Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.708 – 712

REVIEW ARTICLE

A Review on the Role of Domain Driven Data Mining

Madeeha Aslam[1], Ramzan Talib[2], Humaira Majeed[3]

[1, 2, 3] College of Computer Science and Information Studies, Government College University, Faisalabad, Pakistan

[1] aslam.madiha@gmail.com

Abstract - Knowledge Discovery and Data Mining (KDD) refer to the overall process of discovering useful knowledge from data. It involves evaluation and possibly interpretation of the patterns to make decision of what qualifies as knowledge and gives choice of encoding schemes, preprocessing, sampling, and projections of data prior to data mining step. KDD applications in the real world can be as diverse as the real world big databases that exist today thus it lead us to poor knowledge of rich data. Therefore, domain driven data mining (D^3M), allow data mining and domain experts to complement each other in regard to in-depth granularity through interactive interfaces. The advantage of D^3M over traditional KDD is that, the involvement of domain experts and their knowledge can assist in developing highly effective domain-specific data mining techniques and can reduce the complexity of the knowledge-producing process in the real world business needs. In this study, an attempt is made to identify the role of D^3M in the business.

Keywords - Data mining, actionable knowledge discovery; domain-driven data mining; domain driven in-depth pattern discovery

I. INTRODUCTION

Data mining or knowledge discovery has emerged as one of the most vigorous areas in information and communication technologies in last few decades. Data mining is an iterative process involving a combination of various techniques of different disciplines (e.g. bioinformatics, medical, agriculture). When these techniques are applied to large data sets, data mining generates interesting knowledge, patterns, or high-level information of any dimension. The discovered knowledge can be applied to decision making, process control and information management. This has pushed data mining into the forefront of recent developments in information and communication technology. [1, 2].

The versatility of data mining motivates its research and development in academia and its applications in the business community [3]. To further increase this versatility, latest developments in data mining presented as publications in recent journals and conferences should be integrated into business applications in order to get better results as has been set as goals of most recent conferences such as SIGKDD [4].

The developments and applications of actionable knowledge discovery (AKD), a new paradigm shift in data mining, to real-world businesses and applications are based on Domain Driven Data Mining. Studies and research in this regard will make a huge difference in Business intelligence [5]. The final goal is to have data mining well integrated into the decision-making process for real life businesses by generating more accurate, timely and relevant information. For that purpose need arises for a better framework that produce results from existing data mining methodologies, techniques, tools and applications.

This leads to the emergence of domain driven data mining which primarily aims to deliver better decision making solutions for businesses by presenting tools for actionable knowledge that can be passed on to business people for direct decision-making and action-taking.[1] Domain driven data mining goes beyond the conventional data mining methods. It involves the application of relevant intelligence surrounding the business i.e., human intelligence, domain intelligence, network intelligence and organizational/social intelligence, and the combination of such relevant intelligence into a complete human Computer-cooperated problem-solving system [6].

II. MATERIAL AND METHODS

This paper is based on the randomly selected journal articles in the field of domain driven data mining. Eleven articles were identified positive to be fit for the study and were read completely for the review. In this review our emphasis is on paradigm shift from "data-cantered knowledge discovery" to "domain-driven actionable knowledge delivery and the role of domain driven data mining in in real world business. Main article opens with section III issues of traditional data mining are discussed and requirement is identified as being multidimensional. Later we discuss the key elements of domain-driven actionable knowledge discovery. In Section IV we discuss the concept and fundamental framework (domain driven in-depth pattern discovery) of D^3M, involving ubiquitous intelligence for D3M. In Section V we confer two techniques supporting D^3M and a case study is included to illustrate the use of D^3M in handling real-world problems. Section VI points some open issues for further research investigations in D^3M. Finally we conclude this paper work in Section VII. It's worth mentioning that all Calculations included in these articles were carefully rerun to guarantee information accuracy.

