Application of Data Mining Techniques in Knowledge Management System

Mohnish Mahamune[#1], Sandip Ingle[*2]

[#1]*School of Computational Science, S.R.T.M.University,Nanded (M.S.), India*[2]
[1]mohnish.mahamune@gmail.com

* *Dept of Computer Science, Hi-Tech Institute of Technology, Aurangabad (M.S.),India*
[2]seingle@gmail.com

***Abstract*— The primary social and economic value of modern societies is knowledge. For this reason, mastery of Information technology for structured or unstructured information recorded in Data warehouses such as the web is indispensable for the development of both the individuals and the society. Data mining and knowledge management (DMKM) has become essential for improving the competitiveness of businesses and increasing access to knowledge. DMKM still, however, comes up against major scientific and technological obstacles.**

Data mining is one of the most important steps of the knowledge discovery in databases process and is considered as significant field in knowledge management. This paper explores the applications of data mining techniques which have been developed to support knowledge management process.

***Keywords*— Data Mining; Data mining applications; Knowledge management.**

I. INTRODUCTION

Knowledge Managers in any organization need to integrate Information Systems Strategies with Business Strategies in order to attain their vision and mission. In information era, knowledge is becoming a crucial organizational resource that provides competitive advantage and giving rise to knowledge management (KM) initiatives. Many organizations have collected and stored vast amount of data. However, they are unable to discover valuable information hidden in the data by transforming these data into valuable and useful knowledge [2]. Managing knowledge resources can be a challenge. Many organizations are employing information technology in knowledge management to aid creation, sharing, integration, and distribution of knowledge. Knowledge management is a process of data usage [6]. The basis of data mining is a process of using tools to extract useful knowledge from large datasets; data mining is an essential part of knowledge management [6]. Data mining can be useful for KM in two main manners: (i) to share common knowledge of business intelligence (BI) context among data miners. (ii) to use data mining as a tool to extend human knowledge.

Thus, data mining tools could help organizations to discover the hidden knowledge in the enormous amount of data. As a part of data mining research, this paper focuses on surveying data mining applications in knowledge management.

II. KNOWLEDGE MANAGEMENT

As technologies play an important role in KM, technologies stand to be a necessary tool for KM usage [1]. Knowledge is an expensive commodity, which if managed properly is a major asset to the company. Knowledge is a complex and fluid concept. It can be either explicit or tacit in nature. Explicit knowledge can be easily articulated and transferred to others. In contrast tacit knowledge, which is personal knowledge, residing in individual's heads, is very difficult to articulate, codified and communicate "Knowledge management (KM) is an effort to increase useful knowledge within the organization. Ways to do this include encouraging communication, offering opportunities to learn, and promoting the sharing of appropriate knowledge artifacts" This definition emphasizes the interaction aspect of knowledge management and organizational learning.

Knowledge management process focuses on knowledge flows and the process of creation, sharing, and distributing knowledge [5]. Each of knowledge units of capture and creation, sharing and dissemination, and acquisition and application can be facilitated by information technology.

KM technologies classifies in seven categories:

1. KM Framework
2. Knowledge-Based Systems (KBS)
3. Data Mining
4. Information and Communication Technology
5. Artificial Intelligence (AI)/Expert Systems (ES)
6. Database Technology (DT)
7. Modelling

III. DATA MINING

"Data Mining is the process of discovering actionable and meaningful patterns, profiles and trends by sniffing through your data using pattern recognition technologies such as neural networks, machine learning and genetic algorithms". DM tools can answer business questions that traditionally were too time consuming to resolve. They search databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectation. Data

mining is an essential step in the knowledge discovery in databases (KDD) process that produces useful patterns or models from data (Figure 1) [7]. The terms of KDD and data mining are different. KDD refers to the overall process of discovering useful knowledge from data. Data mining refers to discover new patterns from a wealth of data in databases by focusing on the algorithms to extract useful knowledge [7].

