case, it is even called restraint-based data mining.

b) *Incorporate Field Awareness*

It is accepted gradually that field awareness can play noteworthy roles in real-world data mining. For instance, in trade (buy and sell) pattern mining, brokers often take "beating market" as an individual liking to judge a recognized rule's actionability. In this case, stock mining system requires to set in the formulas calculating market return and rule return, and set an interface in order for traders to specify a most wanted threshold and comparison relationship between the two returns in the evaluation process. Therefore, the key is to take advantage of field awareness in the KDD process.

The incorporation of field awareness is subject to how it can be signifying and filled in to the knowledge discovery process. Ontology-based field awareness representation, transformation, and mapping between business and data mining systems is one of the proper approaches to form field awareness.

c) *Collaboration Among Human beings and Mining Systems*

The genuine requirements for discovering actionable knowledge in restraint-based framework is more expected to be human involved rather than automated. Human involvement is embodied through the collaboration among humans (including users and business analysts, essentially domain experts) and data

mining systems. This is accomplished through the complementation between human qualitative brainpower, such as field awareness and field supervision, and mining quantitative brainpower like computational potential. Therefore, real-world data mining possibly presents as a human-machine-cooperated interactive knowledge discovery process.

For example, skills, metaknowledge, and invented philosophy of field experts can lead or help out with the selection of characteristics and models, adding industry features into the modelling, generating high-quality assumptions, designing interestingness measures by injecting business concerns, and rapidly estimating mining results. This help basically can progress the effectiveness and competence of drawing out actionable knowledge.

d) *Mining Exhaustively Patterns*

Sometime there is a situation that numerous mined patterns are attention-grabbing more to data miners than to businesspersons and such type of situations slowed down the deployment and implementation of data mining in real applications. For that reason, it is vital to estimate the actionability of a pattern in addition to to further find out actionable patterns to hold up smarter and more effectual decision making. This leads to exhaustively pattern mining.

Mining exhaustively patterns should think as how to get better both scientific and business interestingness in the previous restraint-based framework. Technically, it could be through enhancing or generating more effective interestingness measures. Additional awareness has to be remunerated to business desires, intentions, field acquaintance, and qualitative intelligence of field experts for their impact on taking out deep patterns.

e) *Improving Knowledge Actionability*

Patterns that are motivating to data miners may not guide essentially to business reimbursement, if deployed. For instance, a large number of association rules often are found, while most of them might not be workable in business state of affairs. These rules are generic patterns or technically interesting rules. Further actionability upgrading is mandatory for producing actionable patterns which is practically useful to commerce.

The measurement of actionable patterns is to follow the actionability of a pattern. Both technical and business interestingness measures must be satisfied from both objective and subjective point of view. For those generic patterns identified based on technical measures, business interestingness needs to be checked and emphasized so that the business requirements and user preference can be put into proper consideration.

f) *Loop - clogged repetitive Improvement*

Actionable knowledge discovery in a restraint - based framework is probably to be a clogged rather

than an open course of action. It includes repetitive feedback to varying phases such as sampling, assumption, feature selection, modelling, evaluation, and interpretation in a human-involved approach. On the other hand, real-world mining process is highly repetitive, because the evaluation and refinement of features, models, and outcomes cannot be completed once but, rather, is based on repetitive feedback and interaction before reaching the concluding juncture of knowledge and decision-support report delivery.

The previous key elements indicate that real-world data mining cannot be dealt by means of just an algorithm; rather, it is truly essential to assemble a suitable data mining infrastructure in order to find out actionable knowledge from restraint-based situations in a loop-clogged repetitive manner.

g) *Interactive and Concurrent Mining Supports*

To support domain-driven data mining, it is noteworthy to develop interactional mining supports for human-mining dealings and to estimate the findings. And also concurrent mining supports often are necessary and can deeply promote the real-world data mining performance.

For interactional mining supports, clever agents and service-oriented computing are a number of high-quality technologies. They can support flexible, business-friendly, and user-oriented human-mining interaction through building facilities for user modeling; user knowledge achievement; domain knowledge modeling; personalized user services and recommendation; run-time supports; and mediation and management of user roles, interaction, security, and cooperation.

The facilities for interactional and concurrent mining supports largely can improve the performance of real-world data mining in aspects such as human-mining interaction and cooperation, user modeling, domain knowledge capturing, reducing computation complexity, and so forth. They are a few crucial ingredients of subsequent generation of KDD infrastructure.

VII. CONCLUSION

Real-world data mining applications have projected critical desires for discovering actionable knowledge especially for real-users and industry needs. Actionable knowledge discovery is significant and also very challenging. It is listed as one of great challenges of KDD. The research on this issue has latent to revolutionize the alive state of affairs in which a huge quantity of rules are mined but still few of them are interesting to business, and to endorse the extensive deployment of data mining into business.

This research paper had tried to provide a novel data mining methodology referred to as Domain-Driven Data Mining. It provides a systematic indication of the issues in discovering actionable knowledge and

advocates the methodology of mining actionable knowledge in restraint-based framework through human mining system cooperation in a loop-clogged repetitive refinement manner. It is useful for promoting the paradigm shift from data-driven hidden pattern mining to domain-driven actionable data discovery.

REFERENCES REFERENCES REFERENCIAS

1. Boulicaut, J-F., & Jeudy, B. (2005). Constraint-based data mining. In O. Maimon, & L. Rokach (Eds.), *The data mining and knowledge discovery handbook* (pp. 399–416). New York: Springer.
2. Omiecinski, E. (2003). Alternative interest measures for mining associations. *IEEE Transactions on Knowledge and Data Engineering*, 15, 57–69.
3. Pohle, C. (n.d.). Integrating and updating domain knowledge with data mining. Retrieved from <http://citeseer.ist.psu.edu/668556.html>
4. S. Sharma and K. Osei-Bryson, *Role of Human Intelligence in Domain Driven Data Mining In: Data Mining for Business Applications*, New York: Springer Science+Business Media, 2009.
5. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining 2009*, Paris, France, June 28 - July 01, 2009.
6. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd edition, London: Morgan Kaufmann, 2006.
7. H. Varian, *Intermediate Microeconomics Fourth Edition*, New York: W. W. Norton & Company, 1996.