

Domain Driven Data Mining – Application to Business

Adeyemi Adejuwon and Amir Mosavi

University of Debrecen, Faculty of Informatics, Egyetem ter1, Debrecen 4032,
Hungary

Abstract

Conventional data mining applications face serious difficulties in solving complex real-life business decision making problems when practically deployed. This 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 (DDDM). 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.

Keywords: business, business intelligence, Domain Driven Data Mining, Data Mining.

1. Introduction

In recent times, the application of information technology (IT) to yield better performance in the business domain is all pervasive. However, past research suggests that not all investments in IT made by businesses result in improved performance. Rather, a high conversion effectiveness of the IT investments to measurable business objectives is necessary before a positive impact can be achieved [1]. Businesses that can efficiently transform data into useful information can use them to make quicker and more effective decisions and thus form better actionable business strategies which will give them a competitive edge.

The development of actionable business strategies is however not an easy task due to domain knowledge constraints and expectations of key decision makers of the business domain [2]. When arrived at, the correct business strategy addresses such issues as personnel selection, accurately identifying target markets, consumer preferences and effectively managing the process by which goods and services are produced and delivered to

the consumer [3]. Through a balanced mix of economic theories and IT, businesses can effectively devise appropriate strategies. The effective use of IT to achieve business intelligence via data mining techniques enable businesses to more quickly and accurately analyze operations in areas such as customer relationship management, personnel management, and finance.

In the following sections, we examine business intelligence, data mining and domain driven data mining and how a combination of these can be used to solve real-life business decision making problems. We thereafter review case studies in which domain knowledge has significantly impacted on the results obtained from the application of data mining to real-life business problems.

1.1 Business Intelligence

One of the primary factors that influence the performance of businesses is the ability to make effective and timely decisions in a consistent manner. Businesses now have access to large amounts of business related data but are unable to utilize them. Business intelligence aims to bridge this gap by providing businesses with tools and methodologies with which to harness the potential in already available business related data, thereby facilitating more effective and timely decisions [4].

The ability of a business to correctly transform and utilize information and knowledge in a timely manner is called “business intelligence”. Business intelligence methodologies are varied and complex and have a wide area of application. The major advantage deriving from the adoption of a business intelligence system is found in the increased effectiveness of the decision-making process.

The complexity of business intelligence real-world problems can be categorized into:

- Human roles and intelligence,
- Domain knowledge and intelligence,
- Network and web intelligence,

- Organizational and social intelligence,
- In-depth data intelligence,
- Metasynthesis of the above intelligences.

As it is mentioned in [5] there is inadequate literature in these regard which leaves lots of room for further research in above categories.

In the business environment, the main objectives of business intelligence are to make effective and timely decisions, and to reduce uncertainty. These objectives are realized through the application of data mining as a tool of business intelligence by providing the means to transform data into useful and actionable knowledge [6].

1.2 Data Mining

Data mining or knowledge discovery has emerged as one of the most active areas in information and communication technologies (ICT). Data mining is an iterative process involving a combination of techniques from several disciplines. When applied to large data sets, data mining yields interesting knowledge, patterns, or high-level information which can be viewed from different angles. 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 ICT [7].

The versatility of data mining motivates its research and development in academia and its applications in the business community [8]. 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 [9]. Our goal would be finding out such developments and techniques to improve the efficiency of business intelligence.

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 [10]. 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. With a more timely and streamlined flow of more accurate, business related information, decision makers across the pyramid structure of businesses have a better idea of what is happening in the world in which they operate. Not only do they more quickly receive reports that are more understandable, but

can also navigate through business related data to answer a multitude of business questions in a timely manner. In the next section, the concept of domain driven data mining is briefly introduced.

1.3 Domain Driven Data Mining

An imbalance exists between the number of data mining algorithms published and the very few that are actually useful in a business setting. Hence conventional data mining has failed to deliver adequate results in decision making for real life business. A need therefore arises for a better framework within which better results can be obtained from existing data mining methodologies, techniques, tools and applications [11].

This led 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.

Domain driven data mining aims to introduce a new paradigm shift; from data-centered hidden pattern mining to domain driven actionable knowledge discovery and delivery [12].

2. Review

Data mining techniques are effective at generating useful statistics and finding patterns in large volumes of data, but as Pohle [13] mentioned, not as effective at interpreting these results, which is crucial for turning them into interesting, understandable and actionable knowledge.