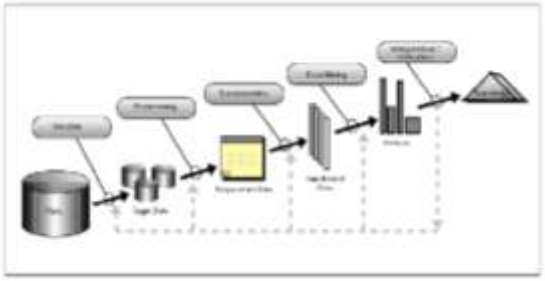

Figure 1 Data Mining and the KDD Process (Source: Fayyad, et.al., 1996)

In figure 1, KDD process consists of iterative sequence methods as follows [7, 9]:

1. *Selection:* Selecting data relevant to the analysis task from the database
2. *Pre-processing:* Removing noise and inconsistent data; combining multiple data sources
3. *Transformation:* Transforming data into appropriate forms to perform data mining
4. *Data mining:* Choosing a data mining algorithm which is appropriate to pattern in the data; extracting data patterns
5. *Interpretation/Evaluation:* Interpreting the patterns into knowledge by removing redundant or irrelevant patterns; translating the useful patterns into terms that human understandable

IV. Data Mining Tasks :

Six main functions of data mining:

1. Classification is finding models that analyze and classify a data item into several predefined classes
2. Regression is mapping a data item to a real-valued prediction variable
3. Clustering is identifying a finite set of categories or clusters to describe the data
4. Dependency Modeling (Association Rule Learning) is finding a model which describe significant dependencies between variables
5. Deviation Detection (Anomaly Detection) is discovering the most significant changes in the data.
6. Summarization is finding a compact description for a subset of data.

Data mining has two primary objectives of prediction and description. Prediction involves using some variables in data sets in order to predict unknown values of other relevant variables (*e.g.classification, regression, and anomaly detection*) Description involves finding human understandable patterns and trends in the data *(e.g. clustering, association rule learning, and summarization)* [8].

Some of the tasks solved by Data Mining are:

1. Prediction: a task of learning a pattern from examples and using the developed model to predict future values of the target variable.
2. Classification: a task of finding a function that maps records into one of several discrete classes.
3. Detection of relations: a task of searching for the most influential independent variables for a selected target variable.
4. Explicit modelling: a task of finding explicit formulae describing dependencies between various variables.
5. Clustering a task of identifying groups of records that are similar between themselves but different from the rest of the data.
6. Market Basket Analysis: processing transactional data in order to find those groups of products that are sold together well
7. Deviation Detection: a task of determining the most significant changes in some key measures of data from previous or expected values.

V. Data Mining techniques

Data Mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. Most companies already collect and refine massive quantities of data. The application areas of DM as contained in recent literatures as corroborated in Jiawei (2003) include: medical treatment/disease symptoms identification, retail industry, telephone calling patterns, DNA sequences, natural disaster, web log click stream, financial data analysis, bioinformatics, melody track selection, content-based e-mail processing systems, analyzes of data from specific experiments conducted over time, analysis of nation's census database, and so on.

DM techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line.

There are three groups of DM users namely, Application users, Designers and Theorists. It is usually common that the theorists based on some principal assumptions usually formulate new ideas. Therefore, some users are primarily interested in this group.

Those concerned with the application of DM such as knowledge Managers which as a direct result of their interest in DM research and design they are referred to as the 'DM researcher /designer'. Finally, the respondents concerned primarily with the using or solving problems, for which DM

offered an effective approach, are referred to as the "DM application group. The most commonly used techniques in data mining are:

1. *Artificial Neural Networks*: this is a nonlinear predictive model that learns through training and resembles biological neural networks in structure.
2. *Decision trees*: tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.
3. *Genetic Algorithms*: They are optimization techniques that use process such as genetics combination, mutation, and natural selection in a design based on concepts of evolution. It tries to mimic the way nature works. It is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetics.
4. *Rule Induction*: the extraction of useful if-then rules from data based on statistical significance.
5. *Regression Methods*: this tries to identify the best linear pattern in order to predict the value of one characteristic we are studying in relation to another.

VI. THE APPLICATIONS OF DATA MINING IN KNOWLEDGE MANAGEMENT

In the study, we divided knowledge resources into eight groups as that which knowledge object to be stored and manipulated in KM and how data mining aids.

1. *Health Care Organization***:** this domain was a use of the disease knowledge management system (KMS) of the hospital case study [10]. Data mining tool was used to explore diseases, operations, and tumors relationships. This tool used to build KMS to support clinical medicine in order to improve treatment quality [10].

2. *Retailing:* this was customer knowledge from household customers for product line and brand extension issues [13]; data mining can help and propose suggestions and solutions to the firm for product line and brand extensions. This doing by extracting market knowledge of customers, brands, products, and purchase data to fulfill the customers' demands behavior [13].

3. *Financial/Banking:* the domain knowledge covered financial and economic data; data mining can assist banking institutions making decision support and knowledge sharing processes to an enterprise bond classification [4].

