Domain Driven Data Mining (D³M)

Ashima Khanna¹ and Zoya Siddiqui²
¹,²Department of MCA, Institute of Technology & Science, Ghaziabad
¹sardana.ashima@gmail.com, ²zoyasiddiqui67@gmail.com

ABSTRACT
The current data mining algorithms and tools faces critical challenges in solving real-world complex problems as they do not provide action to be taken by Business people on basis of delivered technical outputs. Next-generation methodologies, techniques and tools like Domain Driven Data Mining (D³M) will bridge the gap between academic output and business expectations by considering real world factors such as human knowledge, constraints and business expectations. D³M will involve ubiquitous intelligence and meta-synthesis into the mining process, and an actionable knowledge discovery-based problem-solving system as the space for data mining. D³M methodology will be able to cater for organizational factors, user preferences and business needs. In this paper, aim is to identify various directions, issues for research and areas of application etc. related to D³M by reviewing and studying the latest methodological, technical and practical progresses in D³M.

KEYWORDS
Actionable Knowledge Discover (AKD), Domain Driven Data Mining (D³M), meta-synthesis, Next-generation

1. INTRODUCTION
Recent developments in technologies in area of acquisition of data, storage of data, computation and communications make it possible to collect, store and process enormous volumes of data. It creates unprecedented opportunities for knowledge discovery form large-scale database. Data mining technology is a useful tool for these kinds of problems. Data Mining uses various techniques and methodologies to discover knowledge from data and present it in usable form. It has got lot of attention in recent time.

Data mining is a process of identifying understandable patterns from data through searching, mining and utilizing the functionalities of different patterns embedded in various databases.[1,4,6,10] Always it has been attempted to make processes of mining more effective and efficient. In the last decade, data mining has emerged as one of the most vigorous and charming areas in information technology. Current data mining is heavily dependent on data itself, and relies on data-centered methodologies. Existing data mining approaches either view data mining as data-driven trial-and-error process, or analyze business aspects in an isolated and case to case basis. Due to this, most of the time, the knowledge discovered does not always generally satisfy real business requirements.

When data mining is to be deployed into the real-world scenario then it is required to serve for business needs and scenarios, various organizational factors and preferences of user. Most of the time current data mining tools and techniques specific problems, such as financial areas, bridge the gap between data mining research and business expectations, and drive a paradigm shift from traditional data-centered hidden pattern mining to domain-driven actionable knowledge discovery. D³M has re-interpreted, re-visited, weak areas or ignored areas in classical data mining methodologies. D³M will bridge the gap between academic output and business expectations by considering real world factors such as human knowledge, constraints and business expectations. D³M will involve ubiquitous intelligence and meta-synthesis into the mining process, and an actionable knowledge discovery-based problem-solving system as the space for data mining. D³M methodology will be able to cater for organizational factors, user preferences and business needs. In this paper, aim is to identify various directions, issues for research and areas of application etc. related to D³M by reviewing and studying the latest methodological, technical and practical progresses in D³M.

2. ISSUES OF CLASSICAL DATA MINING
Classical data mining research mainly focuses on demonstrating the use of specific algorithms and models. As a
result, many algorithms have been designed out of which only very few are repeatable and executable in the real world scenario. Many patterns are discovered by traditional data mining but very few are of any particular interest to business. Due to this end users generally cannot utilize them for their business use. This makes that the outcome or findings of data mining are not actionable and lack soft power in problem solving. Now-a-days lot of research work is being done on Domain driven data mining (D3M) methodology, which can tackle the above issues, and is a paradigm shift from 'data-centered knowledge discovery'; to 'domain-driven, actionable knowledge delivery'[1]. In D3M, all the intelligence available in surrounding is incorporated into the data mining process and models, and a corresponding problem-solving system is formed as the space for knowledge discovery and delivery. The traditional data mining faces critical challenges during deployment and lacks soft power in solving real-world complex problems. After the paradigm shift from ‘data mining’ to ‘knowledge discovery’, there is a need of wider acceptance and employment of knowledge discovery in real-world decision-making.[1,7] So, another paradigm shift is expected from ‘data-centered knowledge discovery’ to ‘domain-driven actionable knowledge discovery’. In the domain-driven actionable knowledge discovery, all the intelligence available in surrounding must be involved and meta-synthesized into the data mining process, and an actionable knowledge discovery-based problem-solving system is formed as the space for data mining. This is the motivation and aim of developing $D^3M$ and applying it in different domains.