© 2014, IJCSMC All Rights Reserved

III. OVERVIEW OF DRIVING FORCES

A. *Issues of Traditional Data Mining Studies for real world business*

1) It's complicated to identify real word business problems and the main reason is a big gap between a syntactic system and its actual target problem. That makes identified patterns unable to solve that business need.
2) Existing work often stops at pattern discovery, which is mainly based on technical significance and interestingness.
3) Business concerns are not considered in evaluating patterns. Consequently, the identified patterns are mainly of technical interest.
4) Business people cannot themselves obtain interesting patterns for them and mostly, the mined patterns are not that informative and up to requirements of business people.
5) Mostly the mined patterns are of no interest of business needs thus it makes business people obscure why they should spend time on these irrelevant findings.
6) Business people are often naive users, and are also not informed how to interpret and use/execute them and what straightforward actions can be taken to engage them in business operational systems and decision making.

B. *Multidimensional Requirements on AKD :* Macro Level Issues

On the macro-level, issues are related to methodological and fundamental aspects [7]. The following typical macro-level issues need to be addressed:

1) *Environment:* Refer to any factors surrounding data mining models and systems, for instance, domain factors, constraints, expert groups, organizational factors, social factors, business processes, and workflows.

2) *Process:* Real-world problem solving has to cater for dynamic and iterative involvement of environmental elements and domain experts along the way.

3) *Infrastructure*: The engagement of environmental elements and humans at runtime in a dynamic and interactive way requires an open system with closed loop interaction and feedback. AKD infrastructure should provide facilities to support such scenarios.

4) *Dynamics:* To deal with the dynamics in data distribution from training to testing and from one domain to another, in domain and organizational factors, in human cognition and knowledge, in the expectation of deliverables, and in business processes and systems.

5) *Evaluation:* Interestingness needs to be balanced between technical and business perspectives from both subjective and objective aspects; special attention needs to be paid to deliverable formats, and its actionability and generalizable capability, as well as the support from domain experts.

6) *Risk*: Risk needs to be measured in terms of its presence and then magnitude, if any, in conducting an AKD project and system.

7) *Policy*: Data mining tasks often involve policy issues such as security, privacy, and trust existing not only in the data and environment, but also in the use and management of data mining findings in an organization's environment.

8) *Delivery*: Determining the right form of delivery and presentation of AKD models and findings so that end users can easily interpret, execute, utilize, and manage the resulting models and findings, and integrate them into business processes and production systems.

C. *Multidimensional Requirements on AKD : Micro Level Issues*

On the micro-level, issues related to technical and engineering aspects supporting AKD need to be addressed [7]. The following lists a few dimensions that address these concerns:

1) *Architecture*: AKD system architectures need to be effective and flexible for incorporating and consolidating specific environmental elements, AKD processes, evaluation systems, and final deliverables.

2) *Process*: Tools and facilities supporting the AKD process and workflow are necessary, from business understanding, data understanding, and human system interaction to result assessment, delivery, and execution of the deliverables.

3) *Interaction*: To cater for interaction with business people along the way of ADK process, appropriate user interfaces, user modeling, and servicing are required to support individuals and group interactions.

4) *Adaptation:* Data, environmental elements, and business expectations change all the time. AKD systems, models, and evaluation metrics are required to be adaptive for handling differences and changes in dynamic data distributions, cross domains, changing business situations, and user needs and expectations.

5) *Actionability*: What do we mean by "actionability?" How should we measure it? What is the trade-off between technical and business sides? Do subjective and objective perspectives matter? This requires essential metrics to be developed.

6) *Deliverable*: AKD deliverables are required to be easily interpretable, convertible into or presented in a business-friendly way and be compatible with business operational systems and rules. In this sense, AKD deliverables are required to be easily interpretable, convertible into or presented in a business-oriented way such as business rules, and be linked to decision-making systems.

© 2014, IJCSMC All Rights Reserved

According to [8] key elements of domain driven data mining are following:

1) *Restraint-Based framework*: Our society keep us restrained based either at communal of individual situations. Same way actionable knowledge can only be discovered in a restrained based framework, for instance, restraints at mining procedures.

2) *Incorporate Field Awareness:* it is based on ontology-based field awareness representation, transformation, and mapping between real world business and data mining systems, is one of the proper approaches to form field awareness.