A so called knowledge acquisition bottleneck is caused by experts who have gathered a lot of experience over long periods of time in a particular domain but are unable to use this to effectively solve problems in a timely manner. A major benefit of using data mining techniques is that it bypasses the knowledge acquisition bottleneck [14].

The application of data mining techniques alone is not sufficient in solving real-world business problems. Dybowski et al. (2003) [15] investigated how data mining techniques and domain knowledge can be combined to construct more useful, efficient and effective decision support systems. Fayyad et al. [16] also suggested that the use of domain knowledge is important in all stages of the knowledge discovery process.

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 [17].

The significance of human intelligence in data mining was investigated by S. Sharma and K. Osei-Bryson [18]. The researchers identified twelve data mining processes which require human intelligence. They posited that data mining requires human intelligence in order to generate valid and meaningful results.

As a direct application of this, a recent research carried out by Atish et al. (2008) [19] established that there is interaction between the classification method of data mining and domain knowledge. They concluded that the incorporation of domain knowledge has a higher influence on performance for some data mining methods than for others.

Chien et al. (2006) [20] collaborated with domain experts to develop specific recruitment and human resource management strategies using data mining techniques. Their results were successfully applied in a real-world business.

Zhao et al. (2009) examined the effects of feature construction [21] guided by domain knowledge, on classification performance. The results of their study showed that feature construction, guided by domain knowledge, significantly improves classifier performance.

Based on the surveys presented in [17] many more studies exist showing the importance of domain knowledge in data mining. This solidifies the argument for the inclusion of domain knowledge in data mining techniques to increase their relevance, efficiency and efficacy in real-world business decision making.

3. Case Studies

In this section, three case studies are reviewed. These case studies show how domain driven data mining can be applied to the business domain, more specifically in the sectors of risk management in insurance, churn prediction and personnel selection.

3.1 Risk Management in Insurance

Data mining aims to derive valuable business knowledge from patterns in database. In the majority of cases there is theoretical and domain dependent knowledge available. This study, carried out by Daniels and Dissel (2002) [22], investigates risk management in the Insurance business. In

this case study, it is observed that the effectiveness of data mining systems can be substantially improved as compared to data mining systems based on blind search only, by including knowledge about the model to be constructed and knowledge of experienced domain experts. This approach has two advantages. First, the otherwise blind search in databases is now guided by expert experience leading to substantially more accurate results. Second, since in general experts find it difficult to combine decision rules into a single risk score, the framework discussed offers the possibility to combine and fine-tuned expert knowledge using real cases. Figure 1 below shows the flow of claims and how domain experts is added as an extra check after the data mining system has filtered out suspicious claims.

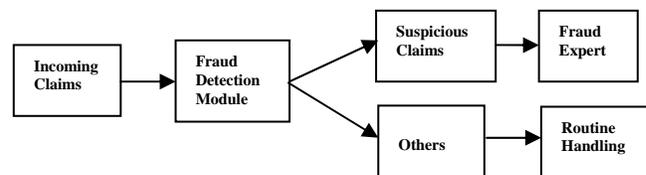


Figure 1: Flow of claims

3.2 Churn Prediction

Several data mining models and algorithms exist that carry out churn prediction [23]. In this case study [24], the researchers showed how to make data mining models developed for churn prediction more understandable and compliant by combining it with relevant domain knowledge. More specifically, they showed how the analysis of coefficient signs in logistic regression and the monotonicity analysis of DTs can be used to check whether the knowledge contained in data mining models is in accordance with domain knowledge, and how to correct any discrepancies found. The idea is to help companies discover which customers are more valuable and also to help them identify the main elements in their data that can contribute positively or negatively to the relationship with the customer, and through that, define strategies that would benefit both company and customer alike.

3.3 Personnel Selection

High-tech companies rely on human capital to maintain competitive advantages. This study developed a data mining framework to extract useful rules from the relationships between personnel profile data and their work behaviors. Furthermore, the researchers developed useful strategies with domain experts in the case company and most of the suggestions have been implemented. With

an effective personnel selection process, organizations can find the suitable talents at the first attempt to improve retention rate and generate better performance [20].

4. Conclusion

This study examined how data mining via domain driven data mining can be applied to businesses in order to yield more useful results. Three case studies were reviewed which show the effectiveness and efficacy of this method in the business domain. In each case study reviewed, domain knowledge was applied in addition to the data mining techniques and this yielded a significant improvement in the results obtained.