3. CONCEPT MAP OF D3M
Following are four layers of concept map of D3M from outer most layer to central core:

**Decision Support Power:**[13]
D3M is concerned with main issues which reflects and enhances decision making power of knowledge identified and deliverables during data mining in terms of performances through adaptability, dynamics, action-ability, workability, operability, dependability, repeatability, trust, explain-ability, transferability and usability

**D3M theoretical foundations:**[13]
D3M supports tools and techniques needed for foundational support from many relevant areas ranging from the informational sciences to social sciences. In particular need to create new scientific fields such as data sciences, web sciences and service sciences which are targeting the establishment of family of scientific foundations, techniques and tools for dealing with emergent complexities and challenges in the respective area.

**D3M supporting techniques:**[13]
D3M engages and consolidates fundamental issues surrounding domain driven actionable knowledge discovery, there is a need to develop corresponding techniques and tools for involving and utilizing ubiquitous intelligence, supporting knowledge representation and deliverables, catering for project and process management, and implementing decision-making pursuant to the findings.

4. PHASES OF D3M PROCESS
The typical phases of lifecycle of the D3M process is as follows, some phases may be bypassed or there may be change in sequence or repetition in dealing with a real problem. Every step of the $D^3M$ process may involve ubiquitous intelligence and interaction with business users and/or domain experts. This D3M Process is already suggested by [7] in continuation we are focusing on these parameters which may be beneficial for beginning of this area:

**Problem Understanding and Definition**
Understanding the problem is the first stage of most descriptions of problem solving Domain experts and user work together to understand the domain knowledge. And answer the following questions: Who uses, why uses, what to use, where to use and how to use.

**Constraint Analysis**
D3M has to deal with following types of Constraints based context like
- Data constraints.
- Domain knowledge constraints.
- Interestingness constraints.
- Rule constraints or action-ability constraint

**Data Preprocesses**
Data Preprocessing is required because we have incomplete data, inconsistent data, Noisy data etc. For this we required quality measure in terms of accuracy, timeliness, completeness, consistence etc. Domain experts can create “Artificial Data” for some situations either data is sufficient or not.

**Modeling**
In modeling the experts uses all the mining the functions on data mining problems. The modeling will be done after preprocessing the data i.e. data preparation, artificial data etc and then uses the algorithms on the data.

**Interesting and Actionable Rules.**
It has been organized that a knowledge discovery can generate lots of patterns/rules most of which are not interest to the expert.[9] As a result, now days the focus of data mining has moved from discovered rules into discovered interesting rules. There are two meaningful measures: Objective Measure and Subjective Measure.[11] Actionable rules algorithm exam the data in the objective way and represent the discovered information. The rules present to a decision maker should only consist of simple, understandable and complete strategies that allow a reasonably easy identification of preferable rules. It supports human-machine-cooperated interactive knowledge discovery and this in-depth pattern discovery is conducted under the cooperation of business analysts and data analysts.[12]

**Feedback from Real World.**
All test will be tested from real world and then confirmed that either these rules are actionable or not. And from the feedback
of the World we pass all the algorithms or modeling techniques to our data.
In this manner, data mining is viewed as a loop-closed iterative refinement process. The data mining process is closed with iterative refinement and feedbacks of hypotheses, features, models, evaluation and explanations in the human-involved or centered context. [1,6]

5. INTELLIGENCE SURROUNDING AKD

Domain Driven Data Mining is a next-generation data mining methodologies, tool, frameworks and algorithm which aim to promote the paradigm shift from data-centred hidden pattern mining to domain-driven actionable knowledge discovery (AKD).[1] To achieve this, D3M needs to involve and integrate all the intelligence in surrounding like 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.

The above ubiquitous intelligence has to be combined for problem solving. The methodology for combining such intelligence is called meta-synthesis, which provides a human-centred and human-machine-cooperated problem solving process by involving, synthesizing and using ubiquitous intelligence surrounding AKD as need for problem solving. D3M serves for the effective involvement of the following six types of ubiquitous intelligence surrounding AKD based problem:

Human role and intelligence

Human role and intelligence refers to explicit or direct involvement of humans such as empirical knowledge, belief, intention and expectation, run-time supervision, evaluating and expert group; It also refers to implicit or indirect involvement of human intelligence such as imaginary thinking, emotional intelligence, inspiration, brain-storm and reasoning inputs.

Domain knowledge and intelligence

Domain knowledge and intelligence is not just only for wrapping a problem and its target data but it also assists in the understanding and problem solving of the problem. Domain knowledge consists of qualitative and quantitative intelligence. Both types of intelligence are instantiated in terms of aspects such as domain knowledge, background information, constraints, organization factors and business process, as well as environmental intelligence, business expectations and interestingness.

Network and web intelligence

Network with web intelligence refers to both web intelligence and broad based network intelligence such as distributed information and resources, linkages, searching and structured information from textual data.