3) *Collaboration Among Human beings and Mining Systems*: involvement of human is embodied with the collaboration among human and data mining systems, e.g. it include user and business analysts, essentially the domain experts to complement each other in substance of human qualitative brain power and mining qualitative brain power.

4) *Mining Exhaustively Patterns*: it must take notice of, how to get better results in both scientific and business interestingness in the previous restraint-based framework. Technically, it could be through enhancing more effective interestingness measures.

5) *Improving Knowledge Actionability*: Both technical and business interestingness measures are satisfies on both objective and subjective point of view by following actionability of pattern.

6) *Loop - clogged repetitive Improvement:* Actionable knowledge discovery is probably a clogged rather than an open course of action. Where it include the repetitive feedback to varying phases. For instance sampling, modeling and evaluation and interpretation in human-involved approach.

7) *Interactional and Concurrent Mining Supports*: to meet this area a high technology is used, such as some clever agents and service oriented computing., that is used to support business friendly and user oriented human-mining interaction through providing facilities of user knowledge achievements, domain knowledge modeling , run time support and management of user roles, security.

IV. D^{3}M THEORETICAL FRAMEWORK

A. *Basic Concepts*

Real-world data mining is a complex problem-solving system. The main objective of D^{3}M is to enhance the actionability of identified patterns for problem solving. The term "actionability" measures the ability of a pattern to prompt a user to take concrete actions to his/her advantage in the real world. It mainly measures the ability to suggest business decision-making actions.

The main task of D^{3}M is to develop AKD-oriented problem-solving systems. AKD- oriented D^{3}M, on top of the data-driven framework, aims to complement the shortcomings of traditional data mining, through developing proper methodologies and techniques to incorporate domain knowledge, user needs, the human role and interaction, as well as actionability measures into KDD process and systems [7]. It is data and domain intelligence working together to disclose a hidden story to business, and to satisfy real user needs. End users hold the final decision in evaluating the findings and business deliverables.

According to [9] Domain-driven data mining consists of a domain-driven in-depth pattern discovery (DDID-PD) framework. The DDID-PD takes *I^{3}D* (i.e., interactive, in-depth, iterative, and domain-specific) as real-world KDD bases. I^{3}D means that the discovery of actionable knowledge is an iteratively interactive in-depth pattern discovery process in domain-specific context. I^{3}D is further embodied through following:

1) Mining constraint-based context,
2) Incorporating domain knowledge through human-machine-cooperation,
3) Mining in-depth patterns,
4) Enhancing knowledge actionability,
5) Supporting loop-closed iterative refinement in order to enhance knowledge actionability.

Mining constraint-based context requests to effectively extract and transform domain-specific datasets with advice from domain experts and their knowledge. A system following the DDID-PD framework can embed effective supports for domain knowledge and experts' feedback and refine the life cycle of data mining in an interactive manner [10].

TABLE 1. COMPARES MAJOR ASPECTS UNDER THE RESEARCH OF TRADITIONAL DATA-DRIVEN AND DOMAIN-DRIVEN DATA MINING

Aspects	Data-Driven	Domain- Driven
Rational	Data tells a story	Data and ubiquitous intelligence disclose problem solving solutions
Objective	Innovative and effective algorithms	Effective problem solving
Data	Abstract, synthetic and refined data	Real life data and surrounding information
Process	One-off	Multiple step, iterative and interactive o demand
Mechanism	Automated	Human cantered or human-mining-cooperated
Infrastructure	Closed pattern mining systems	Closed loop problems-solving system in open environment
Usability	Predefined models and process	Ad-hoc, dynamic and customizable models and process
Deliverable	Patterns	Business-friendly decision support actions
Deployment	Solid validation	Well found art work in problem-solving
Evaluation	Technical merits	Trade-off between technical significance and business expectation

© 2014, IJCSMC All Rights Reserved

B. *Ubiquitous Intelligence Patient*

A key concept [7] in D^3M that defined as *actionable knowledge discovery* (AKD). It involves the following:

1) *In-Depth Intelligence*: Data Intelligence tells interesting stories or uncovers indicators about a business problem hidden in the data. Even though mainstream data mining focuses on substantial investigation of various data for interesting hidden patterns or knowledge, the real-world data and surroundings are usually much more complicated.