Areas of future study could be the expansion of the scope to consider other fields such as agricultural, engineering and medical applications.

Acknowledgments

The supports of the Dr. Laszlo Kozma, Director of the International Studies Centre of Arts, Humanities and Sciences and Ms. Denissza Blanar head office of International Relations at University of Debrecen are strongly acknowledged.

References

- [1] P. Weill, "The Relationship between Investment in Information Technology and Firm Performance", *Information Systems Research*, Vol 3, No 4, 1992, pp 307-333.
- [2] L. Cao and C. Zhang, Knowledge Actionability: Satisfying Technical and Business Interestingness, *International Journal of Business Intelligence and Data Mining*, Vol. 2 No. 4, 2007, pp. 496-514.
- [3] L. Cao, P. S. Yu, C. Zhang and Y. Zhao, *Domain Driven Data Mining*, New York: Springer Publishers, 2010.
- [4] Moss L., Atre S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*.
- [5] Carlo Verzellis, *Business Intelligence: Data Mining and Optimization for Decision Making*, Indianapolis: John Wiley Publishers, 2009.
- [6] S. Kudyba and R. Hoptroff, *Data Mining and Business Intelligence: A Guide to Productivity*, London: Idea Group Publishing, 2001.
- [7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd edition, London: Morgan Kaufmann, 2006.
- [8] H. Varian, *Intermediate Microeconomics Fourth Edition*, New York: W. W. Norton & Company, 1996.
- [9] Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining 2009, Paris, France, June 28 - July 01, 2009.
- [10] 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.

- [11] Ankerst M. Report on the SIGKDD-2002 Panel the Perfect Data Mining Tool: Interactive or Automated? *ACM SIGKDD Explorations Newsletter*, 4(2):110-111, 2002.
- [12] Cao L and et al. Domain-driven data mining: a practical methodology, *Int. J. of Data Warehousing and Mining*, 2(4): 49-65, 2006.
- [13] C. Pohle, Integrating and updating domain knowledge with data mining, In: M.H. Scholl, T. Grust (Eds.), *Proceedings of the VLDB 2003 PhD Workshop (electronic edn.)*, Berlin, Germany, 2003
- [14] R.R. Hoffman, The problem of extracting the knowledge of experts from the perspective of experimental psychology, *AI Magazine*, Vol. 8, 1987, pp. 53-67.
- [15] R. Dybowski, K.B. Laskey, J.W. Myers and S. Parsons, Introduction to the special issue on the fusion of domain knowledge with data for decision support, *Journal of Machine Learning Research* Vol 4, 2003, pp. 293-294.
- [16] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, From data mining to knowledge discovery in databases, *AI Magazine* Vol. 17 No. 3, 1996, pp. 37-54.
- [17] L. Cao, P. S. Yu, C. Zhang and Y. Zhao (eds), *Data Mining for Business Applications*, New York: Springer Publishers, 2009.
- [18] 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.
- [19] A. P. Sinha and H. Zhao, Incorporating domain knowledge into data mining classifiers: An application in indirect lending, *Decision Support Systems* vol.46, 2008, pp.287-299.
- [20] C. Chien and L. Chen, Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry, *Expert Systems with Applications*, Vol. 34, 2008, pp. 280-290.
- [21] H. Zhao, A. P. Sinha and W. Ge, Effects of feature construction on classification performance: An empirical study in bank failure prediction. *Expert Systems with Applications* Vol. 36, 2009, pp. 2633-2644
- [22] 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.
- [23] H. Liu and H. Motoda, Less is more In Feature Extraction, Construction and Selection: a data mining perspective, Norwell: Kluwer Academic Publishers, 1998.
- [24] E. Lima, C. Mues and B. Baesens, Domain knowledge integration in data mining using decision tables: case studies in churn prediction, *Journal of the Operational Research Society* Vol. 60, 2009, pp. 1096-1106.

Adeyemi Adejuwon Adeyemi Adejuwon is doing his masters degree in computer science at University of Debrecen, Faculty of Informatics. He is interested in data mining studies and its applications to business.

Amir Mosavi is a PhD candidate and teacher assistant at university of Debrecen Faculty of Informatics. He is working in engineering MCDM and MOO as well as data mining. He has more than ten published journals and conference proceedings so far in these fields.