4. *Small and Middle Businesses (food company and food supply chain):* there were two methods and processes to obtain knowledge resources: knowledge seeding-the relative knowledge to the problems; knowledge cultivating-the process to find the key knowledge from knowledge seeding [11]. Data mining and knowledge management integrated can help making better decisions [11]. As Death-On-Arrival (DOA) problem encountered in food supply chain networks (FSCN), Li et al. (2010) aimed to build Early Warning and Proactive Control (EW&PC) systems to solve such problems [12]. Knowledge Base was an important part of EW&PC systems. It contained data analysis by managers and organizes in an appropriate way for other managers. Data mining methods were helpful for the EW&PC systems [12].

5. *Entrepreneurial Science:* The knowledge resource was research assets in a knowledge institution [3]; there were three types of the research assets: research products, intellectual capital, and research programs. Data mining facilitated for knowledge extraction and helped guiding managers in determining strategies on knowledge-oriented organization competition [3].

6. Business: data collected from questionnaire, an intensive literature review, and discussions with four KM experts. Data mining can discover hidden patterns between KM and its performance for better KM implementations.

7. Collaboration and Teamwork: Worker's log and documents were analyzed each worker's referencing behavior and construct worker's knowledge flow. Data mining techniques can mine and construct group-based knowledge flows (GKFs) prototype for task-based groups [14].

8. Construction Industry: a large part of this enterprise information was available in the form of textual data formats [15]. This leads to the influence of text mining techniques to handle textual information source for industrial knowledge discovery and management solutions [15].

VII. CONCLUSION:

In this paper, we have discussed the data mining task and techniques can be integrated into KM and enhanced the KM process with better knowledge. It is clear that the data mining techniques will have a major impact on the practice of KM, and will present significance challenges for future knowledge and information systems research.

VII. REFERENCES

[1] An, X. & Wang, W. (2010). Knowledge management technologies and applications: A literature review. IEEE, 138-141. doi:10.1109/ICAMS.2010.5553046

[2] Berson, A., Smith, S.J. &Thearling, K. (1999). Building Data Mining Applications for CRM. New York: McGraw-Hill.

[3] Cantú, F.J. & Ceballos, H.G. (2010). A multiagent knowledge and information network approach for managing research assets. Expert Systems with Applications, 37(7), 5272-5284.

[4] Cheng, H., Lu, Y. & Sheu, C. (2009). An ontology-based business intelligence application in a financial knowledge management system.Expert Systems with Applications, 36, 3614–3622. doi:10.1016/j.eswa.2008.02.047

[5] Dalkir, K. (2005). Knowledge Management in Theory and Practice. Boston: Butterworth-Heinemann.

[6] Dawei, J. (2011). The Application of Date Mining in Knowledge Management.2011 International Conference on Management of e-Commerce and e-Government, IEEE Computer Society, 7-9. doi:10.1109/ICMeCG.2011.58

[7] Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases.AI Magazine, 17(3), 37-54.

[8] Gorunescu, F. (2011). Data Mining: Concepts, Models, and Techniques. India: Springer.

[9] Han, J. &Kamber, M. (2012). Data Mining: Concepts and Techniques. 3rd.ed. Boston: Morgan Kaufmann Publishers.

[10] Hwang, H.G., Chang, I.C., Chen, F.J. & Wu, S.Y. (2008). Investigation of the application of KMS for diseases classifications: A study in a Taiwanese hospital. Expert Systems with Applications, 34(1), 725-733. doi:10.1016/j.eswa.2006.10.018

[11] Li, X., Zhu, Z. & Pan, X. (2010). Knowledge cultivating for intelligent decision making in small & middle businesses.Procedia Computer Science, 1(1), 2479-2488. doi:10.1016/j.procs.2010.04.280

[12] Li, Y., Kramer, M.R., Beulens, A.J.M., Van Der Vorst, J.G.A.J. (2010). A framework for early warning and proactive control systems in food supply chain networks. Computers in Industry, 61, 852–862. Doi:101.016/j.compind.2010.07.010

[13] Liao, S.H., Chen, C.M., Wu, C.H. (2008). Mining customer knowledge for product line and brand extension in retailing.Expert Systems with Applications, 34(3), 1763-1776. doi:10.1016/j.eswa.2007.01.036

[14] Liu, D.R. & Lai, C.H. (2011). Mining group-based knowledge flows for sharing task knowledge. Decision Support Systems,50(2), 370-386. doi:10.1016/j.dss.2010.09.004

[15] Ur-Rahman, N. & Harding, J.A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. Expert Systems with Applications, 39, 4729- 4739. doi:10.1016/j.eswa.2011.09.124