2) *Domain Intelligence:* Domain Intelligence emerges from domain factors and resources that not only wrap a problem and its target data but also assist in problem understanding and problem solving. Domain intelligence involves qualitative and quantitative aspects. These are instantiated in terms of aspects such as domain knowledge, background information, prior knowledge, expert knowledge, constraints, organization factors, business process, and workflow, as well as environment intelligence, business expectation, and interestingness.

3) *Human Intelligence*: Human Intelligence refers to

   - explicit or direct involvement of human pragmatic knowledge, belief, intention, expectation, runtime supervision, evaluation, and expert groups into AKD;
   - implicit or indirect involvement of human intelligence such as imaginary thinking, emotional intelligence, inspiration, brainstorm, reasoning inputs, and embodied cognition like convergent thinking through interaction with other members in dynamic data mining and assessing identified patterns.

4) *Social Intelligence:* Social Intelligence refers to the intelligence that lies behind group interactions, performances, and corresponding parameter. Social intelligence concealments both human social intelligence and agent-based social intelligence. Human social intelligence is related to aspects such as social cognition, emotional intelligence, consensus construction, and group decision. Agent-based social intelligence involves swarm intelligence, action selection, and the foraging procedure. Both sides also engage social network intelligence and collective interaction, as well as business rules, law, trust, and reputation for governing the emergence and use of social intelligence.

V. TECHNIQUES

In this section, we briefly introduce two techniques of D^3M:

C. *Combined Mining*

It's for complex knowledge in complex data, that include the combination of multiple data sources, the combination of multiple features and the combination of multiple methods e.g., association mining and classification.

D. *Agent-driven Data Mining*

It is for enhancing interaction, coordination, and distributed processing in complex data mining applications. Main strength of this technology can greatly complement data mining, in particular complex data mining problems in aspects such as data processing, information processing, pattern mining, user modeling and interaction, infrastructure, and services.

In [11] a case study has been done in which domain driven data mining is applied to business domain specifically in risk management (in business of car insurance company). For regular insurance companies the fraud ratio is 8%. For the WBF this fraud ratio is probably higher, since there is only one party that can be questioned for information. This fraud percentage is estimated by fraud experts to be around 12%. At the moment only 4% of the fraudulent claims are identified. To overcome this issue the effectiveness of data mining systems can be significantly improved as compared to data mining systems based on blind search only. It can be achieved by including domain knowledge about the model to be constructed by help of experienced domain experts. Figure 2 shows the flow of claims and how domain experts is added to framework for an extra check after the data mining system has filtered out suspicious claims.

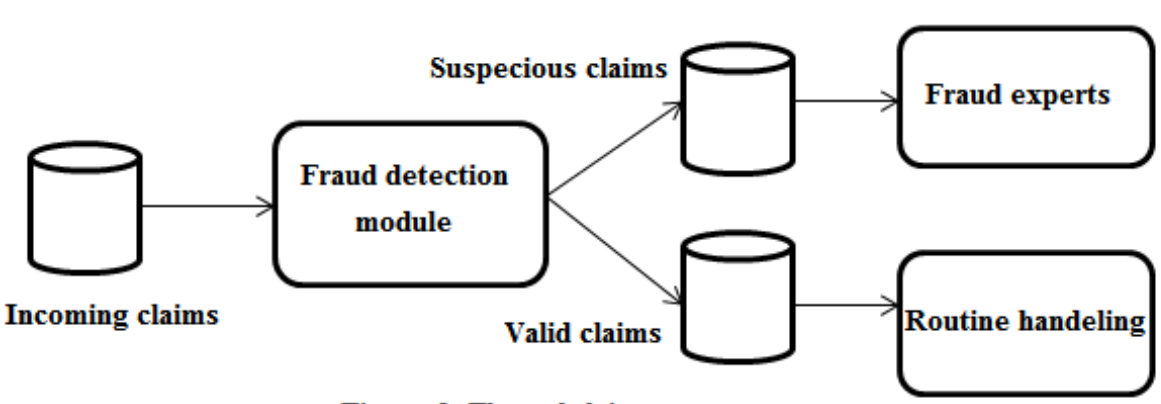

Figure 2. Flow of claims

This approach has an advantage that the blind search in databases is now guided by domain expert and their experience is leading to substantially more accurate results.