Organizational and social intelligence

Organizational and social intelligence onsists of interpersonal intelligence, emotional intelligence, social cognition, consensus construction, group discussion, as well as organizational factors, business process, workflow, project management and delivery, social network intelligence, collective interaction, business rules, law, trust and so on.

In-depth data intelligence

In-depth data intelligence mainly tells about hidden pattern in data for a business problem.

Meta-synthesis of the above intelligence[10]

In the real world, DM is a problem solving process from business problems to problem-solving solutions. From modelling perspective, such a problem-solving process is a state transformation process from source data to resulting pattern set. The methodology for combining above intelligence is called meta-synthesis, which provides a human-centred and human-machine-cooperated problem solving process by involving, synthesizing and using ubiquitous intelligence surrounding AKD as need for problem solving.

6. BUSINESS APPLICATIONS AND RESEARCH AREAS

In the last few years, applications of data mining and knowledge discovery have been increasingly undergoing enormous transformation as it is using many techniques which are statistical methods like statistic based rule induction, knowledge extraction and abstraction which are using decision tree, neural network, artificial intelligence, fuzzy set, linear regression and linear predictive coding. Growth of applications of data mining is influenced by enormous growth of e-commerce, m-commerce tremendous progress in information technology and extensive use of data mining for security. In security area, combined mining is also used. Other applications include financial data mining especially important for capital market, trading strategies and understanding behavior patterns of microstructure of market.[3]

D3M can be developed and deployed in various areas such as

- Stock market,
- Medical & pharmaceutical sciences,
- Biological Data Mining and its Applications in Healthcare
- Steganalysis,
- Transport network,
- Crime detection for credit applications,
- Cross-market surveillance or distributed fraudulent activities
- Blog search,
- Social security area etc.
- Behavior patterns of Micro-structure of Finance
- Climate Data Prediction

An interactive approach with domain experts may be required to prune and filter discovered rules might be required. In many applications it can make possible to use concept of activity mining and combined mining. Applications may require to draw information from multiple and distributed sources. [2]

Research areas of D3M may include knowledge Discovery from extremes and analyzing its impacts; identifying rare and dispersed yet significant links in cross-organization data;
Visual Analytics; Social Interactions Analysis; Opinion Mining. [9]

Data mining methods could be interesting to generate substantial profits for decision makers and to optimize the choice of different business activities.

7. CONCLUSION

It is clearly realized that there is a need for domain driven data mining, and efforts are required to develop corresponding techniques and applications. The research and development is required for discovering actionable knowledge from complex domain problems, enhancing interaction and reducing the gap between academia and business, and driving a paradigm shift from interesting hidden pattern mining to actionable knowledge discovery in varying data mining domains. Data mining in the real world needs to develop innovative methodologies, approaches, and enterprise applications for workable, dependable, and actionable knowledge discovery in the real life. Given the nature of D³M, it can bring about the effective and practical development of many challenging data mining applications in every area. D³M can be developed and deployed in various areas such as stock market, medical & pharmaceutical sciences, steganalysis, transport network, crime detection for credit applications, blog search, social security area etc. An interactive approach with domain experts may be required to prune and filter discovered rules might be required. It can be helpful to develop actionable agents, actionable strategies, and exceptional microstructure behavior patterns. In many applications it can make possible to use concept of activity mining and combined mining.

We expect a new paradigm shift from ‘data-centered knowledge discovery’ to ‘domain-driven actionable knowledge discovery’

8. FUTURE SCOPE

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

- Mining structured knowledge in unstructured data.
- Representation, modeling and involvement of domain knowledge, constraints, organizational factors, and business interestiness.
- Ontological engineering techniques, and web knowledge management to improve the integration of user knowledge in the post-processing task, it can contribute to attribute selection and in the interpretation of mining results to make data mining better aligned with business understanding.
- An interactive framework may be designed for Human-machine interaction and representation and involvement of empirical and implicit knowledge.
- Building metasynthetic interaction as working mechanism, and metasynthetic space as an AKD-based problem-solving system.

Fundamental Research issues are infrastructure capabilities for handling and consolidation of ubiquitous intelligence surrounding a domain problem by D³M and way of presenting it to decision maker.

REFERENCES

[7]. Z. Zhu, J. Gu, W. Yang, X. Li “Towards Domain Driven Data Mining” International Symposium on Intelligent Technology Application Wokshops,2008 pages 44-48
[10]. Longbing Cao: Domain Driven Data Mining (D3M), 2008 IEEE International Conference on Data Mining Workshops
[13]. Domain Drive Data Mining, Springer, Chapter D3M Methodology, Pg No 28