VI. OPEN ISSUES

Accordingly, many open issues await further research investigation. For instance, to effectively synthesize the ubiquitous intelligence in actionable knowledge discovery, many research issues need to be studied or revisited [7].

1) Typical research issues and techniques in Data Intelligence include mining in-depth data patterns disclosing deep knowledge in data and context, combined patterns consisting of information from heterogeneous sources, and structured knowledge in mix-structured data.

© 2014, IJCSMC All Rights Reserved

2) Typical research issues and techniques in Domain Intelligence consist of static and dynamic representation, modeling and involvement of domain knowledge, constraints, organizational factors, and business interestingness into models and the AKD process, and of both syntactic and semantic aspects.

3) Typical research issues and techniques in Network Intelligence include involving not only general techniques such as information retrieval, text mining, web mining, web intelligence into data mining, but also the involvement of web/networked facilities to support web-based data mining, the discovery of networks and communities in networked data, and web/network knowledge management in the data mining process and systems.

4) Typical research issues and techniques in Human Intelligence include representation and involvement of empirical and implicit knowledge, reasoning and situated computing capabilities, as well as runtime human-machine interaction into data mining process and models at both individual and group levels.

5) Typical research issues and techniques in Social Intelligence include collective intelligence, social network analysis, and social cognition interaction in data mining systems, building social data mining software that can cater for ubiquitous intelligence in a social context.

VII. CONCLUSION

This study examined a systematic overview of concepts, challenges, techniques, and prospects of D^3M. It presents an overview of driving forces, theoretical frameworks, techniques, and open issues of D^3M. We tried to study also the aspects of how to apply data mining via domain driven data mining can be applied to real world businesses in order to yield more useful results. A case study is reviewed which shows the effectiveness and efficiency of domain knowledge, that is applied in addition to the data mining techniques yielding a significant improvement in the results obtained.

REFERENCES

[1] A Adejuwon and A Mosavi, "*Domain Driven Data Mining: Application To business*" IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 2, July 2010.
[2] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd edition, London: Morgan Kaufmann, 2006.
[3] H. Varian, *Intermediate Microeconomics Fourth Edition, New York*: W. W. Norton & Company, 1996.
[4] Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining 2009, Paris, France, June 28 - July 01, 2009.
[5] L. Cao, and C. Zhang, *The Evolution of KDD: Towards Domain-Driven Data Mining*, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 21, No. 4, 2007, pp.677-692.
[6] L. Cao, P. S. Yu, C. Zhang and Y. Zhao (eds), *Data Mining for Business Applications*, New York: Springer Publishers, 2009.
[7] L Cao, "Domain Driven Data Mining: Challenges and Prospects" IEEE Transactions On Knowledge And Data Engineering, VOL. 22, NO. 6, June 2010.
[8] M. Kumari, *Data Driven Data Mining to Domain Driven Data Mining*, Global Journal of Computer Science and Technology Volume 11 Issue 23 Version 1.0 December 2011
[9] L. Cao, *Domain-Driven Data Mining: A Practical* Methodology , International Journal of Data Warehousing & Mining, 2(4), 49-65, October-December 2006
[10] L. Cao, L. Lin C. Zhang *Domain-Driven In-Depth Pattern discovery: A Practical Methodology*[1]
[11] H. Daniels and H. Dissel, *Risk Management based on Expert Rules and Data Mining, A Case Study in Insurance*. In Proceedings of the 10th European Conference on Information Systems (ECIS), 2002, Gdansk.

© 2014, IJCSMC All Rights Reserved